

# Sequential Binary Hypothesis Testing with Competing Agents under Information Asymmetry

Aneesh Raghavan, M. Umar B. Niazi, and Karl H. Johansson

**Abstract**—This paper concerns sequential hypothesis testing in competitive multi-agent systems, where agents exchange potentially corrupted information with each other. Specifically, a two-agent scenario is studied where each agent aims to correctly infer the true state of nature while optimizing decision speed and accuracy. At each iteration, agents collect private observations, update their beliefs, and share potentially corrupted belief signals with their counterparts before deciding whether to stop and declare their current state estimate or continue to the next iteration based on their confidence level. The analysis yields three main results: (1) when agents share information strategically, the optimal signaling policy involves equal-probability randomization between truthful and inverted beliefs; (2) agents maximize performance by relying solely on their own observations for belief updating while using received information only to anticipate their counterpart’s stopping decision; and (3) the agent reaching their confidence threshold first cause the other agent to achieve a higher conditional probability of error. Numerical simulations further demonstrate that agents with higher KL divergence in their conditional distributions gain a competitive advantage. Furthermore, our results establish that information sharing, despite strategic signaling, reduces overall system stopping time compared to non-interactive scenarios, which highlights the inherent value of communication even in this competitive setup.

## I. INTRODUCTION

Decision-makers in multi-agent systems face uncertainty regarding both the environment and other agents’ behavior. In this paper, we consider a scenario where agents interact with each other to improve their decisions while balancing two conflicting factors: the potential for complementary information received from their neighboring agents, and the risk that the information received might be misleading. In particular, we study a sequential binary hypothesis testing problem where two agents interact with each other to competitively infer the true state of nature. At each iteration, both agents collect local observations that are statistically related to the true state, update their beliefs derived from these observations, and share (possibly corrupted) beliefs with their counterparts. In this sequential framework, agents must make decisions over time based on the accumulated information, i.e., their own sets of observations and belief signals received from their counterparts. At every iteration,

This work was partly supported by the Swedish Research Council’s Distinguished Professor grant 2017-01078, Knut and Alice Wallenberg Foundation’s Wallenberg Scholar grant, the Swedish Strategic Research Foundation’s FuSS SUCCESS grant, and the European Union’s Horizon Research and Innovation Programme under Marie Skłodowska-Curie grant agreement No. 101062523.

The authors are with the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm SE-100 44, Sweden. Emails: aneesh@kth.se, mubniazi@kth.se, kallej@kth.se

both agents update their confidence levels and choose from two possible actions: stop and declare the state they believe to be true, or continue gathering more information.

Despite the information exchange between agents, the considered framework is competitive as each agent independently aims to infer the true state both quickly and accurately. The information asymmetry arises because agents have access to only their local observations and knowledge of their respective marginal distributions.

Such sequential multi-agent decision-making problems are ubiquitous in both economic and engineering applications. Some examples include financial markets [1], smart grid [2], and federated learning [3]. These applications share a common structure: agents with access to only local information make sequential decisions while incorporating potentially misleading signals from others. The fundamental challenge lies in designing robust decision rules when received information is not reliable, yet contains valuable insights that could improve decision quality.

## A. Related Literature

Decentralized hypothesis testing is a classical problem in distributed control. Tenney and Sandell [4] established the collaborative framework for decentralized detection where sensors make local decisions based on independent observations, showing that optimal strategies typically involve likelihood ratio tests with coupled thresholds. Teneketzis and Varaiya [5] extended these concepts to sequential problems, demonstrating how threshold-based stopping rules operate when agents must decide whether to continue gathering information. Using a game-theoretic framework for multi-agent sequential hypothesis testing, [6] establishes that belief states serve as sufficient statistics for conditionally independent observations. However, these works assume that agents share true beliefs with their counterparts.

Strategic interactions in hypothesis testing under adversarial frameworks have also been studied recently. For instance, [7] analyzes the tradeoff between security and efficiency in binary hypothesis testing with Byzantine sensors, while [8] and [9] formulate a similar problem as a zero-sum partial information game. A related line of research addresses how agents can collectively learn despite limited information sharing. In this regard, a distributed hypothesis testing approach using a “min-rule” rather than belief-averaging has been proposed by [10]. Similarly, [11] examines hypothesis testing as a signaling game with subjective priors or misaligned objectives. Another particularly relevant work is [12], which analyzes distributed hypothesis testing where network nodes

communicate beliefs to neighbors and shows that beliefs on wrong hypotheses converge to zero exponentially fast. While these distributed approaches often focus on consensus-building protocols, our work specifically addresses the strategic tension between information sharing and withholding by modeling the possibility of deceptive signaling between agents and determining optimal response strategies.

### B. Our Contributions

In this paper, we examine competitive sequential binary hypothesis testing in which agents may manipulate the information they share with their counterparts. The reason for deliberately manipulating the information is to delay the other agent in inferring the true state. Our work differs from existing literature in two key aspects: first, unlike adversarial models, we focus on strategic behavior where agents act in self-interest without malicious intent; second, unlike consensus-building protocols or collaborative models, we analyze the strategic nature of information exchange and develop optimal response strategies within a competitive decision-making framework.

Our results are summarized as follows. First, we show that when agents strategically share information, the optimal signaling policy involves randomizing between truthful and inverted beliefs (Proposition 1), which results from effectively maximizing uncertainty while preserving information content. Second, we prove that it is optimal for agents not to incorporate information received from others in their local beliefs, relying solely on their own observations until the stopping time is detected (Proposition 2). However, information received from others helps the agents detect when their counterparts will stop. Finally, we establish that this dynamic creates an advantage for the agent achieving the confidence threshold before the other (Proposition 3). That is, the agent who reaches their confidence threshold first achieves a bounded conditional probability of error, while the other agent faces potentially less accurate inference. Our information-theoretic interpretations through entropy offer an elegant explanation for these somewhat counterintuitive results. While the distributed hypothesis testing literature primarily focuses on fusion of likelihood ratios [12], [10], belief fusion policy and stopping policy account for the signaling policy which is strategic and not adversarial.

## II. PROBLEM FORMULATION

This section describes the model of sequential multi-agent binary hypothesis testing and formulates the problem.

### A. Setup

Suppose there are two possible states of nature,  $\theta^0$  and  $\theta^1$ , that are mutually exclusive. Consider two agents, A and B, each collecting its local set of observations, which are statistically related to the true state of nature. Each agent has access to the joint distribution of their respective observations and the state. Given the observations, the objective of the agents is to find the true state  $\theta$  competitively. This setup is illustrated in Fig. 1 and formalized below:

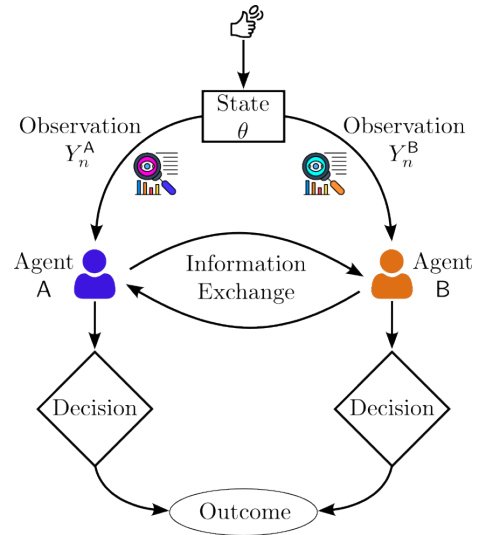


Fig. 1. The two-agent binary hypothesis testing setup.

- Both agents act simultaneously. Time is considered discrete, and each iteration is denoted by a subscript  $n \in \mathbb{N}$ . At every iteration  $n$ , agents perform multiple steps (e.g., interact with each other and make decisions).
- The true state of nature  $\theta \in \Theta \triangleq \{\theta^0, \theta^1\}$  remains fixed.
- The observations collected by agent  $i \in \{A, B\}$  are denoted by  $Y_n^i \in S^i$ , where  $S^i$  is a finite set.
- The state of nature  $\theta$  and the observations  $\{Y_n^A, Y_n^B\}_{n \in \mathbb{N}}$  (and functions of these observations) are considered to be random variables on an abstract probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . That is,  $\theta : \Omega \rightarrow \Theta$  and  $Y_n^i : \Omega \rightarrow S^i$  are measurable functions, for  $i \in \{A, B\}$ .
- None of the agents have access to the joint distribution  $\mathbb{P}(\theta = \theta^k \cap \{Y_t^A = y_t^A\}_{t=1}^n \cap \{Y_t^B = y_t^B\}_{t=1}^n)$ , where  $k \in \{1, 2\}$ . However, each agent  $i$  knows only its marginal distribution  $\mathbb{P}(\theta = \theta^k \cap \{Y_t^i = y_t^i\}_{t=1}^n)$ . Moreover, both agents know a prior  $\mathbb{P}(\theta = \theta^k)$  and their respective conditional probabilities  $\mathbb{P}(\{Y_t^i = y_t^i\}_{t=1}^n \mid \theta = \theta^k)$ .

### B. Belief Signaling and Decision Policies

During iteration  $n \in \mathbb{N}$ , as illustrated in Fig. 2, each agent  $i \in \{A, B\}$  executes the following steps in sequence:

- 1) Update the local belief  $\pi_n^i \in \Delta(\Theta)$  based on the sample  $Y_n^i$  collected. We consider a Bayesian update:

$$\pi_n^i(\theta^k) = \mathbb{P}(\theta = \theta^k \mid \{Y_t^i = y_t^i\}_{t=1}^n).$$

- 2) Reveal a (possibly corrupted) belief  $\tilde{\pi}_n^i$  to the other agent  $\sim i$ , where<sup>1</sup>  $\tilde{\pi}_n^i = \nu^i(\pi_n^i)$  with  $\nu^i : \Delta(\Theta) \rightarrow L^1(\Omega, \mathcal{F}, \mathbb{P})$  the *belief signaling policy*.
- 3) Update the belief to  $\bar{\pi}_n^i \in \Delta(\Theta)$  by incorporating  $\tilde{\pi}_n^{\sim i}$  in the local belief via *belief fusion policy*.
- 4) Decide whether to stop and declare  $D_n^i = \mu_n^i(\bar{\pi}_n^i)$  to be the most likely state of nature according to belief  $\bar{\pi}_n^i$  or continue with collecting the next observation  $Y_{n+1}^i$ . Here,  $\mu_n^i : \Delta(\Theta) \rightarrow \Theta$  is the *decision policy* of  $i$ .

<sup>1</sup>We use the notation  $\sim i$  to denote the agent who is not  $i$ . Similarly,  $\theta^{\sim k}$  means the state that is not  $\theta^k$ .

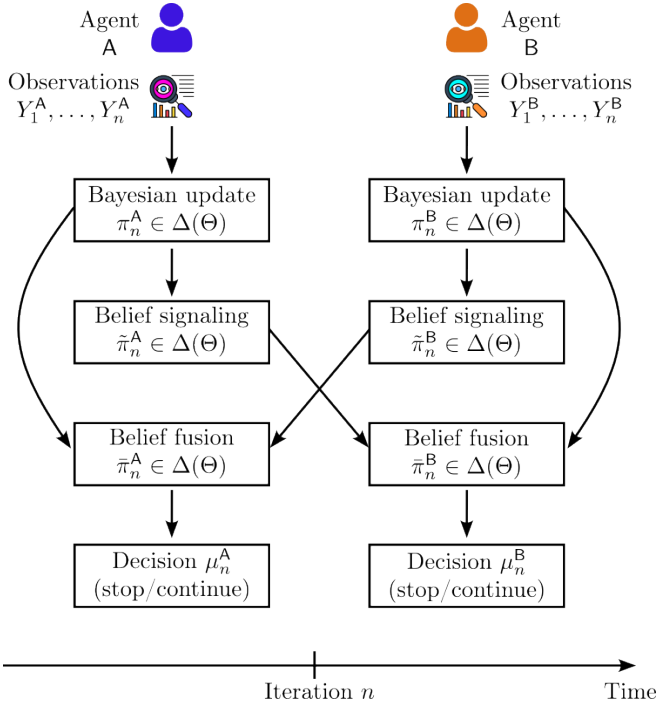


Fig. 2. Steps executed by agents at each iteration of sequential binary hypothesis testing.

*Belief signaling policy.* Let the belief signaling policy be:

$$\nu^i(\pi) = \begin{cases} \pi, & \text{with probability } \alpha^i \\ 1 - \pi, & \text{with probability } 1 - \alpha^i \end{cases} \quad (1)$$

where  $0 \leq \alpha^i \leq 1$  is a design parameter corresponding to how much  $i$  should be truthful. In particular, agent  $i$  flips a biased coin to decide what signal should be sent. If the outcome is head, the agent sends their true belief ( $\tilde{\pi}_n^i = \pi_n^i$ ), else the agent sends the inverted belief ( $\tilde{\pi}_n^i = 1 - \pi_n^i$ ) to agent  $\sim i$ . The outcome of the coin toss for agent  $i$  is denoted by  $Z_n^i$ , where  $Z_n^i : \Omega \rightarrow \{0, 1\}$  is a measurable function. Thus,  $\{Z_n^i\}$  is a Bernoulli process and for agent  $i$ , the process  $\{Z_n^i\}$  is assumed to be independent of the process  $\{Y_n^i\}$ . The belief distribution of  $\theta$  due to  $\tilde{\pi}_n^i$  is denoted by

$$\left[ \frac{\mathbb{P}(\theta = \theta^1 | \{Y_t^i = y_t^i\}_{t=1}^n)}{\mathbb{P}(\theta = \theta^0 | \{Y_t^i = y_t^i\}_{t=1}^n)} \right]_{i \rightarrow \sim i}$$

*Belief fusion policy.* For each agent  $i$ , we consider the belief fusion policy to be a convex combination of its own belief and the belief revealed by the other agent:

$$\begin{aligned} \bar{\pi}_n^i &= w_n^i \pi_n^i \\ &+ (1 - w_n^i) \left[ \frac{\mathbb{P}(\theta = \theta^1 | \{Y_t^{\sim i} = y_t^{\sim i}\}_{t=1}^n)}{\mathbb{P}(\theta = \theta^0 | \{Y_t^{\sim i} = y_t^{\sim i}\}_{t=1}^n)} \right]_{\sim i \rightarrow i} \end{aligned} \quad (2)$$

where  $0 \leq w_n^i \leq 1$  is a design parameter corresponding to the weight  $i$  puts on its own observations as compared to the belief revealed by the other agent. Note that the aim of the belief fusion policy (2) is to efficiently incorporate the belief revealed by the other agent to infer the true state of nature.

*Decision policy.* Following the literature on stopping time problems [13], [14], we consider the decision policy  $\mu_n^i(\bar{\pi}_n^i)$

to be a threshold policy:

$$\mu_n^i(\bar{\pi}_n^i) = \begin{cases} \text{stop, } D_n^i = \theta^1; & \text{if } \bar{\pi}_n^i(\theta^0) \leq T_{\theta^0}^i \\ \text{stop, } D_n^i = \theta^0; & \text{if } \bar{\pi}_n^i(\theta^1) \leq T_{\theta^1}^i \\ \text{continue, } D_n^i = \emptyset; & \text{otherwise} \end{cases} \quad (3)$$

where  $T_{\theta^0}^i, T_{\theta^1}^i \in [0, 1]$  are threshold parameters to be designed. The decision policy (3) aims to infer the true state of nature quickly and achieve the desired level of accuracy.

### C. Problem Definition

Following stochastic control formulations of the hypothesis testing problem in [4], [5], [15], we consider the cost function for agent  $i$  to be

$$C^i(D_n^i) = \mathbb{E}_{\mathbb{P}} \left[ nc^i \mathbf{1}_{D_n^i = \emptyset} + \hat{c}^i \mathbf{1}_{D_n^i \neq \emptyset, D_n^i \neq \theta} \mid \{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n \right]$$

where  $c^i > 0$  is the cost associated with continuing and collecting the next sample  $Y_{n+1}^i$ , and  $\hat{c}^i > 0$  is the cost associated with stopping when the declared state  $D_n^i$  is wrong (i.e.,  $D_n^i \neq \emptyset$  and  $D_n^i \neq \theta$ ). The optimization problem for each agent at iteration  $n$  is thus defined as:

$$\min_{w_n^i, T_{\theta^0}^i, T_{\theta^1}^i} C^i(D_n^i) \quad (4)$$

where  $D_n^i := D_n^i(w_n^i, T_{\theta^0}^i, T_{\theta^1}^i)$ .

Since the joint distribution between  $\theta$ ,  $\{Y_t^A\}_{t=1}^n$ , and  $\{Y_t^B\}_{t=1}^n$  is unknown to agent  $i$ , the joint distribution between  $\theta$  and  $\{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n$  is also unknown to the agent. Therefore, finding the conditional probability  $\mathbb{E}_{\mathbb{P}}[\mathbf{1}_{\theta=\theta^1} \mid \{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n]$  using Bayes theorem is not possible. In order to analyze the optimization problem (4), define

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^i} \left[ \mathbf{1}_{\theta=\theta^1} \mid \{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n \right] &\triangleq \bar{\pi}_n^i(\theta^1), \text{ and} \\ \mathbb{E}_{\mathbb{P}^i} \left[ \mathbf{1}_{\theta=\theta^0} \mid \{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n \right] &\triangleq \bar{\pi}_n^i(\theta^0) = 1 - \bar{\pi}_n^i(\theta^1). \end{aligned}$$

Then, the cost function of (4) can be simplified as

$$C^i(D_n^i) = nc^i \mathbf{1}_{D_n^i = \emptyset} + \hat{c}^i \left[ \mathbf{1}_{D_n^i = \theta^1} \bar{\pi}_n^i(\theta^0) + \mathbf{1}_{D_n^i = \theta^0} \bar{\pi}_n^i(\theta^1) \right].$$

Note that since

$$\begin{aligned} \mathbf{1}_{D_n^i = \theta^1} \bar{\pi}_n^i(\theta^0) + \mathbf{1}_{D_n^i = \theta^0} \bar{\pi}_n^i(\theta^1) \\ = \mathbb{P}(\mu_n^i(\bar{\pi}_n^i) \neq \theta \mid \{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n) \end{aligned} \quad (5)$$

it follows that

$$C^i(D_n^i) = nc^i \mathbf{1}_{D_n^i = \emptyset} + \hat{c}^i \mathbb{P}(\mu_n^i(\bar{\pi}_n^i) \neq \theta \mid \{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n).$$

Let  $\tau^i$  denote the random stopping time of  $i$ , then

$$C^i(D_{\tau^i}^i) = c^i \tau^i + \hat{c}^i \mathbb{P}(\mu_{\tau^i}^i(\bar{\pi}_{\tau^i}^i) \neq \theta \mid \{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n).$$

Thus, the cost is a tradeoff between the number of samples  $\tau^i$  and the conditional probability of error  $\mathbb{P}(\mu_{\tau^i}^i(\bar{\pi}_{\tau^i}^i) \neq \theta \mid \{Y_t^i, \tilde{\pi}_t^{\sim i}\}_{t=1}^n)$ .

Different values of  $c^i$  and  $\hat{c}^i$ , more specifically the ratio  $\hat{c}^i/c^i$ , lead to different probabilities of error for a given

stopping time, and vice versa. When the ratio is high, then the emphasis is on the probability of error, and the agent collects samples until the probability of error is sufficiently small. When the ratio is very low, the emphasis is on the stopping time, and the agent collects fewer samples at the expense of a high probability of error. Thus, problem (4) can be viewed in terms of a Lagrangian formulation:

$$\min_{\{w_n^i\}_{n=1}^{\tau^i}, T_{\theta^0}^i, T_{\theta^1}^i} \tau^i \quad (6a)$$

$$\text{subject to } \mathbb{P}(\mu_{\tau^i}^i(\bar{\pi}_{\tau^i}^i) \neq \theta \mid \{Y_t^i\}_{t=1}^{\tau^i}) \leq \beta^i \quad (6b)$$

where different confidence thresholds  $\beta^i$ 's correspond to different stopping thresholds  $T_{\theta^0}^i, T_{\theta^1}^i$ .

*Signaling problem.* In the competitive setup, the agent who wins is the one achieving the confidence threshold before the other agent. To ensure fairness among the agents, we impose the condition that  $\beta^1 = \beta^2 = \beta$ . In other words, the agent who stops first while ensuring that the probability of error conditioned on its observations is less than or equal to  $\beta$  wins the competition. For any belief fusion policy and thresholds of agent  $\sim i$ , agent  $i$  aims to prevent  $\sim i$  from winning the competition by optimizing over the belief signaling policy. Since the belief signaling policy (1) is parameterized by  $\alpha^i$ , the objective of agent  $i$  is

$$\max_{\alpha^i} \mathbb{P}(\mu_n^{\sim i}(\bar{\pi}_n^{\sim i}) \neq \theta \mid \{Y_t^{\sim i}, \tilde{\pi}_t^{\sim i}\}_{t=1}^n). \quad (7)$$

That is, agent  $i$  chooses the truthfulness probability  $\alpha^i$  such that it maximizes  $\sim i$ 's conditional probability of error. Although solving (7) is not possible for agent  $i$  because it doesn't know the parameters of  $\sim i$ 's belief fusion and decision policies, one can indirectly characterize the optimal value of  $\alpha^i$ . Note that the aim of the signaling policy according to (7) is to not reveal a lot of information so that the other agent gains sufficient information to correctly and rapidly infer the true state, but also to not corrupt it too much so that the signal becomes predictable enough and can be simply inverted.

*Competitive stopping problem.* By incorporating the signaling problem (7) into the stopping problem (6), we obtain a competitive stopping problem as

$$\min_{\{w_n^i\}_{n=1}^{\tau^i}, T_{\theta^0}^i, T_{\theta^1}^i} \tau^i \quad (8a)$$

$$\text{subject to } \max_{\alpha^{\sim i}} \mathbb{P}(\mu_{\tau^i}^i(\bar{\pi}_{\tau^i}^i) \neq \theta \mid \{Y_t^i\}_{t=1}^{\tau^i}) \leq \beta^i. \quad (8b)$$

That is, agent  $i$  minimizes its stopping time subject to achieving a conditional probability of error irrespective of signals it receives from  $\sim i$ . Note that the belief fusion policy is parameterized by  $w_n^i$  and the decision policy is parameterized by  $T_{\theta^0}^i$  and  $T_{\theta^1}^i$ . Thus, the problem boils down to designing parameters  $w^i, T_{\theta^0}^i$ , and  $T_{\theta^1}^i$  of the belief fusion policy (2) and the decision policy (3) such that the above objective is achieved.

### III. ANALYSIS

This section presents our key results on sequential binary hypothesis testing with competing agents. We characterize

the optimal signaling and belief fusion policies. We also derive the optimal stopping thresholds and the conditional probability of errors of the agents.

#### A. Optimal Signaling

We explore how agents should optimally manipulate their beliefs when signaling. Considering the belief signaling policy (1), where each agent has a "truthfulness" parameter  $\alpha^i$ , we observe that setting this truthfulness parameter to exactly 0.5 creates an optimal signaling strategy. This means the agent should randomly choose between sending their true belief or its opposite with equal probability, regardless of the other agent's stopping thresholds or belief fusion parameters.

**Proposition 1.** Consider the belief signaling policy (1) parametrized by  $0 \leq \alpha^i \leq 1$ . Then, for any belief fusion parameter sequence,  $\{w_n^{\sim i}\}$  and thresholds  $T_U^{\sim i}, T_L^{\sim i}$  of agent  $\sim i$ ,  $\alpha^i = 0.5$  is optimal with respect to (7).

*Proof.* See [16] for the proof.  $\square$

*Entropy-based interpretation.* The intuition behind Proposition 1 connects to information entropy. By choosing a 50/50 split between truth and inversion, the agent maximizes uncertainty in their signals while still conveying useful information. This approach effectively hides the agent's true decision while ensuring the communicated belief maintains the same level of uncertainty as the original belief. Since the entropy of a belief is identical to its inverse, this strategy preserves information content while maintaining strategic ambiguity about the agent's actual position.

To elucidate, recall the entropy of a random variable with a Bernoulli distribution  $p$ :

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p).$$

The entropy function possesses two geometric properties: (i) symmetry about  $p = 0.5$  and (ii) concavity in  $p$ . Since the entropy of  $\pi_n^i$  or  $1 - \pi_n^i$  is the same, an agent can communicate to the other agent by transmitting either of them that it has achieved the desired level of uncertainty in the decision while keeping the true decision hidden. For any two independent random variables  $Y$  and  $Z$ , it holds that  $H(Y, Z) = H(Y) + H(Z \mid Y) = H(Y) + H(Z)$ . When an agent is choosing  $\alpha^i$  such that the entropy of the distribution of the random vector  $\tilde{\pi}_n^i$  is maximized, it follows from the previous relation that the agent has to maximize  $H(\alpha)$  because  $H(\tilde{\pi}_n^i) = H(\alpha) + H(\pi_n^i)$ . Since the entropy of the Bernoulli distribution gets maximized at 0.5, it is optimal for agent  $i$  to choose  $\alpha^i = 0.5$ .  $\diamond$

*Remark 1.* The above result states that by choosing  $\alpha^A = 0.5$ , we get

$$\left[ \begin{array}{c} \mathbb{P}(\theta = \theta^1 \mid \{Y_t^A = y_t^A\}_{t=1}^n) \\ \mathbb{P}(\theta = \theta^0 \mid \{Y_t^A = y_t^A\}_{t=1}^n) \end{array} \right]_{A \rightarrow B} = \left[ \begin{array}{c} 0.5 \\ 0.5 \end{array} \right]$$

which implies that

$$\bar{\pi}_n^B = w_n^B \pi_n^B + (1 - w_n^B) \left[ \begin{array}{c} 0.5 \\ 0.5 \end{array} \right] \quad (9)$$

Hence, for any  $0 \leq w_n^B < 1$ , the probability of error based on decision  $\bar{\pi}_n^B$  is greater than or equal to the probability of error based on decision  $\pi_n^B$ .  $\square$

### B. Optimal Belief Fusion and Stopping Time

We analyze how agents should optimally combine their own beliefs with information received from others, and when they should stop the process to make a final decision about the true state of nature. For the analysis, we consider belief fusion policy (2) where each agent uses a weighted combination of their local observations and the beliefs shared by other agents, alongside a threshold-based policy (3) to determine when to stop and declare the state.

In our multi-agent setup, the agent achieving the desired conditional probability of error at iteration  $n$ ,

$$\mathbb{P}(\mu_n^i(\bar{\pi}_n^i) \neq \theta \mid \{Y_t^i\}_{t=1}^n) \leq \beta^i$$

will *initiate* the stopping, i.e.,  $\tau^i = n$ . However, we show that both agents will ultimately stop the process at the same time, regardless of their local observation processes. Expanding on Remark 1, we further show that the optimal strategy for agents is to rely exclusively on their own observations when making decisions until the stopping time. This means agents should essentially ignore the information shared by their counterparts until they detect that one of them has achieved the desired conditional probability of error.

**Proposition 2.** For almost all sample paths, both agents stop at the same time, i.e.,  $\tau^A(\omega) = \tau^B(\omega) = \tau(\omega)$ . Moreover, for all agents  $i \in \{A, B\}$ , the optimal fusion parameter sequence with respect to (8) is given by  $w_n^i = 1$ , for  $n \leq \tau$ . For agent  $i$  initiating the stopping, the decision policy at  $\tau$  is given by

$$\mu_\tau^i(\bar{\pi}_\tau^i) = \begin{cases} D_\tau^i = \theta^1, & \text{if } \bar{\pi}_\tau^i(\theta^0) \leq \beta \\ D_\tau^i = \theta^0, & \text{if } \bar{\pi}_\tau^i(\theta^1) \leq \beta. \end{cases} \quad (10)$$

That is,  $T_{\theta^0}^i = T_{\theta^1}^i = \beta$  is optimal for  $i$  with respect to (8). The policy of the other agent  $\sim i$  at  $\tau$  is

$$\mu_\tau^{\sim i}(\bar{\pi}_\tau^{\sim i}) = \begin{cases} D_\tau^{\sim i} = \theta^1 & \text{if } \bar{\pi}_\tau^{\sim i}(\theta^0) \leq \bar{\pi}_\tau^{\sim i}(\theta^1) \\ D_\tau^{\sim i} = \theta^0 & \text{if } \bar{\pi}_\tau^{\sim i}(\theta^1) \leq \bar{\pi}_\tau^{\sim i}(\theta^0). \end{cases} \quad (11)$$

*Proof.* See [16] for the proof.  $\square$

The above proposition shows that even though  $i$  may initiate the stopping,  $\sim i$  can detect it through  $i$ 's revealed belief  $\bar{\pi}_n^i$  and will also stop at the same time. However, we show in the next section that  $\sim i$ , the agent who did not initiate the stopping, may lose the competition in the sense that it incurs a higher probability of error. Another observation in the proposition is that even when agents receive information from their counterparts, none of the agents find it optimal to incorporate it into their local beliefs. However, they leverage this shared information to anticipate when a counterpart will stop, and only then do they incorporate it into their beliefs. This counterintuitive result stems from the information-theoretic properties of belief fusion. When one agent shares manipulated information to maximize uncertainty (as established in the previous section),

the receiving agent's best response is to deliberately disregard this ambiguous signal entirely.

*Entropy-based interpretation.* This result can also be understood through the lens of entropy. Note that the belief fusion policy utilizes the concavity of the entropy function. From the concavity of  $H(\cdot)$  it follows that

$$H(\bar{\pi}_n^i) \geq w_n^i H(\pi_n^i) + (1 - w_n^i) H\left(\left[\begin{array}{c} \mathbb{P}(\theta = \theta^1 \mid \{Y_t^{\sim i} = y_t^{\sim i}\}_{t=1}^n) \\ \mathbb{P}(\theta = \theta^0 \mid \{Y_t^{\sim i} = y_t^{\sim i}\}_{t=1}^n) \end{array}\right]_{\sim i \rightarrow i}\right).$$

Therefore, for an agent aiming to choose the belief fusion parameter for which  $H(\bar{\pi}_n^i)$  is minimized, it has to choose  $w_n^i = 1$  or  $w_n^i = 0$  depending on which of the terms they multiply has a lower value. Since agents are incentivized to maximize the entropy of their shared beliefs, i.e., agent  $\sim i$  chooses  $\alpha^{\sim i}$  to maximize the term that multiplies  $1 - w_n^i$ , recipients minimize their uncertainty by trusting only their own observations, i.e., agent  $i$  inevitably chooses  $w_n^i = 1$  for  $n < \tau^i$ . The optimal confidence thresholds then balance the cost of continuing to collect samples against the potential cost of making an incorrect decision.  $\diamond$

### C. Conditional Probability of Error

We have established that both agents stop at the same time. However, what's the benefit to the agent initiating the stopping? This section analyzes the accuracy of agents' decisions within this framework. In particular, we derive the conditional probability of error for each agent at the time they stop collecting information and make their final decision.

**Proposition 3.** For almost all sample paths (except a set of  $\mathbb{P}$  measure 0), the conditional probability of error of the agent initiating the stopping is less than or equal to  $\beta$ , while for the other agent is less than or equal to 0.5.

*Proof.* See [16] for the proof.  $\square$

The above proposition shows an important competitive dynamic between agents: the agent who initiates the stopping process (by reaching their confidence threshold first) achieves a conditional probability of error that is bounded by the pre-defined parameter  $\beta$ . Meanwhile, the other agent, who must decide under insufficient information, can only guarantee that their conditional probability of error is bounded by 0.5.

## IV. NUMERICAL EXAMPLE

This section demonstrates the algorithm with an example. Suppose the common prior known to both agents is

$$\mathbb{P}(\theta = \theta^1) = \mathbb{P}(\theta = \theta^0) = \frac{1}{2}.$$

The distributions of agents under the null ( $\theta = \theta^0$ ) and the alternate ( $\theta = \theta^1$ ) hypotheses are listed in Table I. Let the observations be independent in time. We have

$$\begin{aligned} D_{KL}(\mathbb{P}(Y_n^A \mid \theta = \theta^0), \mathbb{P}(Y_n^A \mid \theta = \theta^1)) &= 0.14 \\ D_{KL}(\mathbb{P}(Y_n^B \mid \theta = \theta^0), \mathbb{P}(Y_n^B \mid \theta = \theta^1)) &= 0.496. \end{aligned}$$

TABLE I

DISTRIBUTIONS  $\mathbb{P}(Y_n^i | \theta)$  OF THE AGENTS UNDER DIFFERENT STATES OF NATURE.

Agent A					
$\theta = \theta^0$	0.1	0.2	0.1	0.3	0.3
$\theta = \theta^1$	0.2	0.15	0.25	0.2	0.2
Agent B					
$\theta = \theta^0$	0.15	0.25	0.15	0.25	0.2
$\theta = \theta^1$	0.4	0.05	0.35	0.1	0.1

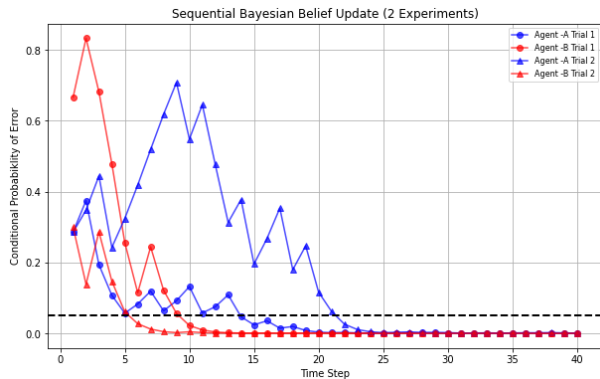


Fig. 3. Decay of conditional probability of error with time

In Fig. 3, we plot the decay of the conditional probability of error for both agents across two different trials. We set  $\beta = 0.05$ . Observe that, in trial 1, agent A achieves the desired probability of error in 14 iterations, while agent B achieves the same in only 10 iterations. In trial 2, agent A takes 21 iterations while agent B takes only 5 iterations. For  $\beta = 0.05$ , the average stopping times over 100 trials are

$$\tau_{\text{avg}} = 6.95, \quad \tau_{\text{avg}}^A = 19.70, \quad \tau_{\text{avg}}^B = 7.95.$$

For  $\beta = 0.01$ , we have

$$\tau_{\text{avg}} = 10.34, \quad \tau_{\text{avg}}^A = 30.52, \quad \tau_{\text{avg}}^B = 10.85.$$

Thus,  $\tau_{\text{avg}}$  for the system is closer to  $\tau_{\text{avg}}^B$ , the average stopping time for agent B. This is because the higher KL divergence between the distributions under the null and alternative hypotheses enables agent B to distinguish between the two hypotheses quickly. Among the 100 trials, there were very few trials where agent A initiated the stopping.

## V. CONCLUSION

This paper studied sequential hypothesis testing in a competitive multi-agent setting under information asymmetry. The analysis showed that when agents interact under strategically manipulated communication channels, conventional information-sharing protocols require fundamental reconsideration. Specifically, we showed that randomizing between truthful and inverted beliefs is the optimal signaling strategy to prevent the other agent from stopping first. We also showed that agents achieve optimal performance by maintaining separate processing streams for their own observations versus information received from others. In particular, agents should update their beliefs based exclusively on their private observations until reaching their confidence thresholds. The information received from counterparts serves pri-

marily as a mechanism to anticipate stopping decisions rather than as a direct input to belief formation. This result provides a practical approach to managing potentially manipulated information in distributed decision-making systems. Finally, we showed that the agent who reaches their confidence threshold first has an advantage, which creates an asymmetry with the first agent achieving a bounded error probability while the second faces compromised accuracy.

Several promising directions exist for extending this research. Extending the analysis to networks with multiple agents would enhance the applicability of our work to organizational and social systems. Moreover, developing adaptive learning mechanisms that enable agents to discover optimal strategies under uncertainty about their counterparts' behavior represents an important practical extension.

## REFERENCES

- [1] F. D. Foster and S. Viswanathan, "Strategic trading with asymmetrically informed traders and long-lived information," *Journal of Financial and Quantitative Analysis*, vol. 29, no. 4, pp. 499–518, 1994.
- [2] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 86–105, 2012.
- [3] N. Gupta, T. T. Doan, and N. Vaidya, "Byzantine fault-tolerance in federated local SGD under  $2f$ -redundancy," *IEEE Transactions on Control of Network Systems*, vol. 10, no. 4, pp. 1669–1681, 2023.
- [4] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, no. 4, pp. 501–510, 1981.
- [5] D. Teneketzis and P. Varaiya, "The decentralized quickest detection problem," *IEEE Transactions on Automatic Control*, vol. 29, no. 7, pp. 641–644, 1984.
- [6] K.-K. Kim and J. S. Shamma, "Multi-agent sequential hypothesis testing," in *53rd IEEE Conference on Decision and Control*, pp. 1942–1947, 2014.
- [7] X. Ren, J. Yan, and Y. Mo, "Binary hypothesis testing with Byzantine sensors: Fundamental tradeoff between security and efficiency," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1454–1468, 2018.
- [8] K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo, "Detection in adversarial environments," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3209–3223, 2014.
- [9] M. Barni and B. Tondi, "Binary hypothesis testing game with training data," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4848–4866, 2014.
- [10] A. Mitra, J. A. Richards, and S. Sundaram, "A new approach to distributed hypothesis testing and non-Bayesian learning: Improved learning rate and Byzantine resilience," *IEEE Transactions on Automatic Control*, vol. 66, no. 9, pp. 4084–4100, 2020.
- [11] S. Saritaş, S. Gezici, and S. Yüksel, "Hypothesis testing under subjective priors and costs as a signaling game," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 5169–5183, 2019.
- [12] A. Lalitha, T. Javidi, and A. D. Sarwate, "Social learning and distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6161–6179, 2018.
- [13] A. N. Shiryaev, *Optimal Stopping Rules*. Springer Berlin, Heidelberg, 2007.
- [14] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Chapman and Hall/CRC, 2014.
- [15] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015.
- [16] A. Raghavan, M. U. B. Niazi, and K. H. Johansson, "Sequential binary hypothesis testing with competing agents under information asymmetry," *arXiv preprint arXiv:2504.02743*, 2025.