

Prior-Agnostic Robust Forecast Aggregation

Zhi Chen*

Cheng Peng[†]

Wei Tang[‡]

Abstract

Robust forecast aggregation combines the predictions of multiple information sources to perform well in the worst case across all possible information structures. Previous work largely focuses on settings with a known binary state space, where the state is either 0 or 1. We study prior-agnostic robust forecast aggregation in which the aggregator observes only experts' reports, yet is ignorant of both the underlying joint information structure and the full prior, including the underlying state space. Unlike the standard model that fixes the binary state space $\{0, 1\}$, we allow the (binary) unknown state values to be arbitrary numbers in $[0, 1]$, so the same reported probability may correspond to very different realized outcome frequencies across environments.

Our main contribution is a simple, explicit, closed-form log-odds aggregator that linearly pools forecasts in logit space, together with (nearly-)tight minimax-regret guarantees across three knowledge regimes. We first show that under conditionally independent (CI) signals, robust aggregation with an unknown state space is strictly harder than in the known-state setting by establishing a larger lower bound, and our aggregation rule can achieve worst-case regret of 0.0255. Along the way, we also characterize tight regret bounds for Blackwell-ordered structures and for general information structures. In the classical setting with known state space $\{0, 1\}$, our aggregator achieves regret strictly below 0.0226 for CI structures. To the best of our knowledge, this is the first explicit closed-form aggregator that achieves a regret upper bound strictly less than 0.0226. Finally, we extend the model to allow the aggregator to additionally know each expert's marginal forecast distribution; in this setting, with the CI structures, we show that a generalized log-odds rule achieves regret of 0.0228, and we complement this guarantee with a lower bound of 0.0225.

*Chinese University of Hong Kong. Email: zhi.chen@cuhk.edu.hk

[†]Chinese University of Hong Kong. Email: chengpeng@link.cuhk.edu.hk

[‡]Chinese University of Hong Kong. Email: weitang@cuhk.edu.hk

1 Introduction

When checking tomorrow’s weather, reputable sources often sharply disagree on the likelihood of rain, leaving the downstream decision-maker unsure how to combine these multiple predictions into a single, reliable forecast. In practice, each expert (e.g., a forecasting service) forms their forecast based on observable covariates describing the current atmospheric state, such as pressure patterns, humidity levels, and radar signatures, which collectively encode a latent “likelihood of rain.” Importantly, these informational signals are generated from such features before the rain outcome is realized: they are informative about rain only through the underlying feature state,¹ rather than being conditionally drawn from the eventual binary rain/no-rain outcome itself.

The aggregator, however, typically observes only the experts’ reported probabilities, not the underlying features or the mapping from features to signals, while still hoping to choose an aggregation rule that performs well uniformly across all possible (and unknown) signal structures that could have produced the experts’ reports. If the features perfectly revealed whether it would rain, then the latent uncertainty collapses to a binary state space $\{0, 1\}$, recovering the classic robust aggregation problem with a known state space. This problem was first studied in [Arieli et al. \(2018\)](#) and later extended by [Neyman and Roughgarden \(2022\)](#); [Kong et al. \(2024\)](#); [Guo et al. \(2025\)](#). In reality, the features only partially predict rain, and because the aggregator does not observe them, they face a genuinely prior-agnostic robust aggregation problem in which the relevant “state” is the (unknown) latent likelihood encoded by those features, rather than a fixed and known binary variable.

This same feature-outcome separation appears in many real-world applications. For example, election forecasters base their predictions on poll responses, turnout proxies, demographic composition, and economic indicators, not on the election outcome that will only be observed later; physicians form prognoses from symptoms, lab tests, imaging, and patient history, not from the future health outcome itself; and financial analysts recommend portfolios using earnings reports, macro signals, order flow, and risk exposures, not the realized returns. In each case, experts’ forecasts are best viewed as posterior beliefs induced by intermediate features, while the aggregator must reconcile conflicting posteriors without observing either those features or the experts’ information-generating processes.

Motivated by the above observations, we generalize the robust forecast aggregation problem studied in [Arieli et al. \(2018\)](#), which assumes a known and fixed binary state space $\{0, 1\}$, to a setting in which the state space is unknown to the aggregator. More formally, we consider a stochastic environment in which there is an unobserved state that captures the underlying “situation” of the world relevant for predicting a binary outcome. Here, the state should be thought of as a latent driver of the outcome’s likelihood, such as an underlying atmospheric configuration that makes rain more or less likely, and, more generally, as any hidden condition that determines the event’s likelihood. To capture such latent uncertainty while keeping the model parsimonious, we allow the latent likelihood to take one of two (a priori unknown) states, drawn from an unknown prior distribution. The outcome is then generated conditionally on the realized state via a Bernoulli distribution, with a success probability equal to the likelihood encoded by that state. There are two experts, each observing a private signal stochastically related to the state through an unknown joint information structure, and then reporting a probabilistic forecast equal to their posterior belief about the outcome. The aggregator observes only the experts’ reported probabilities, and does not observe the state, the experts’ raw signals, the prior, or the signal generation process, and must map the pair of reports into a single forecast. We evaluate the aggregator’s performance by

¹For example, the weather forecasting service may observe a summary of atmospheric measurements, such as upstream radar reflectivity and local humidity profiles, which together provide a noisy cue about tomorrow’s rainfall.

squared-error loss, and regret is defined relative to an omniscient Bayesian benchmark that knows the full environment and has access to the experts’ signals. Our goal is to design an aggregation function that achieves small worst-case regret uniformly over all priors (including the unknown states) and information structures (or a specified class of structures).

1.1 Our Contributions

We initiate the study of prior-agnostic robust forecast aggregation with an unknown state space, motivated by settings in which experts’ reports are induced by latent feature states that are not directly observed by the aggregator. Relative to the classic robust aggregation model with a known binary state space, our formulation treats the “state” as an unknown latent likelihood that corresponds to the event/outcome probability. Our goal is to design an aggregation rule whose performance is uniformly good over all priors and information structures consistent with the experts’ reports, including uncertainty about the underlying state space itself.

Our first technical contribution is a simple, explicit, closed-form log-odds aggregation function that linearly pools experts’ forecasts in logit space. In the setting of conditionally independent signals (i.e., experts’ signals are distributed independently conditional on the realized state), we provide a near-tight minimax regret bound: our aggregation function achieves worst-case regret 0.0255 (Theorem 3.2), and we complement this upper bound by showing that the optimal scheme cannot do much better. In particular, we show a lower bound that no aggregation scheme can achieve regret below $31/1326 \approx 0.0234$ (Theorem 3.4). This lower bound demonstrates a strict separation between the unknown-state problem and the known-state benchmark: under the setting with a known binary state space $\{0, 1\}$, Arieli et al. (2018) have established a regret lower bound of $\frac{1}{8}(5\sqrt{5}-11) \approx 0.0225$, and Guo et al. (2025) present an algorithmic approach that can achieve regret 0.0226, which is strictly below 0.0234. Taken together, these results show that robust aggregation is strictly harder when the state space is not known a priori in environments with conditionally independent signals.

We also evaluate how aggregation schemes designed for the known-state setting perform when the state space is unknown. As shown in Table 2, all methods we evaluate, including those proposed in Arieli et al. (2018); Kong et al. (2024), incur worst-case regret that is strictly above 0.03, which is substantially larger than the regret achieved by our proposed aggregation scheme. We also apply our log-odds aggregation scheme back to the classical known-binary-state setting. In the known state space setting $\{0, 1\}$, the same log-odds family (with an appropriate parameter choice) achieves regret that is strictly below 0.0226 (Proposition 3.5). To the best of our knowledge, this is the first regret upper bound in this regime achieved by an explicit closed-form aggregator.²

Beyond conditional independence, we also obtain sharp minimax characterizations in the unknown-state-space setting for two additional settings with different classes of information structures. For Blackwell-ordered information structures (one expert is uniformly more informative than the other), we show that not knowing the state values does not hurt the worst-case aggregation performance: the minimax regret is exactly $\frac{1}{8}(5\sqrt{5}-11)$, matching the known- $\{0, 1\}$ setting (Corollary 4.1).

²Consistent with the approach adopted in prior works (Arieli et al., 2018; Guo et al., 2025), we numerically estimate and compare the worst-case regrets of our proposed aggregator against other common benchmarks. In all of our numerical experiments, we evaluate the worst-case performance of our proposed aggregation function using a two-stage, high-precision procedure designed to avoid missing the true worst-case instance. First, we run a global heuristic search over the relevant family of information structures to locate a small set of candidate worst-case instances. Second, we independently validate each candidate via a fine-grained grid refinement (discretization step size of 10^{-5}). For example, in the known $\{0, 1\}$ state-space setting, this pipeline consistently identifies a worst-case regret of 0.022599, and the subsequent refinement confirms that this value is numerically stable and strictly below 0.0226.

Moreover, this bound is achieved by directly applying the precision-weighted aggregation scheme proposed in Arieli et al. (2018), as the scheme depends solely on the observed forecasts (via their “precision,” i.e., inverse variance) and not on the unknown state values or prior. For fully general information structures (no restrictions on correlation), the robust aggregation is fundamentally limited: the minimax regret is exactly $1/4$. This is achieved by the trivial constant rule that always outputs $1/2$, and the matching lower bound follows from an inclusion argument (the adversary can restrict attention to the classic binary-state hard instance) (Corollary 4.2), where a lower bound $1/2$ in the known state space setting is known (Arieli et al., 2018).

Finally, we explore the setting in which the aggregator additionally knows each expert’s marginal forecast distribution (i.e., the unconditional distribution of the expert’s posterior mean). Note that this extra information can indeed make some environments essentially trivial. In Blackwell-ordered settings, the aggregator can compare the informativeness implied by the two marginal distributions (via the induced Blackwell order), identify the better-informed expert, and simply follow that expert, achieving zero regret. Likewise, in the classical conditionally independent model with a known binary state space $\{0, 1\}$, observing an expert’s marginal forecast distribution pins down the prior mean via Bayes plausibility. Since the omniscient Bayesian benchmark under conditional independence depends on the information structure only through this prior mean, aggregation again becomes trivial once the marginals are known.

Somewhat surprisingly, however, knowing the marginals does not, in general, eliminate worst-case difficulty in our unknown-state setting. For general information structures, the minimax regret remains exactly 0.25 even with known marginals. Even under conditionally independent structures, we prove a minimax lower bound of $\frac{1}{8}(5\sqrt{5}-11)$ (Theorem 5.1): unlike the known- $\{0, 1\}$ setting, the adversary can exploit different state spaces that induce the same marginal forecast distributions, so the marginals do not reveal which environment is realized. We then show that a generalized log-odds aggregation function, which in fact uses only the inferred prior mean (rather than the full marginals), achieves a nearly matching guarantee: for a suitable choice of parameters, it attains the worst-case regret of at most 0.022763 (Proposition 5.3).

We summarize most of our results in the following Table 1.

Setting	Conditionally Independent		Blackwell-Ordered	General
	LB	UB	Tight Bound	Tight Bound
Known state space $\Theta = \{0, 1\}$	$\frac{5\sqrt{5}-11}{8}^\dagger$	0.022599 (Proposition 3.5)	$\frac{5\sqrt{5}-11}{8}$	0.25
Unknown state space	$\frac{31}{1326}^\ddagger$ (Theorem 3.4)	0.025512 (Theorem 3.2)	$\frac{5\sqrt{5}-11}{8}$ (Corollary 4.1)	0.25 (Corollary 4.2)
Unknown state space Known marg. predic. dist.	$\frac{5\sqrt{5}-11}{8}$ (Theorem 5.1)	0.022763 (Proposition 5.3)	0	0.25 (Proposition 5.4)

$^\dagger \approx 0.022542$; $^\ddagger \approx 0.023379$.

Table 1: Summary of regret bounds. Baseline results are rounded to 4 decimal places; our results are reported to 6 decimal places. All rounding follows the standard nearest value rule.

1.2 Related Work

Our work closely relates to the literature on robust forecast aggregation, which aims to find aggregators that perform well in the worst case across all possible information structures. The foundational question of designing such an aggregator was first studied in [Arieli et al. \(2018\)](#). Focusing on the two-expert setting with a binary state space $\{0, 1\}$ and conditionally independent signals, they propose low-regret aggregation rules and establish corresponding upper and lower bounds on the additive regret. Specifically, they show that an average prior scheme achieves a worst-case regret of 0.0260, while proving a lower bound of $\frac{5\sqrt{5}-11}{8} \approx 0.0225$. Building on this setting, recent work by [Guo et al. \(2025\)](#) develop an efficient algorithmic framework that can achieve a regret of 0.0226. These two works are the closest to ours, since we also focus on conditionally independent signals. Our contribution differs in that we study an unknown state-space setting and show that previously proposed aggregation schemes can perform poorly in this regime.

Complementing these approaches, several studies consider settings in which the aggregator is endowed with additional information. [Levy and Razin \(2022\)](#) study robust prediction when the aggregator understands each information source in isolation but is uncertain about the correlation across sources.

[De Oliveira et al. \(2021\)](#) study the robust decision making under uncertain correlations among information sources where the uncertainty lies solely in the correlation structure among information sources. [Levy and Razin \(2021\)](#) propose a maximum-likelihood method for combining multiple forecasts (given a prior) that rationalizes the observed forecasts via the most likely underlying information structure.

Beyond additive regret, other works explore alternative robustness criteria. For example, [Neyman and Roughgarden \(2022\)](#); [Frongillo et al. \(2025\)](#) study a ratio-based robustness notion for a class of information structures they term projective substitutes. [Arieli et al. \(2025\)](#) study binary decision aggregation problem under different standard robustness notions, including maximin, regret minimization, and approximation ratio, and establish that the simple random dictator rule (which selects one agent’s recommendation uniformly at random) is the unique optimal mechanism when the number of agents is large.

More recently, the literature has broadened its scope to account for behavioral or strategic deviations by experts. For example, [Kong et al. \(2024\)](#) study aggregation where experts exhibit base rate neglect. [Guo and Kong \(2025\)](#) consider settings with adversarial experts and characterize the robustly optimal aggregation mechanisms under various loss functions. Complementary to this strand, another line of research leverages second-order information, e.g., agents’ beliefs about others’ answers, to improve aggregation performance. This line of work is initiated with Bayesian Truth Serum ([Prelec, 2004](#)) and the “surprisingly popular” rule ([Prelec et al., 2017](#)), and has more recently developed a broader toolkit that uses second- and higher-order belief elicitation to aggregate opinions in finite populations, including settings with heterogeneous agent types (see, e.g., [Palley and Soll, 2019](#); [Wang et al., 2021](#); [Wilkening et al., 2022](#); [Pan et al., 2024](#); [Ai et al., 2025](#)).

In addition to aggregating forecasts in a robust manner, forecast aggregation has traditionally been also studied through two complementary lenses: axiomatic and Bayesian. The axiomatic literature posits normative desiderata for aggregation rules, e.g., unanimity preservation and variants of independence, and characterizes the (often essentially unique) rules that satisfy them, with linear pooling (averaging) emerging as a canonical solution under standard axioms (see, e.g., [Aczél and Wagner, 1980](#); [Genest, 1984](#); [Dietrich and List, 2016](#)). In contrast, the Bayesian literature models how forecasts are generated from underlying signals and prescribes aggregation by Bayesian updating; most results in this tradition are parametric, optimizing aggregation within a specified family

of information structures (e.g., common-knowledge correlation patterns or exponential-family sampling models), which can limit robustness when the true dependence structure is misspecified (see, e.g., Satopää et al., 2014; Frongillo et al., 2015; Ernst et al., 2016; Satopää et al., 2016).

2 Preliminary

We consider a stochastic environment where the world randomly generates a *state* θ from a finite state space Θ . In this work, we focus on a binary state space, so that $|\Theta| = 2$. Once a state is realized, it further induces a random *binary outcome*, which is unobservable to all players. Specifically, for each state $i \in [2]$, the binary outcome $\omega \in \Omega \triangleq \{0, 1\}$ is drawn from a Bernoulli distribution with mean $\theta_i \in [0, 1]$. Without loss of generality, we assume that the states are sorted in non-decreasing order of their Bernoulli means: $\theta_1 \leq \theta_2$. Since each state can be identified by its Bernoulli mean, we henceforth refer to the states directly as θ_1 and θ_2 , and write the state space as $\Theta = \{\theta_1, \theta_2\}$. We denote by $\lambda_i \in [0, 1]$ the prior probability of the state θ_i for $i \in [2]$ and let μ be the prior mean of the state, i.e., $\mu \triangleq \mathbb{E}_{\theta \sim \lambda}[\theta]$.

Information structure There are two experts, each receiving a private signal that provides information about the current world state. We use an *information structure* to characterize the relationship between the experts' *private* signals and the underlying states. Formally, an information structure for two experts is a tuple (S, π) , where $S = S_1 \times S_2$ is a 2-dimensional signal space and $\pi \in \Delta(\Theta \times S)$ is a joint distribution over states and signal profiles. Each expert $i \in [2]$ receives a private signal $s_i \in S_i$ drawn according to π . Upon receiving their private signal s_i and knowing the information structure π , each expert i provides his forecast $x_i(s_i | \pi)$ about the underlying outcome ω . These forecasts are their Bayesian posterior for outcome: since the outcome space is binary, each expert's forecast is simply their conditional probability that $\omega = 1$, i.e.,

$$x_i(s_i | \pi) \triangleq \Pr_{\pi}[\omega = 1 | s_i] .$$

Aggregation and loss An aggregator, who does not observe the signals or the true state, must form her own forecast about the underlying outcome based solely on the experts' reported forecasts $x(s) = (x_1(s_1), x_2(s_2))$. Unlike in prior work, the aggregator is both information-structure-agnostic and prior-agnostic: she knows only that the state space is binary but does not know the specific values of θ_1 and θ_2 , nor the prior probabilities λ_1, λ_2 , nor the information structure. More formally, the aggregator uses a deterministic aggregation function $f : [0, 1]^2 \rightarrow [0, 1]$ that is a mapping from the two experts' forecasts $(x_1, x_2) \in [0, 1]^2$ to her own forecast.

We evaluate the aggregator's performance using a loss function $\ell(\cdot, \cdot) : [0, 1] \times \Omega \rightarrow \mathbb{R}_+$ where $\ell(x, \omega) \in \mathbb{R}_+$ quantifies the loss when the induced outcome is $\omega \in \Omega$ and the aggregator's prediction is $x \in [0, 1]$. Throughout this work, we adopt the squared loss $\ell(x, \omega) = (x - \omega)^2$ as our loss function.

We benchmark the aggregator's performance against the omniscient (Bayesian optimal) aggregator, who knows the full model: the exact state space $\Theta = \{\theta_1, \theta_2\}$, the prior λ , the information structure π , and observes all the signals of all the experts. With the square loss function, the omniscient aggregator's optimal forecast is her Bayesian posterior expectation of ω given the received reports from the experts and the underlying information structure. We denote by $f^{\text{BAYES}} : [0, 1]^2 \rightarrow [0, 1]$ such Bayesian optimal aggregation:

$$f^{\text{BAYES}}(x(s)) \triangleq \Pr_{\pi}[\omega = 1 | s]$$

where $s = (s_1, s_2)$ is the full signal profile. Note that while f^{BAYES} is a function of the signals s , it can be equivalently viewed as a function of the reported forecasts $x(s)$, since the reports are deterministic functions of the signals under a fixed π .

Prior-agnostic robust forecasting aggregation problem With these definitions, we are ready to formulate the aggregator’s prior-agnostic robust forecasting problem.

Given a state space $\Theta \subset [0, 1]^2$ and an information structure $\pi \in \Delta(\Theta \times S)$, the expected loss of the aggregation function f is

$$L(f \mid \Theta, \pi) \triangleq \mathbb{E}_{(\theta, s) \sim \pi, \omega \sim \text{BERN}(\theta)} \left[\ell(f(x(s)), \omega) - \ell(f^{\text{BAYES}}(x(s)), \omega) \right] .$$

Here $\text{BERN}(\theta)$ denotes a Bernoulli distribution with the mean θ .

Following the prior work (see, e.g., Arieli et al., 2018; Guo et al., 2025), we allow the aggregator to have partial knowledge of the information structure, modeled as a known class Π of possible structures. The prior-agnostic robust forecasting problem solves the following the minimax problem, which aims to find an aggregation function f that minimizes this worst-case regret over all admissible state spaces and information structures in Π :

$$\min_f \mathbf{Reg}_{\Pi}[f] .$$

where $\mathbf{Reg}_{\Pi}[f] \triangleq \sup_{\Theta \subset [0, 1]^2, \pi \in \Pi} L(f \mid \Theta, \pi)$. When the state space Θ is known and fixed to the aggregation and set to be $\{0, 1\}$, then our prior-agnostic robust forecasting problem reduces to the robust forecasting aggregation problem studied in Arieli et al. (2018); Guo et al. (2025).

An exemplification of our framework We illustrate the prior-agnostic robust forecasting problem through an emergency-department triage setting, where a *human* decision maker aggregates two probabilistic risk assessments.

Consider a human clinician (the aggregator) who must combine two probability forecasts from two sources (the experts), e.g., an AI risk score and a specialist’s judgment, to assess whether a patient will experience an adverse event. The latent state $\theta \in \Theta = \{\theta_1, \theta_2\}$ represents the patient’s underlying risk regime (low vs. high), and the realized binary outcome $\omega \in \Omega = \{0, 1\}$ indicates whether the adverse event occurs. Each expert $i \in \{1, 2\}$ observes a private signal $s_i \in S_i$ (such as model-specific features or clinical cues) and reports a forecast $x_i \in [0, 1]$; the aggregator observes only the pair (x_1, x_2) and outputs a final forecast using an aggregation rule.

In practice, the clinician typically does not know the numerical values of the baseline risks θ_1, θ_2 (they vary across hospitals, populations, protocols, and over time), nor the prior likelihood of each regime (e.g., case mix shifts during an outbreak). Moreover, she rarely knows the information structure linking the experts’ private signals to the latent state: the two sources may share data, overlap in cues, or change after retraining or workflow updates, creating unknown dependence and calibration patterns. The prior-agnostic robust forecasting problem therefore asks for an aggregation rule that performs well against an omniscient benchmark uniformly over a plausible class of such information structures.

The expected loss simplification With the squared loss, we can simplify the expected loss of the aggregation function $L(f \mid \Theta, \pi)$ as follows:

Lemma 2.1 (Regret simplification Arieli et al., 2018; Kong et al., 2024; Guo et al., 2025). *Given any state space $\Theta \subset [0, 1]^2$ and any information structure π , we have that*

$$L(f \mid \Theta, \pi) = \mathbb{E}_{(\theta, s) \sim \pi, \omega \sim \text{BERN}(\theta)} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right)^2 \right] .$$

The proof is deferred to Section A.

3 A Robust and Prior-Agnostic Aggregation Strategy

In this section, we focus on a widely studied and economically relevant class of information structures (see, e.g., Arieli et al., 2018; Kong et al., 2024; Guo et al., 2025) in which experts' signals are independent conditional on the realized state. We refer to such structures as conditionally independent (CI), and denote by Π_{CI} the set of all CI information structures. Formally, for a conditionally independent information structure $\pi \in \Delta(\Theta \times S)$ with $S = S_1 \times S_2$, we say π is CI if for every state $\theta \in \Theta$ and every signal profile $(s_1, s_2) \in S$, we have

$$\pi(s_1, s_2 \mid \theta) = \pi_1(s_1 \mid \theta) \times \pi_2(s_2 \mid \theta),$$

where $\pi_i(\cdot \mid \theta)$ is the conditional signal distribution for expert i . When convenient, we write $\pi = \pi_1 \circ \pi_2$ to emphasize this conditional-independence factorization.

3.1 A Log-Odds Aggregation Function

We propose the following family of aggregation functions:

Definition 3.1 (Log-odds aggregation function). Fix a parameter $\alpha \in [0, 1]$, a *log-odds aggregation scheme* is a function $f_\alpha(\cdot, \cdot) : [0, 1]^2 \rightarrow [0, 1]$ mapping from the two experts' reports $(x_1, x_2) \in [0, 1]^2$ to a forecast as follows:

$$f_\alpha(x_1, x_2) = \frac{1}{1 + \left(\frac{1-x_1}{x_1}\right)^\alpha \cdot \left(\frac{1-x_2}{x_2}\right)^\alpha}.$$

Equivalently, let $\text{LOGIT}(p) \triangleq \ln\left(\frac{p}{1-p}\right)$ denote the log-odds function of probability $p \in [0, 1]$, aggregator f_α satisfies that

$$\text{LOGIT}(f_\alpha(x_1, x_2)) = \alpha \cdot \text{LOGIT}(x_1) + \alpha \cdot \text{LOGIT}(x_2).$$

To obtain intuitions behind the above aggregation function, we start with characterizing the omniscient Bayesian forecast under the conditionally independent information structure. Fix a state space $\Theta = \{\theta_1, \theta_2\}$ with the prior mean $\mu = \mathbb{E}_{\theta \sim \lambda}[\theta]$:

$$\begin{aligned} f^{\text{BAYES}}(x_1, x_2) &= \Pr_\pi[\omega = 1 \mid x_1, x_2] \\ &= \mathbb{E}_{\omega \sim \text{BERN}(\theta_1)}[\omega] \cdot \Pr_\pi[\theta = \theta_1 \mid x_1, x_2] + \mathbb{E}_{\omega \sim \text{BERN}(\theta_2)}[\omega] \cdot \Pr_\pi[\theta = \theta_2 \mid x_1, x_2] \\ &= \theta_1 \cdot \left(1 - \frac{1}{1 + \frac{\mu - \theta_1}{\theta_2 - \mu} \cdot \frac{\theta_2 - x_1}{x_1 - \theta_1} \cdot \frac{\theta_2 - x_2}{x_2 - \theta_1}}\right) + \theta_2 \cdot \frac{1}{1 + \frac{\mu - \theta_1}{\theta_2 - \mu} \cdot \frac{\theta_2 - x_1}{x_1 - \theta_1} \cdot \frac{\theta_2 - x_2}{x_2 - \theta_1}}. \end{aligned} \quad (1)$$

As we can see, this expression highlights two key structural features. First, the joint posterior depends on the two reports only through the product of two generalized odds ratios

$$\frac{\theta_2 - x_i}{x_i - \theta_1},$$

which quantifies how strongly expert i 's report tilts beliefs toward the low state relative to the high state. Second, under conditional independence, the experts' evidence combines multiplicatively in odds space: the Bayesian update multiplies their odds contributions and then applies a logistic-type transformation (i.e., the outer $\frac{1}{1+}$ transformation).

An additive log-odds representation of Bayesian forecast Recall that for any probability $p \in (0, 1)$, the log-odds (or logit) is given by $\text{LOGIT}(p) \triangleq \ln\left(\frac{p}{1-p}\right)$. Then the omniscient posterior probability of the high state satisfies

$$\text{LOGIT}(\Pr[\theta = \theta_2 \mid x_1, x_2]) = \ell_{\theta_1, \theta_2}(x_1) + \ell_{\theta_1, \theta_2}(x_2) - \ln \frac{\mu - \theta_1}{\theta_2 - \mu}$$

where $\ell_{\theta_1, \theta_2}(x) \triangleq \ln \frac{x - \theta_1}{\theta_2 - x}$ is the generalized log-odds induced by the endpoints θ_1 and θ_2 . Thus, Bayesian aggregation under conditional independence amounts to linear pooling in log-odds: the aggregator adds the experts' log-odds contributions (plus the prior log-odds), and then maps back via the logistic function.

Our aggregation function in Definition 3.1 implements the same linear-in-log-odds principle on the canonical $[0, 1]$ scale by taking $\theta_1 \leftarrow 0, \theta_2 \leftarrow 1$. In this sense, f_α can be viewed as a symmetric “tempered Bayes” rule: it still adds evidence on the log-odds scale, but uses a parameter α to control the strength of aggregation.

Why f_α may work for unknown Θ When the state space $\Theta = \{\theta_1, \theta_2\}$ is unknown, an aggregator cannot form the correct generalized odds $\frac{x - \theta_1}{\theta_2 - x}$ nor the correct prior term $\ln \frac{\mu}{1 - \mu}$. In other words, the exact Bayesian forecast is unavailable because the aggregator cannot normalize a report x relative to the true endpoints (i.e., θ_1, θ_2). The key observation from the Bayesian formula above is that what matters is not the endpoints per se, but a “monotone transformation” that turns reports into comparable units of evidence and then adds them. The log-odds aggregator provides a robust surrogate for this transformation: it uses the universal mapping $x \mapsto \ln \frac{x}{1-x}$, which preserves the essential ordering

“higher reports \Rightarrow stronger evidence for the high state,”

and combines two experts by adding their evidence due to conditional independence. Moreover, the parameter $\alpha \in [0, 1]$ plays a central robustness role. While the canonical scale preserves the ordering of evidence, simply summing logits (i.e., $\alpha = 1$) fails to subtract the unknown prior log-odds, effectively double-counting the prior information.³ Additionally, full summation implicitly assumes perfect conditional independence, which may not hold in practice. Hence α can be interpreted as an “influence weight” or “trust weight” on each expert: higher α implies higher confidence in the independence and informativeness of the signals relative to the prior; lower α is more conservative, correcting for the double-counting of the prior and potential correlation between experts. Importantly, this rule remains implementable without knowing the primitives: it requires only the reported forecasts (x_1, x_2) and a single parameter α , and still mirrors the Bayesian logic that independent evidence should be combined by multiplying odds.

³For binary states $\{\theta_1, \theta_2\}$, expert i 's report $x_i = \mathbb{E}[\theta \mid s_i]$ decomposes as $\ell_{\theta_1, \theta_2}(x_i) = \ell_{\theta_1, \theta_2}(\mu) + \ln(\pi_i(s_i \mid \theta_2) / \pi_i(s_i \mid \theta_1))$, where $\ell_{\theta_1, \theta_2}(z) \triangleq \ln \frac{z - \theta_1}{\theta_2 - z}$. Thus, the prior log-odds $\ell_{\theta_1, \theta_2}(\mu)$ is embedded in each report. Under conditional independence, the joint posterior log-odds satisfies

$$\ell_{\theta_1, \theta_2}(f^{\text{BAYES}}) = \ell_{\theta_1, \theta_2}(\mu) + \sum_i \ln(\pi_i(s_i \mid \theta_2) / \pi_i(s_i \mid \theta_1)) = \ell_{\theta_1, \theta_2}(x_1) + \ell_{\theta_1, \theta_2}(x_2) - \ell_{\theta_1, \theta_2}(\mu) .$$

Summing log-odds *without* subtracting $\ell_{\theta_1, \theta_2}(\mu)$ thus double-counts the prior. Numerically, if $(\theta_1, \theta_2) = (0, 1)$, $\mu = 0.8$ (odds 4), and $x_1 = x_2 = 8/9$ (odds 8), the naive sum yields odds 64 ($p \approx 0.985$), whereas the correct Bayesian update yields odds 16 ($p \approx 0.941$).

3.2 Establishing the Regret Bounds

In this section, we first establish that there exists a log-odds aggregator that achieves a regret upper bound of 0.025512, and we then establish a regret lower bound of 0.02344 for our prior-agnostic robust forecasting problem.

Upper bound analysis We first present the regret upper bound of our proposed aggregation function.

Theorem 3.2. *With $\alpha = 0.585$, the aggregation function f_α guarantees worst-case regret at most 0.025512; that is,*

$$\mathbf{Reg}_{\Pi_{CI}}[f_{0.585}] \approx 0.025512 < 0.0256 .$$

We prove Theorem 3.2 in two steps: Step 1 shows that, without loss of generality, it suffices to consider environments in which each expert's information structure generates at most two possible signals. Step 2 leverages this reduction to reformulate the worst-case regret of f_α as a finite-dimensional optimization problem involving only seven parameters. Optimizing over these parameters, together with α , yields a computable regret upper bound, which we solve numerically to obtain the stated guarantee.

Lemma 3.3 (Reduction to binary signals). *Fix any state space $\Theta = \{\theta_1, \theta_2\}$, given any aggregation function f , we have*

$$\sup_{\pi \in \Pi_{CI}} L(f \mid \Theta, \pi) \leq \sup_{\pi \in \Pi_{CI}: |\text{supp}(\pi_i)| \leq 2, \forall i \in [2]} L(f \mid \Theta, \pi) .$$

That is, the worst case regret can be achieved for the information structures that at most two signals are generated for each expert.

Proof. Fix any state space $\Theta = \{\theta_1, \theta_2\}$. Define $\Delta \triangleq \theta_2 - \theta_1$.

If $\Delta = 0$, then the prior distribution is degenerate and we have $f^{\text{BAYES}} \equiv \theta_1$, so the lemma statement is trivial. Hence we may assume $\Delta > 0$.

Define an indicator variable $\delta \in \{0, 1\}$ such that $\theta = \theta_1 + \Delta \cdot \delta$. For each expert, consider its posterior probability

$$p_i(s_i) \triangleq \Pr_{\pi}[\delta = 1 \mid s_i] \in [0, 1] .$$

Then by definition, we have

$$x_i(s_i) = \mathbb{E}[\theta \mid s_i] = \theta_1 + \Delta \mathbb{E}[\delta \mid s_i] = \theta_1 + \Delta p_i(s_i). \quad (2)$$

Similarly, with the joint posterior probability $p_{12}(s_1, s_2) \triangleq \Pr_{\pi}[\delta = 1 \mid s_1, s_2]$, we have

$$f^{\text{BAYES}}(x_1(s_1), x_2(s_2)) = \mathbb{E}[\theta \mid s_1, s_2] = \theta_1 + \Delta p_{12}(s_1, s_2) . \quad (3)$$

We now construct a new information structure π^\dagger with the state space $\Theta^\dagger \triangleq \{0, 1\}$ where its prior distribution λ^\dagger satisfies that $\Pr_{\theta \sim \lambda^\dagger}[\theta = 0] = \Pr_{\theta \sim \lambda}[\theta = \theta_1]$, and

$$\pi^\dagger(0, s) \leftarrow \pi(\theta, s), \quad \text{if } \theta = \theta_1, \forall s \in S ; \quad \pi^\dagger(1, s) \leftarrow \pi(\theta, s), \quad \text{if } \theta = \theta_2, \forall s \in S . \quad (4)$$

Then given $(\Theta^\dagger, \pi^\dagger)$, we next define the aggregator $f^\dagger : [0, 1]^2 \rightarrow \mathbb{R}$ as follows:

$$f^\dagger(p_1, p_2) \triangleq \frac{f(\theta_1 + \Delta p_1, \theta_1 + \Delta p_2) - \theta_1}{\Delta} . \quad (5)$$

Then combining (2)–(5),

$$\begin{aligned} f(x_1(s_1), x_2(s_2)) - f^{\text{BAYES}}(x_1(s_1), x_2(s_2)) &= \Delta \cdot f^\dagger(p_1(s_1), p_2(s_2)) + \theta_1 - (\theta_1 + \Delta p_{12}(s_1, s_2)) \\ &= \Delta \left(f^\dagger(p_1(s_1), p_2(s_2)) - p_{12}(s_1, s_2) \right) , \end{aligned}$$

and thus we have

$$L(f \mid \Theta, \pi) = \Delta^2 \cdot \mathbb{E}_{\pi^\dagger} \left[\left(f^\dagger(p_1(s_1), p_2(s_2)) - p_{12}(s_1, s_2) \right)^2 \right] = \Delta^2 \cdot L(f^\dagger \mid \{0, 1\}, \pi^\dagger) .$$

The conditional independence $s_1 \perp s_2 \mid \theta$ is equivalent to $s_1 \perp s_2 \mid \delta$ under the bijection $\theta = \theta_1 + \Delta\delta$, and the signal-space is unchanged. Let $\pi^\dagger = \pi_1^\dagger \circ \pi_2^\dagger$. Thus, it suffices to prove the corresponding statement for the normalized $\{0, 1\}$ state space setting:

$$\sup_{\pi^\dagger \in \Pi_{\text{CI}}} L(f \mid \{0, 1\}, \pi^\dagger) \leq \sup_{\pi^\dagger \in \Pi_{\text{CI}}: |\text{supp}(\pi_i^\dagger)| \leq 2, \forall i \in [2]} L(f \mid \{0, 1\}, \pi^\dagger) . \quad (6)$$

To prove Eqn. (6), we follow the analysis used in Arieli et al. (2018). Let P^i be the marginal distribution of $p_i(s_i)$ on $[0, 1]$:

$$P^i(A) \triangleq \Pr_{\pi_i^\dagger}[p_i \in A], \quad \forall \text{ Borel set } A \subseteq [0, 1] .$$

Then Bayes plausibility implies that

$$\int_0^1 p P^i(dp) = \mathbb{E}[p_i(s_i)] = \Pr[\delta = 1] = \lambda_2 .$$

Then we can reformulate the expected relative loss $L(f^\dagger \mid \{0, 1\}, \pi^\dagger)$ as follows:

$$B_{\lambda_2}(f^\dagger \mid P^1, P^2) \triangleq L(f^\dagger \mid \{0, 1\}, \pi^\dagger) = \iint_{[0,1]^2} \Phi_{\lambda_2}(p_1, p_2) P^1(dp_1) P^2(dp_2),$$

where

$$\Phi_{\lambda_2}(p_1, p_2) \triangleq \left(\frac{p_1 p_2}{\lambda_2} + \frac{(1-p_1)(1-p_2)}{1-\lambda_2} \right) \left(f^\dagger(p_1, p_2) - f^{\dagger, \text{BAYES}}(p_1, p_2) \right)^2 .$$

As we can see, $(P^1, P^2) \mapsto B_{\lambda_2}(f^\dagger \mid P^1, P^2)$ is a bilinear function, i.e., it is affine in each marginal P^i . Now fix $\lambda_2 \in (0, 1)$ and define the convex set of mean- λ_2 probability measures:

$$K(\lambda_2) \triangleq \left\{ P \text{ probability measures on } [0, 1] : \int_0^1 p P(dp) = \lambda_2 \right\} .$$

Let $K_2(\lambda_2) \subseteq K(\lambda_2)$ be the subset consisting of measures supported on at most two points. By the fact that every extreme point of the set $K(\lambda_2)$ must be supported on at most two points, we know

$$\sup_{P^1, P^2 \in K(\lambda_2)} B_{\lambda_2}(f^\dagger \mid P^1, P^2) = \sup_{P^1, P^2 \in K_2(\lambda_2)} B_{\lambda_2}(f^\dagger \mid P^1, P^2) .$$

We thus finish the proof of Eqn. (6). □

Proof of Theorem 3.2. Fix any state space $\Theta = \{\theta_1, \theta_2\}$ with the prior distribution λ . With Lemma 3.3, it is without loss to consider an information structure where each expert receives two signals, namely, $S_1 = S_2 = \{0, 1\}$. Given any aggregation rule $f : [0, 1]^2 \rightarrow [0, 1]$, the expected loss under the above binary-signal conditionally independent structure $\pi = \pi_1 \circ \pi_2$ is

$$L(f \mid \Theta, \pi) = \sum_{(s_1, s_2) \in \{0, 1\}^2} \Pr_{\pi}[s_1, s_2] \cdot \left(f(x_1(s_1), x_2(s_2)) - f^{\text{BAYES}}(x_1(s_1), x_2(s_2)) \right)^2 .$$

where $\Pr_{\pi}[s_1, s_2] = \lambda_1 \cdot \pi_1(s_1 | \theta_1) \cdot \pi_2(s_2 | \theta_1) + \lambda_2 \cdot \pi_1(s_1 | \theta_2) \cdot \pi_2(s_2 | \theta_2)$.

Now define the variables $a_1 \leftarrow \pi_1(s_1 | \theta_1), b_1 \leftarrow \pi_2(s_1 | \theta_1), a_2 \leftarrow \pi_1(s_1 | \theta_2), b_2 \leftarrow \pi_2(s_1 | \theta_2)$. Then we can see that we can express $L(f | \Theta, \pi)$ using only these 4 variables. Thus, fixing any aggregation function f_{α} for some $\alpha \in [0, 1]$, we can reformulate the program $\mathbf{Reg}_{\Pi_{\text{CI}}}[f_{\alpha}]$ of identifying its worst-case regret as the following optimization that involves 7 variables:

$$\mathbf{Reg}_{\Pi_{\text{CI}}}[f_{\alpha}] = \sup_{\Theta \in [0,1]^2: \theta_1 \leq \theta_2} \sup_{\lambda_2 \in [0,1]; a_1, b_1, a_2, b_2 \in [0,1]^4} L(f_{\alpha} | \Theta, \pi) .$$

After discretizing the space $[0, 1]$ for the parameter α , we solve the above seven-variable optimization problem at each grid point. We then select the value of α that minimizes the resulting worst-case regret $\mathbf{Reg}_{\Pi_{\text{CI}}}[f_{\alpha}]$. Numerical calculations reveal that the worst-case regret is minimized at $\alpha \approx 0.585$. Under this optimal aggregator, the global maximum of the regret function is attained at the configuration where $\theta_1 = 0.21097, \theta_2 = 1.000$, and the prior mean $\mu = 0.65866$. Specifically, the experts' signal structures a_i, b_i correspond to the following reports: expert 1 provides $x_1^{\dagger} = 0.21097$ and $x_1^{\ddagger} = 0.87002$; expert 2 provides $x_2^{\dagger} = 0.21097$ and $x_2^{\ddagger} = 0.87002$. The resulting minimum worst-case regret is equal to 0.025512. \square

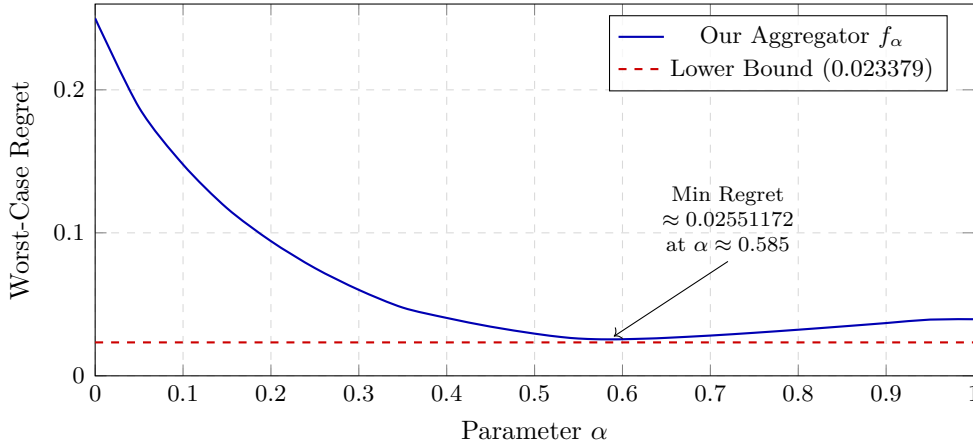


Figure 1: The blue curve is the worst-case regret for the aggregation function f_{α} for every $\alpha \in \{0, 0.001, 0.002, \dots, 1\}$, which achieves a global minimum of approximately 0.0255 when $\alpha \approx 0.585$. The dashed red line represents the theoretical lower bound of 0.023379.

In Arieli et al. (2018), the authors have proposed three aggregation schemes when the state space is known and fixed to be $\Theta = \{0, 1\}$. In this known-state benchmark, the omniscient Bayesian aggregator, who knows the information structure and the prior mean μ , outputs the forecast

$$\begin{aligned} f^{\text{BAYES}}(x_1, x_2) &= \frac{x_1 x_2 (1 - \mu)}{x_1 x_2 (1 - \mu) + (1 - x_1)(1 - x_2) \mu} \\ &= \frac{1}{1 + \left(\frac{1-x_1}{x_1}\right) \left(\frac{1-x_2}{x_2}\right) \cdot \frac{\mu}{1-\mu}} \triangleq g(\mu, x_1, x_2) . \end{aligned}$$

The three methods proposed in Arieli et al. (2018) are summarized as follows:

- The Simple Averaging aggregation scheme (i.e., always output the forecast $\frac{x_1+x_2}{2}$).
- The Average Prior aggregation scheme that outputs a Bayesian forecast by assuming the underlying prior mean is $\frac{x_1+x_2}{2}$, namely, it outputs $g\left(\frac{x_1+x_2}{2}, x_1, x_2\right)$.

- The Heuristic Prior (SOTA) aggregation scheme that utilizes a heuristic function $ep(x_1, x_2)$ to estimate the prior mean, namely, it outputs $g(ep(x_1, x_2), x_1, x_2)$.⁴

Aggregator	Prior Estimate ($\hat{\mu}$)	Worst-Case Regret
Simple Averaging	-	0.0625
Average Prior (Arieli et al., 2018)	$\frac{x_1 + x_2}{2}$	0.0311
Heuristic Prior (Arieli et al., 2018)	$ep(x_1, x_2)^\dagger$	0.0303
Kong et al. (2024) ($\lambda = 0.8$)	$\frac{x_1 + x_2}{2}$	0.0298
Our Aggregator (f_α)	$\alpha \approx 0.585$	0.025512

Table 2: The regret comparison in the unknown state space setting.

Additionally, we consider the parameterized aggregator family proposed by Kong et al. (2024). For reported forecasts (x_1, x_2) , their aggregator is defined as:

$$f_{ap}^{\hat{\lambda}}(x_1, x_2) = \frac{(1 - \hat{\mu})^{2\hat{\lambda}-1} x_1 x_2}{(1 - \hat{\mu})^{2\hat{\lambda}-1} x_1 x_2 + \hat{\mu}^{2\hat{\lambda}-1} (1 - x_1)(1 - x_2)}, \quad (7)$$

where the prior proxy $\hat{\mu}$ is set to the arithmetic mean $\frac{x_1 + x_2}{2}$. For this scheme, we performed a grid search with a step size of 0.1 to optimize the parameter $\lambda \in [0, 1]$ to minimize the worst-case regret. In our setting, the optimal value was found at $\lambda = 0.8$.

These schemes can also be applied in our unknown-state-space setting by treating the reported forecasts (x_1, x_2) as inputs to the same formulas. However, as we demonstrate in Table 2, even with the optimized λ , the methods from Arieli et al. (2018) and Kong et al. (2024) incur worst-case regrets that are substantially larger than the regret achieved by our proposed aggregation scheme.

Lower bound analysis Recall that our prior-agnostic robust forecasting aggregation problem strictly generalizes the canonical setting studied in prior work, which assumes a fixed and known binary state space $\{0, 1\}$. Consequently, any lower bound that holds in that more restrictive model immediately carries over to our setting. In particular, Arieli et al. (2018) prove that, under the known $\{0, 1\}$ -state-space assumption, every aggregation function incurs worst-case regret at least $\frac{5\sqrt{5}-11}{8} \approx 0.022542$ for this known state space $\{0, 1\}$ setting, directly applies as a valid lower bound in our problem. This value therefore remains a valid lower bound for our problem as well. We next show that allowing an unknown state space (i.e., prior-agnosticism) leads to a strictly stronger impossibility result, yielding a strictly larger lower bound on the minimax regret.

Theorem 3.4. *For any aggregation function f , it follows that*

$$\inf_f \mathbf{Reg}_{\Pi_{CI}}[f] \geq \frac{31}{1326} \approx 0.023379 .$$

That is, no aggregation function can guarantee a regret lower than 0.023379.

⁴The heuristic function is defined as $ep(x_1, x_2) = 0.49(x_1 + x_2) + 0.02 \cdot \mathbb{I}(x_1 + x_2 > 1)$ in Arieli et al. (2018).

Structure	State (θ)	Prior (λ)	Signal Probabilities $\Pr[s \mid \theta]$	
			$s = s_L$	$s = s_H$
π^A	0	$\frac{1}{2}$	1	0
	$\frac{5}{6}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{4}$
π^B	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{4}$
	1	$\frac{1}{2}$	0	1

Table 3: The two symmetric adversarial information structures. π^A yields noisy low signals and revealing high signals; π^B yields revealing low signals and noisy high signals.

To establish the regret lower bound, we construct an adversarial distribution over two conditionally independent information structures such that even an aggregator who knows the adversary’s mixing rule cannot guarantee regret below 0.023379.

Specifically, the adversary draws an expert’s information structure uniformly at random, i.e., choosing π^A or π^B with probability $1/2$ each. In both structures, the signal space is binary, $S = \{s_L, s_H\}$, and the induced posterior means are $x_L = 1/6$ and $x_H = 5/6$, respectively (see the detailed construction in Table 3). These two information structure differs in their state space and the way that how they generate signals. Under π^A , the state space is $\Theta^A = \{0, 5/6\}$ with the uniform prior, whereas under π^B , the state space is $\Theta^B = \{1/6, 1\}$ with the uniform prior. After the information structure π^I for some $I \in \{A, B\}$ is realized, both experts receive independent private signals generated according to that realized structure π^I , namely, the joint information structure is $\pi = \pi^I \circ \pi^I$.

Proof of Theorem 3.4. Fix any aggregation rule f . We construct an adversary’s mixed strategy over two conditionally independent information structures π^A, π^B . The adversary draws π^A and π^B uniformly. By Yao’s principle, we have that

$$\mathbf{Reg}_{\Pi_{\text{CI}}}[f] = \sup_{\Theta, \pi \in \Pi_{\text{CI}}} L(f \mid \Theta, \pi) \geq \mathbb{E}_{(\Theta^I, \pi^I)} [L(f \mid \Theta^I, \pi^I \circ \pi^I)] .$$

Therefore, we have

$$\inf_f \mathbf{Reg}_{\Pi_{\text{CI}}}[f] \geq \inf_f \mathbb{E}_{(\Theta^I, \pi^I)} [L(f \mid \Theta^I, \pi^I \circ \pi^I)] .$$

It remains to lower bound the right-hand side.

In both structures, the experts receive binary signals $s \in \{s_L, s_H\}$. Let $x \in \{x_L, x_H\}^2$ be the realized report profile, where x_L, x_H , by construction, satisfy that

$$x_L \triangleq \sum_{\theta \in \Theta^I} \theta \cdot \Pr_{\pi^I}[\theta \mid s_L] = \frac{1}{6} ; \quad x_H \triangleq \sum_{\theta \in \Theta^I} \theta \cdot \Pr_{\pi^I}[\theta \mid s_H] = \frac{5}{6} .$$

Given the expert’s information structure π^I for $I \in \{A, B\}$, we use $f_{\pi^I \circ \pi^I}^{\text{BAYES}}(x)$ to denote the omniscient aggregator’s forecast given the received experts’ reports $x \in \{x_L, x_H\}^2$ under the joint information structure $\pi^I \circ \pi^I$. By construction, we have that

$$f_{\pi^A \circ \pi^A}^{\text{BAYES}}(x) = \begin{cases} \frac{5}{102} & \text{if } x = (x_L, x_L) \\ \frac{5}{6} & \text{otherwise ;} \end{cases} \quad f_{\pi^B \circ \pi^B}^{\text{BAYES}}(x) = \begin{cases} \frac{97}{102} & \text{if } x = (x_H, x_H) \\ \frac{1}{6} & \text{otherwise .} \end{cases}$$

We compute the marginal probability for the experts' reports $x \in \{x_L, x_H\}^2$ as follows:

$$\begin{aligned} \Pr_{\pi^A \circ \pi^A}[x_L, x_L] &= \frac{17}{32}, & \Pr_{\pi^A \circ \pi^A}[x_H, x_H] &= \frac{9}{32}, & \Pr_{\pi^A \circ \pi^A}[x_L, x_H] &= \Pr_{\pi^A \circ \pi^A}[x_H, x_L] = \frac{3}{32}; \\ \Pr_{\pi^B \circ \pi^B}[x_L, x_L] &= \frac{9}{32}, & \Pr_{\pi^B \circ \pi^B}[x_H, x_H] &= \frac{17}{32}, & \Pr_{\pi^B \circ \pi^B}[x_L, x_H] &= \Pr_{\pi^B \circ \pi^B}[x_H, x_L] = \frac{3}{32}. \end{aligned}$$

We next analyze the aggregator's optimal forecast given that he knows the mixed strategy of the adversary. For a report profile x , the optimal forecast $f^*(x)$ minimizes the expected squared loss:

$$f^*(x) = \frac{\Pr_{\pi^A \circ \pi^A}[x] f_{\pi^A \circ \pi^A}^{\text{BAYES}}(x) + \Pr_{\pi^B \circ \pi^B}[x] f_{\pi^B \circ \pi^B}^{\text{BAYES}}(x)}{\Pr_{\pi^A \circ \pi^A}[x] + \Pr_{\pi^B \circ \pi^B}[x]}.$$

Calculating this for each profile yields:

$$f^*(x_L, x_L) = \frac{7}{78}, \quad f^*(x_H, x_H) = \frac{71}{78}, \quad f^*(x_L, x_H) = f^*(x_H, x_L) = \frac{1}{2}.$$

Thus, the aggregator's regret is calculated as:

$$\begin{aligned} \mathbb{E}_{(\Theta^I, \pi^I)} \left[L(f \mid \Theta^I, \pi^I \circ \pi^I) \right] &= \sum_{I \in \{A, B\}} \Pr[\pi^I] \sum_{x \in \{x_L, x_H\}^2} \left(f^*(x) - f_{\pi^I \circ \pi^I}^{\text{BAYES}}(x) \right)^2 \cdot \Pr_{\pi^I \circ \pi^I}[x] \\ &= \frac{31}{1326}. \end{aligned}$$

This yields the lower bound $\inf_f \mathbf{Reg}[f] \geq \frac{31}{1326}$. □

3.3 The Superior Performance of Aggregator f_α in Known-State Setting

We also evaluate our proposed aggregation function f_α in the setting where the state space is known and fixed to be $\Theta = \{0, 1\}$. In this specific setting, we show that there exists an $\alpha \in [0, 1]$ such that the aggregation function achieves a regret upper bound that almost matches the theoretical lower bound $\frac{5\sqrt{5}-11}{8} \approx 0.0225425$ that has been established in [Arieli et al. \(2018\)](#).

Proposition 3.5. *In the state space setting with $\Theta = \{0, 1\}$, the aggregation function with f_α $\alpha = 0.5168$ guarantees a worst-case regret at most 0.022599;*

$$\mathbf{Reg}_{\Pi_{\text{CI}}}[f_{0.5168}] \approx 0.022599 < 0.0226.$$

As we can see, the above upper bound (0.022599), together with the theoretical lower bound derived from [Arieli et al. \(2018\)](#), demonstrate that the optimality gap of our proposed aggregation function is almost negligible:

$$\inf_{\alpha \in [0, 1]} \mathbf{Reg}_{\Pi_{\text{CI}}}[f_\alpha] - \frac{5\sqrt{5}-11}{8} \approx 6 \times 10^{-5},$$

which implies that in the known binary-state setting, our proposed strategy is near-optimal, extracting essentially all achievable information from the experts' forecasts without requiring knowledge of the underlying signal structure.

To corroborate the tightness of our theoretical bounds, we also conduct a numerical sensitivity analysis of the aggregation strategy with respect to parameter α . Figure 2 plots the worst-case regret when the parameter α varies. The curve exhibits a global minimum in the neighborhood of $\alpha \approx 0.5168$. Numerical calculations reveal that the worst-case instance for this known $\{0, 1\}$ state space setting is with the prior mean $\mu = 0.81293$. Specifically, the experts' signal structures correspond to the following reports: expert 1 provides $x_1^\dagger = 0.20043$ and $x_1^\ddagger = 0.99783$; expert 2 provides $x_2^\dagger = 0.81293$ and $x_2^\ddagger = 0.81293$, which gives a maximum regret 0.022599.

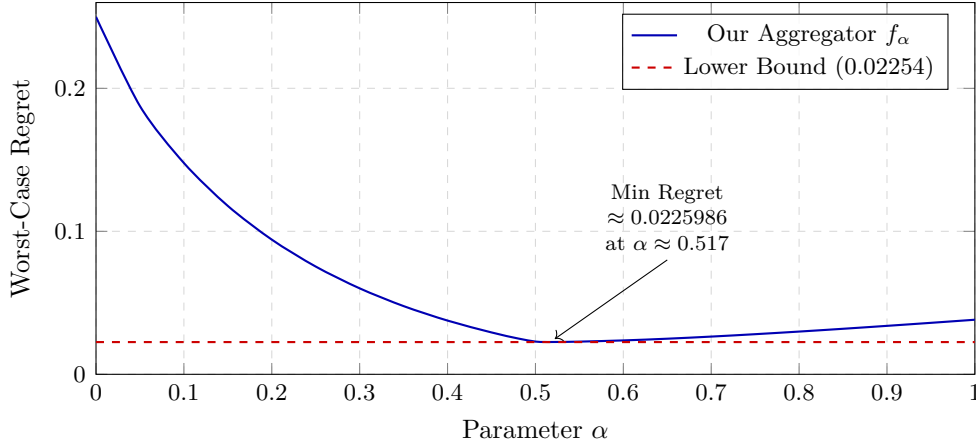


Figure 2: Sensitivity analysis of the worst-case regret with respect to the parameter α . The blue curve depicts the performance of f_α . The regret decreases sharply as α approaches 0.517, achieving a global minimum that converges to the theoretical lower bound (dashed red line).

Compare to other aggregation schemes As mentioned, [Arieli et al. \(2018\)](#) propose three aggregation schemes for the known state space $\Theta = \{0, 1\}$ setting. However, it is known that none of these aggregation schemes achieve a regret strictly lower than 0.0250. We also consider the aggregator family from [Kong et al. \(2024\)](#). In this known-state benchmark, we optimized the parameter $\lambda \in [0, 1]$ and found that the minimum worst-case regret is achieved at $\lambda = 1$. At this value, the formula in (7) simplifies to the Bayesian aggregator with a prior $\hat{\mu}$, effectively degenerating to the Average Prior method of [Arieli et al. \(2018\)](#).

In recent work, [Guo et al. \(2025\)](#) have developed an algorithmic framework for this known-state setting that computes a near-optimal aggregator via efficient approximation procedures (e.g., based on discretized reports or Lipschitz-type regularity restrictions). They show that their framework produces an aggregator whose regret nearly matches the known lower bound. However, their aggregator is obtained through a computational pipeline involving discretization, numerical optimization, and interpolation, which leads to an implicit mapping rather than an explicit closed-form aggregation rule. In contrast, we propose a simple, one-parameter aggregator f_α (Definition 3.1) that is directly implementable and easy to use: one simply plugs in the experts' report (x_1, x_2) and set $\alpha \leftarrow 0.517$ to achieve performance that is comparably close to the theoretical lower bound (within 6×10^{-5}), without the overhead of running a separate approximation algorithm or constructing the aggregation rule numerically.

We summarize these comparisons in Table 2.

Aggregator	Prior Estimate ($\hat{\mu}$)	Worst-Case Regret
Simple Averaging	-	0.0625
Average Prior (Arieli et al., 2018)	$\frac{x_1+x_2}{2}$	0.0260
State-of-the-art (Arieli et al., 2018)	$ep(x_1, x_2)^\dagger$	0.0250
Kong et al. (2024) ($\lambda = 1$)	$\frac{x_1+x_2}{2}$	0.0260
Guo et al. (2025)	-	0.0226
Our Aggregator (f_α)	$\alpha \approx \mathbf{0.5168}$	0.022599

† The heuristic function is $ep(x_1, x_2) = 0.49(x_1 + x_2) + 0.02 \cdot \mathbb{I}(x_1 + x_2 > 1)$.

Table 4: The regret comparison in the known state space setting with $\Theta = \{0, 1\}$.

4 Blackwell-Ordered and General Information Structures

In this section, we analyze the other two classes of information structures: the Blackwell-ordered setting where the experts’ signals can be ranked according to the Blackwell order; and the general setting where no assumption is imposed on the information structures.

4.1 Blackwell-Ordered Information Structures

In the Blackwell-ordered setting, we show that in this setting, the lack of knowledge regarding the state values Θ and the prior λ does not penalize the aggregator. By leveraging the results from Arieli et al. (2018), we demonstrate that, even without knowledge of the state values, the aggregator achieves the same minimax regret bound as established in the known state space setting.

Let Π_{BO} be the class of Blackwell-ordered information structures. Formally, for any $\pi \in \Pi_{\text{BO}}$, we assume without loss of generality that Expert 2 is *Blackwell-more-informative* than Expert 1. This implies that Expert 1’s signal s_1 can be viewed as a noisy garbling of Expert 2’s signal s_2 (or equivalently, s_2 is a sufficient statistic for the pair (s_1, s_2) with respect to the state θ).

Corollary 4.1. *For the Blackwell-ordered information structures, we have that $\inf_f \mathbf{Reg}_{\Pi_{\text{BO}}}[f] \geq \frac{5\sqrt{5}-11}{8} \approx 0.0225$. There exists an aggregation function that achieves $\mathbf{Reg}_{\Pi_{\text{BO}}}[f] = \frac{5\sqrt{5}-11}{8}$.*

To establish the upper bound, we directly adopt the aggregation scheme proposed by Arieli et al. (2018) for the Blackwell-ordered setting. We observe that, although their algorithm was originally derived for a known state space, the construction of the aggregation function relies *solely* on the experts’ reported forecasts. Consequently, we can explicitly reuse this algorithm in our prior-agnostic robust forecasting problem without any modification.

Specifically, we employ the *precision-weighted average* scheme as analyzed in Arieli et al. (2018). We define the precision of a forecast $x \in (0, 1)$ as the inverse of its variance, denoted by $\phi(x) \triangleq \frac{1}{x(1-x)}$. The aggregation function $f_{\text{pre}} : [0, 1]^2 \rightarrow [0, 1]$ is defined as follows:

$$f_{\text{pre}}(x_1, x_2) = \begin{cases} \frac{\sqrt{\phi(x_1)x_1} + \sqrt{\phi(x_2)x_2}}{\sqrt{\phi(x_1)} + \sqrt{\phi(x_2)}} & \text{if } |x_1 - x_2| > 0.4 \\ \frac{\phi(x_1)x_1 + \phi(x_2)x_2}{\phi(x_1) + \phi(x_2)} & \text{if } |x_1 - x_2| \leq 0.4 \end{cases}$$

with appropriate boundary handling for extreme forecasts. Crucially, as this aggregation rule depends *only* on the observed reports x and is independent of the unknown state values Θ or the prior λ , it constitutes a valid and robust strategy in our setting.

Proof of Corollary 4.1. We first prove the lower bound, then prove the upper bound.

Lower bound The lower bound follows from an inclusion argument. Since the adversary is free to select the standard binary state space $\Theta = \{0, 1\}$, the set of admissible environments in our framework (where Θ is unknown) is a superset of the environments where the state space is known and fixed to be $\{0, 1\}$. Consequently, the minimax regret in our setting cannot be lower than the minimax regret in the benchmark setting. Note that in both settings, the aggregator is agnostic to the prior distribution. Therefore, the worst-case construction from Arieli et al. (2018) remains valid, which established a lower bound of $\frac{5\sqrt{5}-11}{8}$, that applies to our setting also.

Upper bound The proof relies on establishing an equivalence between our prior-agnostic robust forecasting problem with unknown state space and the optimization problem solved in Arieli et al. (2018).

First, consider the omniscient benchmark. In the class of Blackwell-ordered information structures Π_{BO} , one expert is strictly more informed than the other and the optimal strategy for an omniscient aggregator is simply to follow the prediction of the more-informed expert. Since the aggregator does not know the identity of the more-informed expert, we consider the worst-case scenario where the identity is randomized. Without loss of generality for a specific realization, assume Expert 2 is more informed. The signal s_2 is sufficient for the state θ , implying the Bayesian optimal forecast is simply $f^{\text{BAYES}}(x(s)) = x_2$.

Using Lemma 2.1 (Regret Simplification), our objective function simplifies to:

$$\mathbf{Reg}_{\Pi_{\text{BO}}}[f_{\text{pre}}] = \mathbb{E} \left[\left(f_{\text{pre}}(x(s)) - f^{\text{BAYES}}(x(s)) \right)^2 \right] = \mathbb{E} \left[(f_{\text{pre}}(x(s)) - x_2)^2 \right]. \quad (8)$$

We now observe that the optimization problem in Eq. (8) is independent of the specific values of the state space $\Theta = \{\theta_1, \theta_2\}$. The expectation depends *solely* on the joint distribution of the experts' forecasts (x_1, x_2) . Besides, the set of feasible joint distributions of forecasts corresponds exactly to the set of martingales with a fixed prior mean, regardless of the underlying state values that generated these forecasts.

Consequently, minimizing the worst-case regret in our setting reduces to finding an aggregation function f that minimizes the expected squared distance to the more-informed forecast under the worst-case. This is mathematically identical to the optimization problem formulated and solved in Theorem 1 of Arieli et al. (2018). Therefore, their derived bound applies directly to our setting:

$$\min_f \mathbf{Reg}_{\Pi_{\text{BO}}}[f_{\text{pre}}] \leq \frac{5\sqrt{5} - 11}{8}.$$

This establishes that the precision scheme f_{pre} remains robust optimal in our unknown state setting, as the guaranteed upper bound matches the earlier derived lower bound. \square

4.2 General Information Structures

In this section, we analyze the performance of the aggregator under general information structures, where no structural assumptions (such as Blackwell-order or conditionally independent) are imposed

on the experts' signals. We show that in this unrestricted setting, information aggregation is effectively impossible: the optimal strategy is the trivial forecast of $1/2$, and the minimax regret is exactly $1/4$.

Corollary 4.2. *For the general information structures, we have that $\inf_f \mathbf{Reg}_\Pi[f] = 0.25$. There exists a trivial aggregation function $f \equiv 0.5$ that achieves $\mathbf{Reg}_\Pi[f] = 0.25$.*

Proof. The proof proceeds by establishing matching upper and lower bounds.

Upper bound Consider the trivial aggregation strategy $f_{\text{triv}}(x) = 1/2$ for all inputs. The expected squared loss of this strategy against any outcome $\omega \in \{0, 1\}$ is:

$$\begin{aligned} \mathbf{Reg}_\Pi[f_{\text{triv}}] &= \mathbb{E} \left[\left(\frac{1}{2} - \omega \right)^2 \right] \\ &= \Pr[\omega = 1] \left(\frac{1}{2} - 1 \right)^2 + \Pr[\omega = 0] \left(\frac{1}{2} - 0 \right)^2 \\ &= \Pr[\omega = 1] \cdot \frac{1}{4} + \Pr[\omega = 0] \cdot \frac{1}{4} \\ &= \frac{1}{4}. \end{aligned}$$

Therefore, the regret upper bound under general information structures is 0.25.

Lower bound The lower bound follows directly from an inclusion argument. The adversary is free to choose the standard state space $\Theta = \{0, 1\}$. Therefore, the minimax regret in our framework is at least as high as in [Arieli et al. \(2018\)](#) setting, inheriting the lower bound of 0.25. \square

5 Knowing the Marginal Prediction Distributions

In our previous discussions, we have assumed that the aggregator observes only the experts' reports and has no further information about either the underlying information structure or the state space. In this section, we instead consider a setting in which the aggregator additionally knows the marginal distributions of the experts' forecasts (i.e., the unconditional distribution of each expert's posterior mean) ([De Oliveira et al., 2021](#); [Levy and Razin, 2022](#)), however, the underlying joint information structure and the state space still remain unobservable to the aggregator.

In some environments, this additional knowledge can make aggregation essentially trivial. For instance, in Blackwell-ordered settings, where one expert is uniformly more informative than the other, this additional knowledge can make the problem essentially trivial: the aggregator can identify the better-informed expert by comparing the informativeness implied by their respective marginal forecast distributions (via the induced Blackwell order) and then simply follow that expert and get a regret of 0. Similarly, in the setting of conditionally independent information structures with a fixed and known $\{0, 1\}$ state space, knowing the marginal prediction distribution will also make the problem trivial. In this case, each expert's marginal forecast distribution must satisfy Bayes plausibility with respect to the prior distribution, so observing the marginal distribution allows the aggregator to infer the prior mean. Since, under conditional independence, the omniscient Bayesian benchmark depends on the information structure only through this prior mean (see Eqn. (1)), the aggregation problem again becomes trivial once the marginals are known.

Given these examples, it is natural to expect that knowing the marginal forecast distributions might generally enable the aggregator to design a more sophisticated rule and achieve strictly

smaller worst-case regret. Perhaps surprisingly, this is not the case. We show first that over the class of general information structures, the aggregator does not gain from this additional information: the minimax (worst-case) regret lower bound is unchanged, matching the benchmark setting in which the aggregator observes only the reports (see Proposition 5.4). Moreover, even when we restrict attention to conditionally independent information structures, the minimax regret lower bound remains the same as the lower bound in the setting where the state space is fixed to be $\{0, 1\}$ and known to the aggregator (see Theorem 5.1). We then show that a generalized version of our proposed log-odds aggregation scheme can achieve small regret (see Proposition 5.3).

Below, we first present our results for conditionally independent information structures, and then present our results for the general information structures.

5.1 Conditionally Independent Information Structures

We first present a regret lower bound.

Theorem 5.1. *With knowledge of each expert’s marginal forecast distribution, it follows that for any aggregation function f , we have*

$$\inf_f \mathbf{Reg}_{\Pi_{CI}}[f] \geq \frac{5\sqrt{5} - 11}{8} \approx 0.022542 .$$

That is, no aggregation function can guarantee a regret lower than 0.022542.

Recall that in Arieli et al. (2018), they also established a same lower bound under the conditionally independent information structures when the aggregator knows the state space to be $\{0, 1\}$. However, their lower bound instance cannot be directly applied to our setting as their constructions are all based on known state space $\{0, 1\}$. Together with the knowledge of the marginal forecast distribution, the aggregator can infer the prior mean and thus achieve 0 regret.

To obtain a non-trivial lower bound, one must design information structures with different state spaces. In particular, we construct a mixed adversary that randomizes between two single-expert information structures π^A and π^B in Table 5 (each chosen with probability 1/2). Here $\pi^I \in \Delta(\Theta^I \times S)$ for $I \in \{A, B\}$ specifies one expert’s signal distribution over the common signal space $S = \{s_L, s_H\}$. The induced joint information structure is then $\pi^I \circ \pi^I$, the two experts’ signals are generated (conditionally) independently and identically according to π^I . The key feature is that, although the two structures have *different state spaces*, they induce the *same marginal forecast distribution* for each expert. Indeed, in both π^A and π^B , each expert observes a binary signal $s \in \{s_L, s_H\}$ with $\Pr[s_L] = \Pr[s_H] = 1/2$, and the expert’s report (i.e., his posterior mean) takes the same two values (corresponding to s_L and s_H). Consequently, even when the aggregator knows the marginal forecast distributions, this information does not reveal which information structure has been realized. That is, from the aggregator’s viewpoint, this “additional information” is the same under $\pi^A \circ \pi^A$ and $\pi^B \circ \pi^B$.

The lower bound follows from an indistinguishability argument at the level of joint reports. Conditional on observing a report pair (x_1, x_2) , the aggregator must output a single forecast $f(x_1, x_2)$ that will be used regardless of whether the realized joint structure is $\pi^A \circ \pi^A$ or $\pi^B \circ \pi^B$.⁵ However, the omniscient Bayesian benchmark, who knows the realized structure, produces different posterior means under different realization for some report profiles, because the likelihood ratios induced by the two structures differ even though the marginal distributions of reports coincide. Therefore, on

⁵Recall that we write $\pi = \pi_1 \circ \pi_2$ to denote a conditionally independent joint information structure π obtained by combining the two experts’ individual information structures π_1 and π_2 (i.e., expert i ’s marginal information structure is π_i).

those profiles the aggregator faces an unavoidable tradeoff: choosing $f(x_1, x_2)$ close to the Bayesian forecast under $\pi^A \circ \pi^A$ necessarily incurs a nontrivial squared error under $\pi^B \circ \pi^B$, and vice versa, which leads to a non-trivial regret lower bound.

Structure	State (θ)	Prior (λ)	Signal Probabilities $\Pr[s \theta]$	
			$s = s_L$	$s = s_H$
π^A	$\frac{3-\sqrt{5}}{4}$	$\frac{\sqrt{5}-1}{2}$	$\frac{1+\sqrt{5}}{4}$	$\frac{3-\sqrt{5}}{4}$
	1	$\frac{3-\sqrt{5}}{2}$	0	1
π^B	0	$\frac{3-\sqrt{5}}{2}$	1	0
	$\frac{1+\sqrt{5}}{4}$	$\frac{\sqrt{5}-1}{2}$	$\frac{3-\sqrt{5}}{4}$	$\frac{1+\sqrt{5}}{4}$

Table 5: Adversary’s mixed strategy of the information structures. In both information structures, each signal is observed with a marginal probability of $1/2$.

Proof of Theorem 5.1. Both the information structures share a common signal space $S = \{s_L, s_H\}$. Information structure $\pi^A \in \Delta(\Theta^A \times S)$ is defined over the state space Θ^A with the prior distribution λ^A defined as follows:

$$\Theta^A = \left\{ \frac{3-\sqrt{5}}{4}, 1 \right\}, \quad \Pr_{\theta \sim \lambda^A} \left[\theta = \frac{3-\sqrt{5}}{4} \right] = \frac{\sqrt{5}-1}{2},$$

while information structure $\pi^B \in \Delta(\Theta^B \times S)$ is defined over the state space Θ^B with the prior distribution λ^B defined as follows:

$$\Theta^B = \left\{ 0, \frac{1+\sqrt{5}}{4} \right\}, \quad \Pr_{\theta \sim \lambda^B} \left[\theta = \frac{1+\sqrt{5}}{4} \right] = \frac{\sqrt{5}-1}{2}.$$

As we can see, both information structure shares the same prior mean $\mu = 0.5$ although they have different state spaces.

We will then construct each information structure π^I for $I \in \{A, B\}$, such that whenever the expert receives the signal $s \in S$, the expert’s posterior mean (i.e., his forecast) over the states satisfies that

$$x_L \triangleq \sum_{\theta \in \Theta^I} \theta \cdot \Pr_{\pi^I}[\theta | s_L] = \frac{3-\sqrt{5}}{4}; \quad x_H \triangleq \sum_{\theta \in \Theta^I} \theta \cdot \Pr_{\pi^I}[\theta | s_H] = \frac{1+\sqrt{5}}{4}.$$

In the meanwhile, the marginal forecast distribution is the same under the information structures π^A and π^B , which is,

$$\Pr_{\pi^I}[x_L] = \Pr_{\pi^I}[x_H] = 0.5, \quad \forall I \in \{A, B\}. \quad (9)$$

Since each expert receives their signal from the same realized information structure, the expert’s marginal forecast distribution is also the same.

We next analyze the regret of any aggregation rule, even when the aggregator knows the marginal forecast distribution of each expert. By Yao’s principle, we have that

$$\mathbf{Reg}_{\Pi_{\text{Cl}}}[f] = \sup_{\Theta, \pi \in \Pi_{\text{Cl}}} L(f | \Theta, \pi) \geq \mathbb{E}_{(\Theta^I, \pi^I)} \left[L(f | \Theta^I, \pi^I \circ \pi^I) \right].$$

Therefore, we have

$$\inf_f \mathbf{Reg}_{\Pi_{\text{CI}}}[f] \geq \inf_f \mathbb{E}_{(\Theta^I, \pi^I)} \left[L(f \mid \Theta^I, \pi^I \circ \pi^I) \right].$$

It remains to lower bound the right-hand side.

Given the information structure π^I for $I \in \{A, B\}$, let $f_{\pi^I}^{\text{BAYES}}(x)$ denote the omniscient aggregator's forecast given the received report x :

$$f_{\pi^{A \circ \pi^A}}^{\text{BAYES}}(x) = \begin{cases} \frac{15 - 5\sqrt{5}}{4} & \text{if } x = (x_H, x_H) \\ \frac{3 - \sqrt{5}}{4} & \text{otherwise;} \end{cases} \quad f_{\pi^{B \circ \pi^B}}^{\text{BAYES}}(x) = \begin{cases} \frac{5\sqrt{5} - 11}{4} & \text{if } x = (x_L, x_L) \\ \frac{1 + \sqrt{5}}{4} & \text{otherwise.} \end{cases}$$

We compute the marginal probability for the experts reporting x under each information structure. In this construction, the marginal distributions are symmetric across structures:

$$\begin{aligned} \Pr_{\pi^I \circ \pi^I}[x_L, x_L] &= \Pr_{\pi^I \circ \pi^I}[x_H, x_H] = \frac{1 + \sqrt{5}}{8}; \\ \Pr_{\pi^I \circ \pi^I}[x_L, x_H] &= \Pr_{\pi^I \circ \pi^I}[x_H, x_L] = \frac{3 - \sqrt{5}}{8}. \end{aligned} \tag{10}$$

We next analyze the aggregator's optimal forecast given that he knows the mixed strategy of the adversary. Let the adversary pick $I \in \{A, B\}$ uniformly at random. Fix any aggregation function f that may depend on the known forecast marginal distribution (i.e., Eqn. (9) in our construction). Knowing the adversary's mixed strategy, the expected regret of the aggregation function f is

$$\begin{aligned} & \mathbb{E}_{(\Theta^I, \pi^I)} \left[L(f \mid \Theta^I, \pi^I \circ \pi^I) \right] \\ &= \sum_{x \in \{x_L, x_H\}^2} \frac{1}{2} \cdot \sum_{I \in \{A, B\}} \Pr_{\pi^I \circ \pi^I}[x] \left(f(x) - f_{\pi^I \circ \pi^I}^{\text{BAYES}}(x) \right)^2 \\ &= \sum_{x \in \{x_L, x_H\}^2} \Pr_{\pi}[x] \cdot \sum_{I \in \{A, B\}} \frac{1}{2} \cdot \left(f(x) - f_{\pi^I \circ \pi^I}^{\text{BAYES}}(x) \right)^2 \\ & \hspace{15em} \text{(By Eqn. (10) where } \Pr_{\pi^{A \circ \pi^A}}[x] = \Pr_{\pi^{B \circ \pi^B}}[x]) \\ &= \sum_{x \in \{x_L, x_H\}^2} \Pr_{\pi}[x] \cdot \left(\left(f(x) - \frac{f_{\pi^{A \circ \pi^A}}^{\text{BAYES}}(x) + f_{\pi^{B \circ \pi^B}}^{\text{BAYES}}(x)}{2} \right)^2 + \frac{\left(f_{\pi^{A \circ \pi^A}}^{\text{BAYES}}(x) - f_{\pi^{B \circ \pi^B}}^{\text{BAYES}}(x) \right)^2}{4} \right) \\ &\geq \sum_{x \in \{x_L, x_H\}^2} \Pr_{\pi}[x] \cdot \frac{\left(f_{\pi^{A \circ \pi^A}}^{\text{BAYES}}(x) - f_{\pi^{B \circ \pi^B}}^{\text{BAYES}}(x) \right)^2}{4} \\ &= \frac{1 + \sqrt{5}}{8} \cdot \left(\left(\frac{15 - 5\sqrt{5}}{4} - \frac{1 + \sqrt{5}}{4} \right)^2 + \left(\frac{3 - \sqrt{5}}{4} - \frac{5\sqrt{5} - 11}{4} \right)^2 \right) \cdot \frac{1}{4} + \frac{3 - \sqrt{5}}{8} \cdot 2 \cdot \left(\frac{3 - \sqrt{5}}{4} - \frac{1 + \sqrt{5}}{4} \right)^2 \cdot \frac{1}{4} \\ &= \frac{5\sqrt{5} - 11}{8}, \end{aligned}$$

where in the inequality we have used the fact that for all $a, b, c \in \mathbb{R}$, we have

$$\frac{1}{2} \left((c - a)^2 + (c - b)^2 \right) = \left(c - \frac{a + b}{2} \right)^2 + \frac{(a - b)^2}{4} \geq \frac{(a - b)^2}{4}.$$

We thus have finished the proof. \square

We now show that there exists a generalized log-odds aggregation function that can achieve a regret upper bound of 0.022763.

Definition 5.2 (Generalized log-odds aggregation function). Let $\mu \in [0, 1]$ be the prior mean. Fix a parameter $\alpha \in [0, 1]$ and a parameter $\gamma \in [-1, 1]$, a *generalized log-odds aggregation scheme* is a function $f_{\alpha,\gamma}(\cdot, \cdot) : [0, 1]^2 \rightarrow [0, 1]$ mapping from the two experts' reports $(x_1, x_2) \in [0, 1]^2$ to a forecast as follows:

$$f_{\alpha,\gamma}(x_1, x_2) = \frac{1}{1 + \left(\frac{\mu}{1-\mu}\right)^\gamma \cdot \left(\frac{1-x_1}{x_1}\right)^\alpha \cdot \left(\frac{1-x_2}{x_2}\right)^\alpha}.$$

Equivalently, aggregator $f_{\alpha,\gamma}$ satisfies that

$$\text{LOGIT}(f_{\alpha,\gamma}(x_1, x_2)) = \alpha \cdot \text{LOGIT}(x_1) + \alpha \cdot \text{LOGIT}(x_2) - \gamma \cdot \text{LOGIT}(\mu).$$

It is important to note that our proposed generalized log-odds aggregation function does not fully leverage the knowledge of the expert's marginal prediction distributions. Instead, it only uses the inferred prior mean in the aggregation function. Nevertheless, as we show below, even with this minimal piece of information there exists a choice of parameters (α, γ) that guarantees a uniformly low regret.

Proposition 5.3. *With $\alpha = 0.656089$ and $\gamma = 0.498268$, the aggregation function $f_{\alpha,\gamma}$ guarantees worst-case regret at most 0.0228; that is,*

$$\mathbf{Reg}_{\Pi_{CT}}[f_{0.656089, 0.498268}] \approx 0.022763 < 0.0228.$$

In the known-marginal distribution setting, the aggregator has access to the marginal distribution of each expert's report. Consequently, the prior mean μ is no longer an unknown parameter to be estimated but can be directly identified. To evaluate the performance of our proposed scheme in this benchmark, we consider three baseline algorithms where the prior estimate $\hat{\mu}$ is set to the true prior mean μ . First, we consider a trivial baseline that *always outputs the prior mean*. Second, we evaluate the *Standard Bayesian* aggregator from Arieli et al. (2018) by substituting the true μ into their formula. Finally, we include the parameterized aggregator from Kong et al. (2024), where the parameter λ is optimized to minimize the worst-case regret given the known prior mean. We summarize the regret comparisons in Table 6.

Aggregator	Prior Estimate ($\hat{\mu}$)	Worst-Case Regret
Always Output Prior Mean	μ	0.2500
Standard Bayesian (Arieli et al., 2018)	μ	0.0403
Kong et al. (2024) ($\lambda = 0.8$)	μ	0.0389
Our Aggregator ($f_{\alpha,\gamma}$)	$\alpha \approx 0.656089, \gamma \approx 0.498268$	0.022763

Table 6: The regret comparison in the known-marginal setting.

Proof of Proposition 5.3. The proof follows a similar structure to that of Theorem 3.2. Invoking Lemma 3.3, it suffices to consider information structures where each expert receives at most two signals, i.e., $S_1 = S_2 = \{0, 1\}$.

We parameterize the conditional signal probabilities as $a_1 = \pi_1(s_1 | \theta_1)$, $b_1 = \pi_2(s_1 | \theta_1)$, $a_2 = \pi_1(s_1 | \theta_2)$, and $b_2 = \pi_2(s_1 | \theta_2)$. Accordingly, the problem of identifying the worst-case regret reduces to a finite-dimensional optimization over the state parameters θ_1, θ_2 , the prior λ , and the signal probabilities:

$$\mathbf{Reg}_{\Pi_{\text{CI}}}[f_{\alpha, \gamma}] = \sup_{\Theta \in [0,1]^2: \theta_1 \leq \theta_2} \sup_{\lambda_2 \in [0,1]; a_1, b_1, a_2, b_2 \in [0,1]^4} L(f_{\alpha, \gamma} | \Theta, \pi) .$$

By solving the optimization problem for the known-marginal setting, we identify the optimal parameters as $\alpha \approx 0.656089$ and $\gamma \approx 0.498268$. Numerical calculations reveal that the global maximum of the regret function is attained at the configuration where $\theta_1 = 0$, $\theta_2 = 0.79570$, and the prior mean $\mu = 0.49908$. Specifically, the experts' signal structures correspond to the following reports: expert 1 provides $x_1^\dagger = 0.18165$ and $x_1^\ddagger = 0.79570$; expert 2 provides $x_2^\dagger = 0.18165$ and $x_2^\ddagger = 0.79570$. The resulting maximum regret is equal to 0.022763. \square

5.2 General Information Structures

We next discuss the regret for the general information structures.

Proposition 5.4. *For the general information structures, we have that $\inf_f \mathbf{Reg}_{\Pi}[f] = 0.25$. There exists a trivial aggregation function $f \equiv 0.5$ that achieves $\mathbf{Reg}_{\Pi}[f] = 0.25$.*

Proof. The upper bound is straightforward as one can use the trivial aggregation function $f_{\text{triv}}(x) = 1/2$, which can guarantee a regret of 0.25.

We next show the lower bound. We use the same hard information structure proposed in Arieli et al. (2018). This instance demonstrates that even if the marginal distribution of each expert's forecast is known to the aggregator, the worst-case regret still remains at least $1/4$.

Consider a binary state space $\Theta = \{0, 1\}$ with a uniform prior $\lambda(0) = \lambda(1) = 1/2$. Experts receive binary signals from $S = \{s_L, s_H\}$. We construct this information structure as Table 7.

State (θ)	Prior (λ)	Joint Signal Probabilities $\Pr[s_1, s_2 \theta]$			
		(s_L, s_L)	(s_H, s_H)	(s_L, s_H)	(s_H, s_L)
0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
1	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$

Table 7: The joint information structure.

Under the constructed information structure, each of the four possible signal profiles $(s_1, s_2) \in \{s_L, s_H\}^2$ has an equal unconditional probability of $1/4$. Moreover, it is easy to see that $x_i(s_L) = x_i(s_H) = 1/2$ all $i \in [2]$. In other words, each expert's forecast marginal distribution is a point mass on $1/2$. However, the omniscient forecasts can always infer the true state. Thus, the aggregator has a relative regret at least $1/4$. \square

References

- János Aczél and Carl Wagner. 1980. A characterization of weighted arithmetic means. *SIAM Journal on Algebraic Discrete Methods* 1, 3 (Sept. 1980), 259–260.
- Rui Ai, Yuqi Pan, David Simchi-Levi, Milind Tambe, and Haifeng Xu. 2025. Beyond majority voting: LLM aggregation by leveraging higher-order information. arXiv preprint arXiv:2510.01499.
- Itai Arieli, Yakov Babichenko, and Rann Smorodinsky. 2018. Robust Forecast Aggregation. *Proceedings of the National Academy of Sciences* 115, 52 (Dec. 2018), E12135–E12143.
- Itai Arieli, Yakov Babichenko, Inbal Talgam-Cohen, and Konstantin Zabarnyi. 2025. A Random Dictator Is All You Need. *American Economic Journal: Microeconomics* 17, 1 (Feb. 2025), 66–96. doi:10.1257/mic.20230255
- Henrique De Oliveira, Yuhta Ishii, and Xiao Lin. 2021. Robust Merging of Information. In *Proceedings of the 22nd ACM Conference on Economics and Computation (EC '21)*. ACM, New York, NY, USA, 341–342.
- Franz Dietrich and Christian List. 2016. Probabilistic opinion pooling. In *The Oxford Handbook of Probability and Philosophy*, Alan Hájek and Christopher Hitchcock (Eds.). Oxford University Press, Oxford, UK, 519–542.
- Philip Ernst, Robin Pemantle, Ville Satopää, and Lyle Ungar. 2016. Bayesian aggregation of two forecasts in the partial information framework. *Statistics & Probability Letters* 119 (Dec. 2016), 170–180.
- Rafael Frongillo, Yiling Chen, and Ian Kash. 2015. Elicitation for aggregation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI '15, Vol. 29)*. AAAI Press, Austin, Texas.
- Rafael Frongillo, Mary Monroe, Eric Neyman, and Bo Waggoner. 2025. Robust forecast aggregation via additional queries. arXiv preprint arXiv:2512.05271.
- Christian Genest. 1984. A characterization theorem for externally Bayesian groups. *The Annals of Statistics* 12, 3 (Sept. 1984), 1100–1105.
- Yongkang Guo, Jason D. Hartline, Zhihuan Huang, Yuqing Kong, Anant Shah, and Fang-Yi Yu. 2025. Algorithmic Robust Forecast Aggregation. In *Proceedings of the 26th ACM Conference on Economics and Computation (EC '25)*. ACM, New York, NY, USA.
- Yongkang Guo and Yuqing Kong. 2025. Robust Aggregation with Adversarial Experts. In *Proceedings of the ACM Web Conference 2025 (WWW '25)*. ACM.
- Yuqing Kong, Shu Wang, and Ying Wang. 2024. The Surprising Benefits of Base Rate Neglect in Robust Aggregation. In *Proceedings of the 25th ACM Conference on Economics and Computation (EC '24)*. ACM, New York, NY, USA, 1131–1131.
- Gilat Levy and Ronny Razin. 2021. A maximum likelihood approach to combining forecasts. *Theoretical Economics* 16, 1 (Jan. 2021), 49–83.
- Gilat Levy and Ronny Razin. 2022. Combining forecasts in the presence of ambiguity over correlation structures. *Journal of Economic Theory* 199 (Jan. 2022), 105075.

- Eric Neyman and Tim Roughgarden. 2022. Are you smarter than a random expert? The robust aggregation of substitutable signals. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC '22)*. ACM, New York, NY, USA, 990–1012.
- Asa B. Palley and Jack B. Soll. 2019. Extracting the wisdom of crowds when information is shared. *Management Science* 65, 5 (May 2019), 2291–2309.
- Deng Pan, Zehong Chen, and Yuqing Kong. 2024. Robust Decision Aggregation with Second-order Information. In *Proceedings of the ACM Web Conference 2024 (WWW '24)*. ACM, New York, NY, USA, 390–400.
- Drazen Prelec. 2004. A Bayesian truth serum for subjective data. *Science* 306, 5695 (Oct. 2004), 462–466. [doi:10.1126/science.1102081](https://doi.org/10.1126/science.1102081)
- Dražen Prelec, H. Sebastian Seung, and John McCoy. 2017. A solution to the single-question crowd wisdom problem. *Nature* 541, 7638 (Jan. 2017), 532–535.
- Ville A. Satopää, Jonathan Baron, Dean P. Foster, Barbara A. Mellers, Philip E. Tetlock, and Lyle H. Ungar. 2014. Combining multiple probability predictions using a simple logit model. *International Journal of Forecasting* 30, 2 (April 2014), 344–356.
- Ville A. Satopää, Robin Pemantle, and Lyle H. Ungar. 2016. Modeling probability forecasts via information diversity. *J. Amer. Statist. Assoc.* 111, 516 (Dec. 2016), 1623–1633.
- Juntao Wang, Yang Liu, and Yiling Chen. 2021. Forecast aggregation via peer prediction. In *Proceedings of the 9th AAAI Conference on Human Computation and Crowdsourcing (HCOMP '21, Vol. 9)*. AAAI Press, 131–142.
- Tom Wilkening, Marcellin Martinie, and Piers D. L. Howe. 2022. Hidden experts in the crowd: Using meta-predictions to leverage expertise in single-question prediction problems. *Management Science* 68, 1 (Jan. 2022), 487–508. [doi:10.1287/mnsc.2020.3934](https://doi.org/10.1287/mnsc.2020.3934)

A Missing Proof in Section 2

Lemma 2.1 (Regret simplification [Arieli et al., 2018](#); [Kong et al., 2024](#); [Guo et al., 2025](#)). *Given any state space $\Theta \subset [0, 1]^2$ and any information structure π , we have that*

$$L(f \mid \Theta, \pi) = \mathbb{E}_{(\theta, s) \sim \pi, \omega \sim \text{BERN}(\theta)} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right)^2 \right] .$$

Proof. Recall that the omniscient aggregator reports the true posterior mean given the signal profile s . By the law of iterated expectations and the fact that $\mathbb{E}[\omega \mid \theta] = \theta$ due to $\omega \sim \text{BERN}(\theta)$, we have:

$$f^{\text{BAYES}}(x(s)) = \Pr[\omega = 1 \mid s] = \mathbb{E}[\omega \mid s] = \mathbb{E}[\mathbb{E}[\omega \mid \theta] \mid s] = \mathbb{E}[\theta \mid s] . \quad (11)$$

Thus, the omniscient forecast $f^{\text{BAYES}}(x(s))$ is the conditional expectation of both the outcome ω and the state θ .

Fix any state space, any information structure, and any aggregation function f , by definition, we have

$$\begin{aligned} L(f \mid \Theta, \pi) &= \mathbb{E} \left[(f(x(s)) - \omega)^2 \right] - \mathbb{E} \left[\left(f^{\text{BAYES}}(x(s)) - \omega \right)^2 \right] \\ &= \mathbb{E} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right)^2 \right] + 2\mathbb{E} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right) \cdot \left(f^{\text{BAYES}}(x(s)) - \omega \right) \right] . \end{aligned}$$

We now show that the cross-term equals to 0. Conditioning on the signals s :

$$\begin{aligned} &\mathbb{E} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right) \cdot \left(f^{\text{BAYES}}(x(s)) - \omega \right) \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right) \cdot \left(f^{\text{BAYES}}(x(s)) - \omega \right) \mid s \right] \right] \\ &= \mathbb{E} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right) \cdot \mathbb{E} \left[\left(f^{\text{BAYES}}(x(s)) - \omega \right) \mid s \right] \right] \\ &= \mathbb{E} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right) \cdot \left(f^{\text{BAYES}}(x(s)) - \mathbb{E}[\omega \mid s] \right) \right] \\ &= 0 . \end{aligned} \quad (\text{Due to Eqn. (11).})$$

Consequently, we have $L(f \mid \Theta, \pi) = \mathbb{E} \left[\left(f(x(s)) - f^{\text{BAYES}}(x(s)) \right)^2 \right]$. □