Interactive Bayesian Deception under Strategic Timing

Amitalok J. Budkuley
IIT Kharagpur
Kharagpur, India
amitalok@ece.iitkgp.ac.in

Arya Choudhari
IIT Kharagpur
Kharagpur, India
choudhuriarya@gmail.com

ABSTRACT

We introduce a novel framework of interactive Bayesian deception under strategic timing, which integrates ideas from Bayesian persuasion, sequential decision-making, and stochastic change point detection. Here a receiver (Agent) observes a sequence of signals to detect a latent change point in an underlying state of the world. However, unlike standard settings, the sequence is generated strategically by a selfish sender (Principal), who designs the signal distributions, both before and after the change, with the goal of maximizing their own utility, and thereby possibly inducing timing deception. Their utilities may differ and are defined over the true change point and the receiver's random stopping rule, creating an interactive and coupled dynamic wherein the sender can either align with or mislead the receiver by shaping the observed sequence. We show that, under separable utilities and fixed sender strategy, the receiver's optimal stopping policy reduces to a simple one-sided threshold test. From the sender's perspective, the problem becomes a form of dynamic persuasion, albeit under a novel stopping-time-dependent objective, which we characterize in a spirit similar to the classic work of Kamenica and Gentzkow [AER, 2011]. Our work opens up a new class of dynamic information design problems where strategic timing and deception fundamentally interact.

KEYWORDS

Interactive deception; Change point detection; Sequential inference; Random stopping time; Persuasion.

ACM Reference Format:
Amitalok J. Budkuley and Arya Choudhari. 2026. Interactive Bayesian Deception under Strategic Timing. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/GUCM1758

1 INTRODUCTION

1.1 Motivation and Problem Statement

The heart of many strategic interactions in economics, finance, and security lies in the tension between two agents: a well-informed *Principal (sender)* who designs an information release strategy, and an *Agent (receiver)* who must make a critical, time-sensitive decision based on that information. The Principal seeks to influence, and potentially manipulate, the Agent's beliefs and actions to maximize their own utility, while the Agent seeks to minimize their cost by optimally balancing the delay of an *irreversible* action against the risk of acting *prematurely* on insufficient information.

Such *sequential* decision-making under strategic information asymmetry finds wide application in multi-agent systems, particularly those demanding timely awareness of system failures or state changes by remote observers. We study optimal information design in an environment where the agent's primary task is *change detection*, i.e., timely identification of an abrupt and random state change (Λ), that is known to the Principal but hidden from the Agent.

Motivating Example: *(Financial Deception and Rating Agencies)* Consider a large *Hedge Fund or Investment Bank (the Principal)* holding a distressed asset. The Fund seeks to delay the market's awareness of the asset's failure (Λ) to allow time for offloading the position (i.e., to maximize the detection time T^*). The *Rating Agency (the Agent)* must detect this failure to issue a downgrade, incurring cost c_2 for being late (delay causing regulatory fines) , or cost c_1 for a false alarm (unjustified downgrade and market panic). The Fund controls the financial data shared with the Agency ($\underline{\pi}$), making key metrics ambiguous. Furthermore, the Agent can choose to run costly, high-fidelity auditing or forensic accounting checks ($\mathbf{F}^*$) to actively clarify the ambiguous signals. This scenario illustrates the central conflict (summarized in Fig. 1): the Principal practices *deception* to maximize T^*, while the Agent employs *strategic timing* (T^*) and active *control* ($\mathbf{F}^*$) in their *interaction.*

We formulate this **interactive Bayesian deception under strategic timing** problem as a dynamic Stackelberg game. The Principal seeks to maximize their expected utility $\mathbb{E}[v(\Lambda, T^*)]$ by strategically designing the signal process ($\underline{\pi}$). The Agent, in response, solves an optimal decision problem by choosing both the detection time (T^*) and an adapted information control ($\mathbf{F}^*$) to minimize their own detection costs $\mathbb{E}_{\underline{\pi}}[u(\Lambda, T, \mathbf{F})]$.

1.2 Related Work

Our work is positioned at the intersection of three key domains: *Bayesian Persuasion (BP)*, *Sequential Analysis for Change Detection*, and *Stochastic Control.*

Bayesian Persuasion and Sequential Deception Inspiration for our strategic communication model lies in the seminal work of Kamenica and Gentzkow [7], which introduced optimal information design in a static game (see also Kamenica [6] for a survey). Following this work, wide research has explored extensions including (but not limited to) costly signaling [13], multiple agents [5, 22], and noisy signaling [9, 19]. These are representative contributions within a vast literature. Closer to our theme, significant effort has been dedicated to *sequential persuasion* [1–3, 23], where authors analyze dynamic signaling and receiver action over time. However, these works, and subsequently related contributions, typically analyze receivers with deterministic, finite and non-random stopping times. In a recent work, Koh et al. [8] explored the case of

---

This work is licensed under a Creative Commons Attribution International 4.0 License.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/GUCM1758

random (optimal) stopping in classic (continuous time) persuasion, presenting interesting results, albeit with a different goal. However, unlike our model, they studied a *non-interactive* setting without any receiver's *active control* over the observation process.

Sequential Analysis and Change Detection The agent's optimal timing decision (T^*) is rooted in the classic theory of sequential analysis founded in Wald's seminal work [20] (see also [21]). However, closer to the theme of this work, is the *quickest change point detection (QCPD)* problem. Procedures like the CUSUM and Shiryaev-Roberts methods [18] define the classic benchmarks for optimal stopping rules when seeking to *detect a random time of change.* The vast majority of this literature, see for example [15, 16, 18] for a detailed account of the several extensions studied, typically assumes the signal process is *passive and exogenous*, meaning the observer cannot influence the signal characteristics. Our model differs by incorporating the principal's strategic design of $\underline{\pi}$ (making the signal *endogenous* to the game) and by incorporating the agent's control $\mathbf{F}$. The classic Bellman principle, and dynamic programming are standard tools in their study, and in particular, appear through the generalized *Wald-Bellman principle* for characterizing solutions for such optimal stopping (and free boundary) problems (see Osękowski [14], Shiryaev [16]).

Stochastic Control and Active Sensing The agent's adaptive probing strategy ($\mathbf{F}$) connects directly to problems in stochastic control and active sequential hypothesis testing [11, 12]. These works focus on optimally selecting measurements from a set of available sensors or experiments to reduce uncertainty. Typically, these models capture the decision-maker's ability to exert *control* over signal quality, but do not incorporate a second, *self-interested* strategic agent (the principal) who is simultaneously *designing* the signal characteristics (in our case, for optimal deception).

Our Work To the best of our knowledge, our problem framework achieves a unique synthesis by bringing together the above three thrust areas, and embedding two distinct layers of strategic endogeneity into that sequential decision problem. Such a synthesis was demonstrated in the context of persuasion and sequential hypothesis testing in [4] (see also [17]). The existing literature typically addresses only one: (i) *Principal-Side Endogeneity* (persuasion), where the signal process $\underline{\pi}$ is strategically *designed* by the sender, but the receiver is passive; or (ii) *Agent-Side Endogeneity* (active sensing), where the receiver's control $\mathbf{F}^*$ *selects* the information to be observed, but the underlying statistical environment ($\underline{\pi}$) is fixed. In essence, we introduce a *two-sided* game of *layered endogeneity*, and seek to characterize a Stackelberg equilibrium in this dynamic setting. The receiver's best-response strategy is a pair comprising a controlled observation sequence and a random stopping rule, and determining it completely requires a significant technical extension of the concavification principle of [7].

1.3 Technical contributions

We provide a complete characterization of the Stackelberg equilibrium $(\underline{\pi}^*, \underline{\Delta}^*_{\underline{\pi}^*})$ for our game. Our key contributions are as follows:

Agent's Coupled Strategy (Theorem 1): For Principal's signalling scheme, say $\underline{\pi}$, we determine agent's optimal best-response strategy $\underline{\Delta}^*_{\underline{\pi}} = (T^*, \mathbf{F}^{T^*})$. The optimal detection time T^* is governed by a *generalized threshold rule* $B(\mu_t)$, and the optimal control $\mathbf{F}^{T^*}$ is selected dynamically to minimize the expected one-step continuation cost.

Principal's Optimization: (Theorem 2): The principal's complex optimization over the high-dimensional signaling scheme $\underline{\pi}$ is simplified to a maximization over the space of terminal posterior beliefs μ_{T^*}. The maximum achievable utility is given by the *concavification*, denoted by $\hat{\hat{V}}$, of the principal's utility function $\hat{V}(\mu_{T^*})$ seen as a function over the belief space.

Contrast with Bayesian Persuasion (BP): The classic result of BP concavifies the utility from a single, static action $\hat{v}(\mu)$. Our result establishes that the concavification principle extends to the dynamic setting, where the function being concavified, $\hat{V}(\mu_{T^*})$, *implicitly incorporates the agent's full optimal sequential strategy* (timing T^* and dynamic control $\mathbf{F}^{T^*}$). In essence, etstablishing that concavification remains the sufficient condition for optimal persuasion under this complex embedded utility is a core technical extension of this work.

Impact of Active Probing (Proposition 2): We show that agent's ability to actively probe may act as a constraint on the principal's deceptive power. The agent's optimal control choice $\mathbf{F}^{T^*}$ reduces the uncertainty available to the principal, strictly lowering the maximum achievable utility compared to any non-interactive setup.

1.4 Paper Structure

The remainder of the paper is organized as follows. Section 2 formally defines the system dynamics and the game structure (referencing Fig. 1). In Section 3, we present our main results, including an illustrative example. The proof of our key results are presented in Section 4, and we make concluding remarks in Section 5.

2 PROBLEM SETUP

2.1 Notation

Let $\mathbb{R}$, $\mathbb{N}$ denote the set of real numbers and natural numbers resp.. Let $[n] := \{1, 2, \cdots, n\}$ for $n \in \mathbb{N}$. Let us denote random variables by upper case letters (eg. X), their values by lower case letters (eg., x), and their alphabets by calligraphic letters (eg. $\mathcal{X}$). Let us denote by $\mathscr{P}(\mathcal{X})$ the simplex of all distributions on set $\mathcal{X}$. Random vectors and their accompanying values are denoted by boldface letters, for example, $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ for some $n \in \mathbb{N}$. We also sometimes use the notation $\mathbf{X}_i^j$ to explicitly indicate the collection $(X_i, X_{i+1}, \cdots, X_j)$, where $i < j \in [n]$. Joint distributions are similarly denoted.

We present a framework for an interactive sequential decision game between a *Principal* (sender) and an *Agent* (receiver). Our framework builds upon elements of sequential decision-making but fundamentally differs from classical Bayesian Persuasion of [7] in two critical aspects: (i) the agent's decision is an optimal *stopping time* T to detect a change point Λ, rather than a final, one-shot action, and (ii) the uncertainty is defined by a sequential, random *change point* Λ, not a static state of the world known *a priori.* Both of these requirements make this problem necessarily *interactive* unlike the one-shot setting in [7].

2.2 System setup and interaction

Refer to the block diagram depicting the problem in Fig. 1. The timing uncertainty centers around a change point Λ, a discrete random variable on $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. The prior distribution $\mu_0 \sim P_\Lambda$, with $\Lambda \sim \texttt{Geo}(q, p)$, is the (zero-modified) Geometric distribution, and is assumed to be fixed and known to both parties:

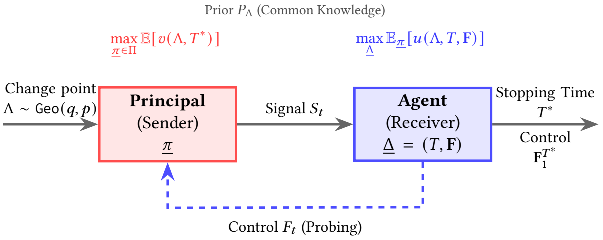

Figure 1: Our problem – Interactive deception under timing

$$\mathbb{P}_\Lambda(\Lambda = k) = \begin{cases} q & \text{if } k = 0 \\ (1-q)p(1-p)^{k-1} & \text{if } k \geq 1 \end{cases} \tag{1}$$

where $q \in [0, 1)$ and $p \in (0, 1)$ are fixed change point parameters.

The interaction unfolds over discrete time $t \in \{1, 2, \dots\}$. The agent's control or *feedback* F_t is chosen at the start of stage t from a finite set $\mathcal{F}$. The principal commits *a priori* to the *signaling scheme*, denoted by $\underline{\pi}$, which defines the fixed conditional distributions of the signal S_t given the true regime (pre-change $\mathbf{P}$ or post-change $\mathbf{Q}$) and the agent's feedback F_t:

$$\underline{\pi} \equiv \{\mathbf{P}_f, \mathbf{Q}_f\}_{f \in \mathcal{F}} \tag{2}$$

The commitment is *ex-ante* and unconditional on future feedback realizations. The signal S_t is drawn based on the current system regime and the agent's control F_t:

$$S_t \sim \begin{cases} \mathbf{P}_{F_t} & \text{if } t < \Lambda \text{ (Pre-change)} \\ \mathbf{Q}_{F_t} & \text{if } t \geq \Lambda \text{ (Post-change)} \end{cases}$$

where $\mathbf{P}_f, \mathbf{Q}_f \in \mathscr{P}(\mathcal{S}), \forall f \in \mathcal{F}$, denote the (conditionally memoryless) distributions of S_t in $\mathscr{P}(\mathcal{S})$ under the pre- and post-change regimes, respectively, when the agent chooses feedback $F_t = f$.

Remark 1 (Absolute Continuity and Separation). *We make typical assumptions (for all $f \in \mathcal{F}$) for analytical tractability:* absolute continuity $\mathbf{Q}_f \ll \mathbf{P}_f$, *(for existence of the likelihood ratio for Bayesian updates), and* separation $\mathbf{P}_f \neq \mathbf{Q}_f$ *(under an appropriate statistical distance) to allow the agent to actually distinguish the regimes.*

The agent's decision-making is governed by the filtration $\mathscr{F}_t$, which represents all information available to agent up to time t:

$$\mathscr{F}_t = \sigma(\{S_1, F_1, \dots, S_t, F_t\})$$

and $\sigma(\cdot)$ denotes the sigma-algebra. The feedback process $(F_t)_{t \geq 1}$ is an *adapted process* where F_t must be measurable with respect to $\mathscr{F}_{t-1}$. The *posterior belief*, denoted by μ_t, is a distribution over Λ conditioned on the history $\mathscr{F}_t$:

$$\mu_t(\cdot) = \mathbb{P}(\Lambda \in \cdot \mid \mathscr{F}_t) \tag{3}$$

In fact, $\mu_t(\cdot)$ is updated at each time step t under the standard notion of *Bayes plausibility*, which we define next (specialized to our problem notation).

Definition 1 (Bayes plausibility). *Let the prior distribution on the change point be P_Λ. A stochastic process of posterior beliefs $(\mu_t)_{t \geq 0}$ induced by the signaling scheme $\underline{\pi}$ and the agent's strategy $\underline{\Delta}$ is* Bayes plausible *or $\underline{\pi}$-*Bayes plausible *if, at any time $t \in \mathbb{N}_0$, the unconditional expectation of the posterior distribution μ_t equals the prior distribution P_Λ. That is, for any measurable set $A \subseteq \mathbb{N}_0$:*

$$\mathbb{E}_{\underline{\pi}, \underline{\Delta}}[\mu_t(A)] = P_\Lambda(A) =: \mu_0(A), \tag{4}$$

where $\mu_t(A) = \mathbb{P}(\Lambda \in A \mid \mathscr{F}_t)$ is the posterior probability of the event $\Lambda \in A$ given the information at time t.

The agent's strategy is a pair $\underline{\Delta} = (T, \mathbf{F}^T)$, where the detection time T must be an *$\mathcal{F}_t$-stopping time* and $\mathbf{F}^T = (F_1, \dots, F_T)$ is the sequence of controls up to T.

2.3 Optimization Problems

The game's solution is a Stackelberg equilibrium defined by a pair comprising strategies of the principal, namely $\underline{\pi}^*$, and the agent, namely $\underline{\Delta}^*_{\underline{\pi}^*}$.

Agent's Problem (Follower): The agent's strategy $\underline{\Delta}$ maximizes expected utility u. The generic agent utility $u : \mathbb{N}_0 \times \mathbb{N}_0 \times \mathcal{F}^\infty \to \mathbb{R}$ depends on the true change point Λ, the terminal action $\hat{\Lambda}$, and the control $\mathbf{F}^T$. We specify it by $u(\Lambda, \hat{\Lambda}, \mathbf{F}^T)$.

Remark 2. *In general sequential decision problems, the stopping time T and the terminal action $\hat{\Lambda}$ can be distinct. Following QCPD literature, in particular Peskir and Shiryaev [15], Shiryaev [16], we impose, for ease of presentation, the simplification that the stopping time is the estimate itself: $\hat{\Lambda} = T$. Thus, the utility is simplified to $u(\Lambda, T, \mathbf{F}^T)$. Note that one may consider more general* separable *utility functions inspired by other goals [18, Ch. 6].*

Specifically, we assume that the utility is the negation of the *separable cost function C_A*, i.e., $u(\Lambda, T, \mathbf{F}^T) = -C_A(\Lambda, T, \mathbf{F}^T)$, where:

$$C_A(\Lambda, T, \mathbf{F}^T) := \underbrace{c_1 \cdot \mathbf{1}_{\{T < \Lambda\}}}_{\text{False Alarm Cost}} + \underbrace{c_2 \cdot (T - \Lambda)^+}_{\text{Detection Delay Cost}} + \underbrace{\sum_{t=1}^{T} C_f(F_t)}_{\text{Probing Cost}} \tag{5}$$

where $c_1, c_2 > 0$ are fixed constants, and $C_f : \mathcal{F} \to \mathbb{R}_+$, captures the cost receiver's incurred cost on generating a certain feedback $F_t = f_t$. Hence, the agent solves the following optimization problem:

$$\underline{\Delta}^*_{\underline{\pi}} \in \arg\max_{\underline{\Delta} = (T, \mathbf{F}_1^T)} \mathbb{E}_{\underline{\pi}}[u(\Lambda, T, \mathbf{F}_1^T)] \tag{6}$$

Remark 3. *The choice of cost function C_A in* (5) *ensures the agent's optimal strategy $\underline{\Delta}^*$ captures the essential trade-offs required in our setting as follows:*

- *The terms involving c_1 (False Alarm) and c_2 (Detection Delay) define the core problem, forcing the Agent to balance the risk of stopping too early (high false alarm cost) against the cost of stopping too late (high delay cost).*
- *The final term, $\sum_{t=1}^{T} C_f(F_t)$, introduces a strategic element by penalizing the Agent's active information gathering (or control F_t). The cost structure compels the Agent to strategically select*

the optimal feedback (probe) intensity at each step, thereby justifying the interactive nature of the problem.

Our canonical separable *structure is inspired from standard literature in sequential analysis, cf. [16, 18], as it facilitates meaningful analysis via classic dynamic programming techniques (unlike in non-separable settings).*

Principal's Problem (Leader): The principal maximizes expected utility $v(\Lambda, T)$. The *principal's utility* v depends only on the true change point Λ and the agent's detection time T:

$$v : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{R}, \quad v(\Lambda, T) \tag{7}$$

The principal's optimization problem is:

$$\underline{\pi}^* \in \arg\max_{\underline{\pi} \in \mathbf{\Pi}} \mathbb{E}_{\underline{\pi}}[v(\Lambda, T^*_{\underline{\pi}})] \tag{8}$$

where $T^*_{\underline{\pi}}$ is the agent's optimal stopping time derived from (6). $\mathbf{\Pi}$ is the admissible set of signaling schemes, and may be unconstrained as well. Here is our equilibrium notion for this strategic interaction.

Definition 2 (Equilibrium). *An equilibrium for the interactive sequential detection game is a pair of strategies* $(\underline{\pi}^*, \underline{\Delta}^*_{\underline{\pi}^*})$ *comprising the Principal's optimal signaling scheme* $\underline{\pi}^* \in \mathbf{\Pi}$ *and the Agent's optimal strategy* $\underline{\Delta}^*_{\underline{\pi}^*} = (T^*, \mathbf{F}^{T^*})$*, which satisfies the coupled Stackelberg optimality conditions in* (8) *and* (6)*.*

For clarity, we summarize the two-party interaction below.

Time Evolution and Nature of Play

The game proceeds in two phases:

1. **Initial Commitment (Ex-Ante):** The principal fixes $\underline{\pi} \in \mathbf{\Pi}$, and nature chooses the change point Λ according to P_Λ.
2. **Sequential Interaction Phase (Stage $t \in \{1, 2, \dots\}$):** This phase repeats until the agent stops at time T. In any stage t:
   - (a) The agent chooses *control* $F_t = f_t \in \mathcal{F}$ based on $\mathcal{F}_{t-1}$.
   - (b) The system transmits *signal* $S_t = s_t$ based on $\underline{\pi}$, Λ, and f_t.
   - (c) The agent observes s_t, updates the belief μ_t, and applies the strategy $\underline{\Delta}$: *stop* ($T = t$, game ends) or *continue* (proceed to next stage).

3 MAIN RESULTS

In this section, we present the key analytical contributions of our work. Our results characterize the Agent's optimal combined strategy (stopping time and adapted control) and the Principal's optimal signaling scheme in the resulting Stackelberg equilibrium $(\underline{\pi}^*, \underline{\Delta}^*)$. Furthermore, our propositions formally establish the connection to Bayesian persuasion theory by characterizing the optimal utility as a concavification and quantifying the benefit of deception in this sequential, interactive framework.

3.1 Optimal Agent strategy: Characterization via the belief process

The Agent's optimal strategy $\underline{\Delta}^*$ solves the inner maximization problem (6) via dynamic programming. The Agent's maximum expected utility is captured by the value function $u(\mu)$, which depends on the full posterior belief process, μ_t.

The core statistic governing the Agent's decision is the (Shiryaev) statistic, ρ_t, the marginal probability that the change has occurred by time t:

$$\rho_t = P(\Lambda \leq t \mid \mathcal{F}_t).$$

The Agent's total cost is composed of a per-period probing cost, $C_f(F_t)$, and a terminal stopping cost, $\mathbf{C}_{\text{Stopping}}(T^*, \Lambda)$, which penalizes detection delays and false alarms (see (5)).

The Agent's decision relies on two utility components based on the continuation strategy $\underline{\Delta}_t = (T^* \mid T^* \geq t, \{F_k\}_{k=t}^{T^*})$ (the optimal stopping time and probing path from time t onward):

1. Optimal continuation utility ($u(\mu_t)$): The value function representing the maximum expected utility achievable from time t, conditional on the current belief μ_t:

$$u(\mu_t) := -\min_{\underline{\Delta}_t} \mathbb{E}_{\underline{\pi}, \underline{\Delta}_t} \left[\mathbf{C}_{\text{Stopping}}(T^*, \Lambda) + \sum_{k=t}^{T^*} C_f(F_k) \mid \mu_t \right] \tag{9}$$

2. Per-period probing utility ($u_f(f)$): The immediate utility (negative cost) from selecting control $f \in \mathcal{F}$:

$$u_f(f) := -C_f(f) \tag{10}$$

We now present our first result.

Theorem 1 (Optimal Agent strategy). *For a fixed principal's signaling scheme* $\underline{\pi}$*, the agent's optimal strategy* $\underline{\Delta}^*_{\underline{\pi}} = (T^*, \mathbf{F}^{T^*})$ *is characterized as follows:*

1. Optimal control (probing): The optimal feedback control $F_t^* \in \mathcal{F}$ *is chosen at time t to maximize the expected one-step continuation utility.* F_t^* *is a deterministic function of the full posterior* μ_{t-1}*:*

$$F_t^* = f^*(\mu_{t-1}) \in \arg\max_{f \in \mathcal{F}} \left\{ u_f(f) + \mathbb{E}_{\underline{\pi}}[u(\mu_t) \mid \mu_{t-1}, F_t = f] \right\}. \tag{11}$$

2. Optimal stopping: The optimal detection time T^* *is a first-passage time given by a threshold rule on the Shiryaev statistic* ρ_t*:*

$$T^* = \inf\{t \geq 1 : \rho_t \geq B(\mu_t)\}, \tag{12}$$

where $B(\mu_t)$ *is the optimal stopping boundary, determined by comparing the immediate stopping utility with the optimal continuation utility* $u(\mu_t)$*.*

The proof of this result appears in Sec. 4.

Remark 4 (Connection to Generalized Shiryaev's Stopping Rule). *Due to the separable nature of the agent's cost function* (5)*, the optimal stopping time* T^* *corresponds to a generalized Shiryaev stopping procedure [16]. The presence of non-trivial probing costs* $C_f(F_t)$ *results in a time-varying and belief-dependent boundary* $B(\mu_t)$*, distinguishing it from the classic SR rule, which typically uses a constant, time-invariant threshold when the cost function is simplified [16].*

3.2 Principal's problem: Reduction and concavification

The principal's problem is the maximization of $\mathbb{E}[v(\Lambda, T^*_{\underline{\pi}})]$ subject to the constraint that the agent plays the optimal strategy $\underline{\Delta}^*_{\underline{\pi}}$.

Let $\hat{V}(\mu_{T^*})$ be the principal's expected utility given the terminal posterior belief μ_{T^*} that leads the agent to stop at $T^* = T^*_{\underline{\pi}}$:

$$\hat{V}(\mu_{T^*}) := \mathbb{E}_{\mu_{T^*}}[v(\Lambda, T^*_{\underline{\pi}})] \tag{13}$$

Definition 3 (Concave Closure). *The concave closure of the Principal's utility function $\hat{V} : \mathscr{P}(\mathbb{N}_0) \to \mathbb{R}$, denoted $\hat{\hat{V}}$, is the minimum concave function that majorizes $\hat{V}$. Formally, for any belief $\mu \in \mathscr{P}(\mathbb{N}_0)$, $\hat{\hat{V}}(\mu)$ is defined as the supremum of the expected payoff:*

$$\hat{\hat{V}}(\mu) = \sup_{\alpha} \left\{ \mathbb{E}_{\alpha}[\hat{V}(\mu')] \mid \alpha \in \mathscr{P}(\mathscr{P}(\mathbb{N}_0)), \mathbb{E}_{\alpha}[\mu'] = \mu \right\}, \quad (14)$$

where the supremum is taken over all probability distributions α over the belief space $\mathscr{P}(\mathbb{N}_0)$, and $\mathbb{E}_{\alpha}[\mu'] = \mu$ is the Bayes plausibility (martingale) constraint.

Theorem 2 (Optimal deception). *The Principal's maximal expected utility is given by the concave closure of $\hat{V}$ evaluated at the prior belief μ_0:*

$$\max_{\underline{\pi} \in \Pi} \mathbb{E}_{\underline{\pi}}[v(\Lambda, T^*_{\underline{\pi}})] =: \hat{\hat{V}}(\mu_0). \quad (15)$$

Remark 5 (Equivalence to belief optimization). *Theorem 2 establishes the fundamental equivalence between optimizing the signaling scheme $\underline{\pi}$ (a set of conditional probability kernels) and optimizing over the set of terminal posterior belief distributions $\alpha \in \mathscr{P}(\mathscr{P}(\mathbb{N}_0))$. The maximum value, $\hat{\hat{V}}(\mu_0)$, is achieved by an optimal distribution α^* that satisfies the Bayes Plausibility constraint. The Principal chooses the optimal $\underline{\pi}^*$ that generates this distribution α^*.*

Remark 6 (Comparison to classic Bayesian persuasion). *This result extends the seminal work of Kamenica and Gentzkow (2011) to the interactive, sequential change detection setting. In the classic BP model, the principal maximizes $\mathbb{E}_{\alpha}[\hat{v}(\mu)]$ where $\hat{v}(\mu)$ is the receiver's utility from a single, final action given posterior μ. In our problem, the principal maximizes $\mathbb{E}_{\alpha}[\hat{V}(\mu_{T^*})]$ where $\hat{V}(\mu_{T^*})$ is the principal's expected utility derived from the agent's entire optimal sequential strategy (stopping time $T^*_{\underline{\pi}}$ and control $\mathbf{F}^{T^*}$).*

3.3 The Stackelberg equilibrium and benefit of deception

Corollary 1 (Stackelberg Equilibrium). *The game admits a Stackelberg equilibrium $(\underline{\pi}^*, \underline{\Delta}^*_{\underline{\pi}^*})$ where:*

1. *The principal's optimal scheme $\underline{\pi}^*$ is any scheme that achieves the maximum expected utility defined by the concavification $\hat{\hat{V}}(\mu_0)$ in Theorem 2.*
2. *The agent's optimal strategy $\underline{\Delta}^*_{\underline{\pi}^*} = (T^*, \mathbf{F}^{T^*})$ is the best response to $\underline{\pi}^*$, determined by the coupled optimal control* (11) *and optimal stopping rule* (12) *in Theorem 1.*

Proof. The existence of the equilibrium is guaranteed by the solutions to the nested optimization problems. The Agent's best response $\underline{\Delta}^*_{\underline{\pi}}$ is characterized by Theorem 1 via backward induction on the optimal value function. The Principal's problem is then the Bayesian Persuasion problem over the terminal beliefs, which, by Theorem 2, has a solution $\underline{\pi}^*$ that generates a distribution of terminal beliefs α^* achieving the concave closure $\hat{\hat{V}}(\mu_0)$. The pair $(\underline{\pi}^*, \underline{\Delta}^*_{\underline{\pi}^*})$ constitutes the Stackelberg equilibrium. □

We now use the language of the *benefit of deception*, similar to [7], to compare the Principal's optimal utility against the utility achievable under any of Principal's *deception-neutral* signaling schemes. A signalling scheme $\underline{\pi} \in \Pi$ is said to be *deception-neutral* if the Principal's resulting utility equals the one under prior belief $\hat{V}(\mu_0)$, i.e. $\mathbb{E}_{\underline{\pi}}[v(\Lambda, T^*_{\underline{\pi}})] = \hat{V}(\mu_0)$. Let $\hat{V}_{\text{NoDeception}}$ denote this utility achieved without the benefit of belief manipulation:

$$\hat{V}_{\text{NoDeception}} := \hat{V}(\mu_0) \quad (16)$$

Note that such a *deception-neutral* signalling scheme, often referred to as an *uninformative signaling scheme*, is always possible; for instance, the Principal sends no signal, and the Agent acts based solely on it prior.

We are now in a position to state the following elementary result on the (im)possibility of benefit of deception in our problem.

Proposition 1 (Benefit of deception). *The Principal's optimal expected utility $\hat{\hat{V}}(\mu_0)$ is greater than or equal to the utility achievable under any deception-neutral scheme:*

$$\max_{\underline{\pi} \in \Pi} \mathbb{E}_{\underline{\pi}}[v(\Lambda, T^*_{\underline{\pi}})] = \hat{\hat{V}}(\mu_0) \geq \hat{V}_{\textit{NoDeception}} \quad (17)$$

The benefit of deception is strictly positive ($\hat{\hat{V}}(\mu_0) > \hat{V}_{\textit{NoDeception}}$) if and only if the function $\hat{V}(\mu_{T^*})$ is not concave at the prior μ_0.

Proof. The inequality $\hat{\hat{V}}(\mu_0) \geq \hat{V}_{\text{NoDeception}}$ follows directly from the definition of the concave closure (Definition 3) since $\hat{V}_{\text{NoDeception}} = \hat{V}(\mu_0)$ is the value obtained by choosing the belief distribution α that places all mass on the prior belief $\mu' = \mu_0$. This is a feasible choice, making $\hat{V}(\mu_0)$ a lower bound for the supremum $\hat{\hat{V}}(\mu_0)$. The condition for strict inequality ($\hat{\hat{V}}(\mu_0) > \hat{V}_{\text{NoDeception}}$) is a direct consequence of the core result of Bayesian Persuasion: persuasion is beneficial if and only if the function $\hat{V}$ is not concave at the prior μ_0. □

The following simple corollary follows immediately.

Corollary 2 (Impact of Agent's Probing). *The agent's ability to adaptively choose F^*_t results in the following:*

1\. Efficiency gain: *The agent's realized cost in the interactive game is strictly lower than the cost realized in any non-interactive game where the probing sequence is fixed a priori, provided $\mathcal{F}$ contains at least one informative and non-zero-cost control.*

2\. Constrained Principal: *Principal's optimal utility $\mathbb{E}[v(\Lambda, T^*_{\underline{\pi}^*})]$ in the interactive game is upper-bounded by the maximum utility achievable in the non-interactive counterpart. The agent's optimal probing behavior $\mathbf{F}^*$ acts as an additional constraint on the principal's ability to manipulate the belief process.*

Proof. We prove each part individually.

1\. *Efficiency gain:* This follows from the principle of optimality in dynamic programming. Since the agent *chooses* the optimal control sequence $\mathbf{F}^*$ adaptively at every step, their realized cost must be less than or equal to the cost achieved under any fixed, non-adaptive sequence $\mathbf{F}$ (i.e., $\min_{\mathbf{F}} C_A(\cdot, \cdot, \mathbf{F})$). The strict inequality holds if the ability to adapt $\mathbf{F}^*$ based on the evolving belief μ_t provides a better trade-off between information gain and probing cost.

2\. *Constrained Principal:* The principal's utility $\hat{\hat{V}}(\mu_0)$ is the maximum achieved subject to the agent's best-response constraint $\underline{\Delta}^*_{\underline{\pi}}$. If the agent's action space $\mathbf{F}$ were fixed exogenously, the principal would have a less constrained problem, yielding a utility

$\hat{\hat{V}}_{\text{non-interactive}}(\mu_0)$. Since the agent's optimal probing $\mathbf{F}^*$ depends on $\underline{\pi}$, it tightens the constraint on the belief process, implying $\hat{\hat{V}}(\mu_0) \leq \hat{\hat{V}}_{\text{non-interactive}}(\mu_0)$. □

3.4 Illustrative example: On strict benefit of deception

We conclude with a simple finite-state example illustrating the concavification principle in action: it shows how the Agent's threshold stopping rule can generate a (non-concave) principal value function $\hat{V}$, leading to a strict benefit of deception as guaranteed by Theorem 2 and Proposition 1.

Setup: Let $\Lambda \in \{0, 1, 2\}$ with prior $\mu_0 \sim P_\Lambda$. Both prior μ_0 and posterior μ are such that $\mu_0, \mu \in \mathscr{P}(\{0, 1, 2\}) \subseteq \mathbb{R}^3_+$. Let $\mu(1)$ be the posterior probability that the change occurs at time 1. Suppose the Principal's primitive utility is

$$v(\Lambda, T) := \mathbf{1}\{\Lambda = 1, T = 1\}, \tag{18}$$

and assume the Agent follows a threshold rule

$$T^*(\mu) := \begin{cases} 1, & \mu(1) \geq \bar{\mu}, \\ 2, & \mu(1) < \bar{\mu}, \end{cases} \quad \text{with } \bar{\mu} = 0.3. \tag{19}$$

Then, the Principal's (induced) expected utility is

$$\hat{V}(\mu) = \mathbb{E}_\mu[v(\Lambda, T^*(\mu))] = \begin{cases} \mu(1), & \mu(1) \geq 0.3, \\ 0, & \mu(1) < 0.3, \end{cases} \tag{20}$$

and can be easily verified to be not concave in μ.

Strict benefit from timing deception: Let the prior μ_0 be such that $\mu_0(1) = 0.2 < 0.3$. Then

$$\hat{V}(\mu_0) = 0. \tag{21}$$

The concave closure of $\hat{V}$ on $[0, 0.3]$ is the linear function $f(x) = x, \quad x \in [0, 0.3]$. Evaluating at the prior gives

$$\hat{\hat{V}}(\mu_0) = f(\mu_0(1)) = 0.2 > 0, \tag{22}$$

which guarantees a strict benefit of deception.

Signaling scheme $\underline{\pi}$. This utility improvement is achieved by a Bayes-plausible signaling scheme, say $\underline{\pi}$, that induces two posteriors

$$\mu^{(1)}(1) = 0, \qquad \mu^{(2)}(1) = 0.3, \tag{23}$$

with probabilities 1/3 and 2/3, respectively, ensuring

$$\frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 0.3 = 0.2 = \mu_0(1), \tag{24}$$

so the scheme satisfies Bayes plausibility.

It can be verified that one possible choice for the conditional distribution $\underline{\pi} = \mathbb{P}(S \mid \Lambda)$ is

$\mathbb{P}(S = 2 \mid \Lambda = 1) = 1, \mathbb{P}(S = 2 \mid \Lambda = 0) = 2/3,\ \mathbb{P}(S = 2 \mid \Lambda = 2) = 0,$

with $\mathbb{P}(S = 1 \mid \Lambda = i) = 1 - \mathbb{P}(S = 2 \mid \Lambda = i)$, $i = 0, 1, 2$. Under this scheme, the Agent stops at $T^* = 1$ following signal $S = 2$ and at $T^* = 2$ following signal $S = 1$, yielding expected payoff $\hat{\hat{V}}(\mu_0) = 0.2$.

Other signaling schemes that achieve the same posterior splitting also exist. Note that a binary signal is sufficient here because the posterior only needs to be split across the single threshold $\bar{\mu} = 0.3$; additional signal values do not further increase the Principal's expected utility.

In summary, the Agent's endogenous stopping threshold creates a "kink" in $\hat{V}$, and belief splitting across this threshold strictly increases the Principal's utility, providing a concrete illustration of the concavification principle at play.

4 PROOFS

4.1 Proof of Theorem 1

Agent's optimal strategy $\underline{\Delta}^*_{\underline{\pi}} = (T^*, \mathbf{F}^{T^*})$ for a fixed signaling scheme $\underline{\pi}$ is the solution to a combined *optimal control* and *optimal stopping* problem in discrete time. We solve this problem via backward induction using dynamic programming on the belief state μ_t. Our key contribution is the reduction of our problem to an optimal (random) stopping problem under active control, and we focus our presentation on that aspect. We state the known results we use, and make clear their correspondence to the scenario in our problem.

Problem Reduction and Value Function. The problem is reduced to the belief space $\mu_t \in \mathscr{P}(\mathbb{N}_0)$ because μ_t is the sufficient statistic for all information contained in the filtration $\mathcal{F}_t$ [18]. The Agent's objective is to maximize the *value function* $u(\mu_t)$, which is defined as the negative of the minimum expected continuation cost; refer (9).

The utility from stopping immediately at time t is the negative expected terminal cost:

$$u_S(\mu_t) := -\mathbb{E}_{\mu_t}[\mathrm{C}_{\text{Stopping}}(t, \Lambda)] \tag{25}$$

Having setup the value function, we now proceed to the next step of characterizing its solution.

Using the Wald-Bellman principle. The Agent's inner problem is a classical sequential control and stopping problem [15, 18], combining Wald's theory of sequential analysis with Bellman's principle of optimality. We characterize $u(\mu_t)$ by drawing upon standard results for the existence and structure of the solution to this coupled problem. Our specific presentation is particularly inspired from [11, 14, 18]. First, we present the following short recap.

Classical optimal stopping and control [15, 18]. Consider a general discrete-time problem with state $\tilde{Z}_t$, control $\tilde{f}_t \in \tilde{\mathcal{F}}$, and stopping time $\tilde{T} \in \mathcal{T}$. The objective is to maximize the expected payoff $\tilde{J}$ over $\mathcal{T}$, the set of all stopping times adapted to the underlying filtration:

$$\max_{\tilde{T} \in \mathcal{T}, \{\tilde{f}_t\}_{t=1}^{\tilde{T}-1} \in \tilde{\mathcal{F}}} \tilde{J} = \max_{\tilde{T} \in \mathcal{T}, \{\tilde{f}_t\}_{t=1}^{\tilde{T}-1} \in \tilde{\mathcal{F}}} \mathbb{E}\left[\tilde{g}(\tilde{Z}_{\tilde{T}}) + \sum_{t=1}^{\tilde{T}-1} \tilde{h}(\tilde{Z}_t, \tilde{f}_t)\right]$$

Here $\tilde{g}(\cdot)$ and $\tilde{h}(\cdot, \cdot)$ are fixed, real-valued and fixed functions.

The Wald-Bellman equation solves this problem.[1]

Lemma 1 (Wald-Bellman equation [14, 15]). *The optimal value function $\tilde{V}(\tilde{Z})$ for a combined optimal control and optimal stopping problem is the unique Wald-Bellman solution :*

$$\tilde{V}(\tilde{Z}_{t-1}) = \max\left\{\tilde{g}(\tilde{Z}_{t-1}), \max_{\tilde{f} \in \tilde{\mathcal{F}}} \left\{\tilde{h}(\tilde{Z}_{t-1}, \tilde{f}) + \mathbb{E}[\tilde{V}(\tilde{Z}_t) \mid \tilde{Z}_{t-1}, \tilde{f}]\right\}\right\}.$$

We refer the reader to [14, 15] for a proof of this result.

We now apply Lemma 1 by making the following correspondence with our problem: $(\tilde{V}, \tilde{Z}, \tilde{g}, \tilde{h}, \tilde{f}) \leftrightarrow (u, \mu, u_S, u_f, F)$. Given

---

[1] This is also called the generalized version of HJB or Bellman equation, which generalizes the classic HJB equation under a random stopping time.

this correspondence, the Agent's value function $u(\mu)$ satisfies the Wald-Bellman equation in Lemma 1 as follows:

$$u(\mu_{t-1}) = \max\left\{ u_S(\mu_{t-1}), \max_{f \in \mathcal{F}} \left\{ u_f(f) + \mathbb{E}_{\underline{\pi}}[u(\mu_t) \mid \mu_{t-1}, F_t = f] \right\} \right\} \tag{26}$$

We now present proofs for the two sub-parts of Theorem 1.

Proof of "Part 1" of Theorem 1: Optimal control. The optimal control F_t^* is determined by the inner maximization in the Wald-Bellman equation (26), conditional on the decision to continue. This directly yields the characterization for F_t^* in (11). The Agent chooses the probing action F_t^* that optimally trades off the immediate probing utility $u_f(f)$ (specified in (10)) against the expected benefit of the resulting, information-rich belief update.

This completes the proof of *Part 1* of the Theorem 1.

Proof of "Part 2" of Theorem 1: Optimal Stopping Rule. The Agent stops at the *first time* T^* when the utility from immediate stopping, $u_S(\mu_{t-1})$, is greater than or equal to the maximum expected continuation utility.[2] The optimal stopping region is the set of all posterior beliefs given by $\mathcal{S} := \{\mu : u(\mu) = u_S(\mu)\}$. As such, it follows that similar to the optimal control vector $\mathbf{F}$, the optimal stopping time $T^* = T^*(\mu)$ will also depend on the posterior μ. However, we considerably simplify that dependence through a simple one-sided threshold test next. We first recapitulate the classic quickest change point detection problem by Shiryaev (cf. [16]).

Recap of Classic Bayesian Change Point Detection [16] The classical Bayesian change point problem seeks the optimal stopping time $\tilde{T}^* \in \mathcal{T}$ that minimizes the expected cost $\tilde{C}$:

$$\min_{\tilde{T} \in \mathcal{T}} \tilde{C} = \min_{\tilde{T} \in \mathcal{T}} \mathbb{E}\left[\tilde{c}_1(\tilde{T} - \tilde{\Lambda})^+ + \tilde{c}_2(\tilde{\Lambda} - \tilde{T})^+ \mid \tilde{\mathcal{F}}_{\tilde{T}} \right]$$

where $\tilde{\Lambda}$ is the change point, and the cost function involves penalties for late detection ($\tilde{c}_1$) and false alarms ($\tilde{c}_2$). The optimal stopping time $\tilde{T}^*$ for this problem (under a zero-modified Geometric prior) is defined by a threshold $\tilde{B}$ on the Shiryaev statistic $\tilde{\rho}_t$:

$$\tilde{T}^* = \inf\{t \geq 1 : \tilde{\rho}_t \geq \tilde{B}\} \tag{27}$$

In the following, we present a re-statement (convenient to our use) of Shiryaev's classic result on QCPD see [16].[3]

Lemma 2 (Shiryaev's test [16]). *In sequential analysis problems, if the stopping utility $\tilde{g}(\tilde{Z})$ and continuation function $\tilde{V}(\tilde{Z})$ are monotonic with respect to the marginal probability $\tilde{\rho}_t$, the optimal stopping region $\tilde{\mathcal{S}}$ is characterized by a threshold rule on the scalar statistic $\tilde{\rho}$.*

In our interactive problem, the Agent's terminal stopping cost $\mathbf{C}_{\text{Stopping}}(T^*, \Lambda)$ (recap of original problem statement) has the same structure as the classic cost $\tilde{C}$. Therefore, it immediately follows that the optimal stopping time T^* is still the first-passage time defined by the critical statistic ρ_t.

However, there is one critical difference: due to the coupling with the optimal control F_t^*, the continuation value $u(\mu_t)$ (and thus the boundary of the stopping region $\mathcal{S}$) depends on the full belief μ_t, although the test retains it single-sidedness. This principle has found application in extensions of change point detection under the setting, for instance, under multi-change points [18, Ch. 6], and other sequential problems like [10, 12].

The optimal stopping rule (refer (12)) is thus:

$$T^* = \inf\{t \geq 1 : \rho_t \geq B(\mu_t)\} \tag{28}$$

The optimal stopping boundary $B(\mu_t)$ is implicitly defined by the Wald-Bellman equation in (26) and the optimal control F_t^*. This completes the proof of *Part 2* of the theorem.

4.2 Proof of Theorem 2

Recall that Theorem 2 characterizes the Principal's optimal utility in the outer optimization problem. The proof crucially relies on showing the equivalence of the Principal's sequential optimization problem to the static or one-shot Bayesian persuasion problem of Kamenica and Gentzkow [7], which they solve via their classic concavification. Similar to our proof of Theorem 1, we bring focus to our contribution by specifically clarifying the reduction that allows us to utilize their characterization. This is not obvious, as our setting involves a random stopping time unlike the single-shot, and deterministic setting of [7].

We quickly recap some key details of the Principal's problem.

Principal's problem. Recall that the Principal's problem:

$$\max_{\underline{\pi} \in \mathbf{\Pi}} \mathbb{E}_{\underline{\pi}}[v(\Lambda, T^*)] \tag{29}$$

where T^* and $\mathbf{F}^{T^*}$ are derived from the Agent's optimal strategy (see Theorem 1). It is pertinent to note that the expectation $\mathbb{E}_{\underline{\pi}}[\cdot]$ is taken over all random variables in the game: the change point Λ, the Agent's observations $\mathbf{X}$, the Agent's controls $\mathbf{F}$, and the Principal's signals $\mathbf{S}$. This will be crucial to complete our reduction by using the law of iterated expectation.

Reduction to belief optimization. The key analytical step in our proof (as also in Kamenica and Gentzkow [7]) is demonstrating that the Principal's utility, though dependent on the full path of sequential actions, can be expressed as a function of the *terminal posterior belief* μ_{T^*} induced by the optimal strategy.

Induced utility function $\hat{V}$. Recall (13) to note that the function $\hat{V}(\mu_{T^*})$ is the Principal's expected utility conditional on the Agent stopping at a belief μ_{T^*}, and given by:

$$\hat{V}(\mu_{T^*}) = \mathbb{E}_{\mu_{T^*}}[v(\Lambda, T^*)] \tag{30}$$

Crucially, the value $\hat{V}(\mu_{T^*})$ incorporates the Agent's *entire* optimal sequential strategy $\underline{\Delta}^*$ that resulted in the terminal belief μ_{T^*}. Since $\underline{\Delta}^*_{\underline{\pi}}$ is completely determined by $\underline{\pi}$, the Principal's optimization over $\underline{\pi}$ is equivalent to optimizing over the distribution of the terminal belief μ_{T^*}, denoted $\alpha \in \mathscr{P}(\mathscr{P}(\mathbb{N}_0))$. We reiterate that α is a distribution on the space of distributions. This is important as the feasible set of such $\alpha \in \mathscr{P}(\mathscr{P}(\mathbb{N}_0))$ are constrained by the Bayesian plausibility constraint, stated next in a form relavant to the proof.

The Bayes Plausibility constraint . The signaling scheme $\underline{\pi}$ must be feasible, which means the resulting distribution of posterior beliefs α must satisfy the Bayes plausibility constraint : the expected value of the posterior belief μ' must equal the prior belief μ_0:

$$\mathbb{E}_\alpha[\mu'] = \mu_0 \tag{31}$$

---

[2] The *first time* event is also called the *first passage time* in literature.

[3] This monograph is a modern presentation of Shiryaev's original work in 1967 in Russian, and its several extensions.

Combining the induced utility $\hat{V}$ with the the above constraint in (31), the Principal's problem is reduced to optimizing over all feasible distributions of terminal beliefs α:

$$\max_{\underline{\pi}\in\Pi} \mathbb{E}_{\underline{\pi}}[v(\Lambda, T^*)] \iff \sup_{\alpha} \left\{ \mathbb{E}_{\alpha}[\hat{V}(\mu')] \mid \mathbb{E}_{\alpha}[\mu'] = \mu_0 \right\} \tag{32}$$

This reduction confirms the statement in Remark 5.

Concavification theorem and its application. The maximization problem on the right side of (32) is the standard optimization problem solved in [7]. To proceed, we quickly recap the basic problem for context and correspondence.

Recap of classic Bayesian persuasion [7]. Consider a static game where the Principal chooses a signal $\tilde{\underline{\pi}}$ to induce a distribution of posterior beliefs $\tilde{\alpha}$ that maximizes $\mathbb{E}_{\tilde{\alpha}}[\tilde{v}(\tilde{\mu})]$, where $\tilde{v}(\tilde{\mu})$ is the Principal's expected utility given the Agent's optimal *static* action $\tilde{a}(\tilde{\mu})$ at posterior $\tilde{\mu}$. The problem is:

$$\sup_{\tilde{\alpha}} \left\{ \mathbb{E}_{\tilde{\alpha}}[\tilde{v}(\tilde{\mu})] \mid \mathbb{E}_{\tilde{\alpha}}[\tilde{\mu}] = \tilde{\mu}_0 \right\}$$

Here, $\tilde{\mu}_0$ is the prior belief. Then, the following result holds.

Lemma 3 (Concavification theorem [7]). *The supremum of the expected utility $\mathbb{E}_{\tilde{\alpha}}[\tilde{v}(\tilde{\mu})]$ subject to the Bayes Plausibility constraint is equal to the concave closure of the utility function $\tilde{v}$, evaluated at the prior belief $\tilde{\mu}_0$:*

$$\sup_{\tilde{\alpha}} \left\{ \mathbb{E}_{\tilde{\alpha}}[\tilde{v}(\tilde{\mu})] \mid \mathbb{E}_{\tilde{\alpha}}[\tilde{\mu}] = \tilde{\mu}_0 \right\} = \hat{\hat{\tilde{v}}}(\tilde{\mu}_0). \tag{33}$$

Using the correspondence: $(\tilde{v}, \tilde{\mu}, \tilde{\mu}_0) \leftrightarrow (\hat{V}, \mu', \mu_0)$, we apply Lemma 3. Here $\hat{V}$ is the Principal's induced sequential utility, and μ' is the terminal belief μ_{T^*}. Applying this lemma to (32):

$$\sup_{\alpha} \left\{ \mathbb{E}_{\alpha}[\hat{V}(\mu')] \mid \mathbb{E}_{\alpha}[\mu'] = \mu_0 \right\} = \hat{\hat{V}}(\mu_0) \tag{34}$$

This result confirms that the Principal's maximum expected utility is the concave closure of $\hat{V}$ evaluated at the prior μ_0, thereby proving Theorem 2. The optimal scheme $\underline{\pi}^*$ is any scheme that generates the distribution α^* achieving this concavification.

5 DISCUSSION AND CONCLUDING REMARKS

The preceding analysis established the Stackelberg equilibrium for the proposed interactive Bayesian detection under strategic timing game. Our work derived the Agent's optimal sequential strategy (Theorem 1) and characterized the Principal's maximal utility using the classic concavification technique from Bayesian persuasion (Theorem 2). Our main contribution lies in developing a novel framework for integrating classical sequential decision-making problems, specifically QCPD, into a strategic information design setting, where *timing* is of essence.

On problem framework. In our opinion, the core value of this work is the *framework*, which models the unique tension over strategic timing between two parties, comprising an informed Principal and an actively probing Agent in a dynamic, sequential environment. The success of our approach hinges on establishing the equivalence between the Principal's optimization over complex signaling schemes and a much simpler optimization over the resulting terminal beliefs, much in the spirit of Kamenica and Gentzkow [7].

Reducing sequential complexity: Our primary technical contribution is demonstrating that the Principal's complex problem, which involves the Agent's optimal sequence of controls ($\mathbf{F}^{T^*}$) and a random, endogenous stopping time (T^*), can be *analytically compressed* into a static optimization over the terminal posterior belief μ_{T^*}. The resulting function $\hat{V}(\mu_{T^*})$, defined solely over the terminal belief, is the crucial step required to apply the concavification result. This successful *dimensionality reduction* is significant given the dynamic nature of the game.

Unique strategic setting: This problem differs fundamentally from existing multi-shot Bayesian persuasion (BP) extensions. Since the Principal's payoff is tied to the Agent's *random stopping time T^**, and not a pre-determined horizon, the deception problem stands on its own. There is no clear equivalent "one-shot" candidate action in this sequential game, demonstrating that simple concatenations of static BP games are insufficient.

Two-sided endogeneity: Our approach provides a generalized procedure to bring any classic optimal sequential problem into a persuasion-inspired framework. This creates a system of *two-sided endogeneity*: the Agent's adaptive control and stopping are determined by the Principal's signal, and conversely, the Principal's optimal signal is determined by the Agent's best-response sequential strategy.

Challenges and possible research directions. The complexity inherent in our proposed framework, and the limitations of some our current results, open several important avenues for future study.

Complexity of the Countable State Space: The classical BP concavification is typically visualized in a two-dimensional simplex corresponding to binary hypotheses (cf. Kamenica and Gentzkow [7], Tsakas and Tsakas [19]). In our setting, the state space $\Lambda \in \mathbb{N}_0$ (the time of change) is *countably infinite*, meaning the belief μ_t lies in a high-dimensional simplex. This significantly increases the complexity of characterizing the optimal signal distribution (α^*) and computing the concave closure $\hat{\hat{V}}(\mu_0)$. Complexity questions related to the necessary support size of the (approximate) optimal signal are thus pertinent.

Adaptive Thresholds: The Agent's optimal stopping rule is defined by the Shiryaev statistic ρ_t crossing a belief-dependent threshold $B(\mu_t)$. This time-varying and state-dependent optimal boundary is an expected consequence of placing QCPD within a control framework. This outcome is consistent with structural results found in the generalized Wald-Bellman problems, even for basic extensions of passive QCPD, where optimal boundaries cease to be constant. Further numerical simplifications will prove to be handy to reveal structural insights in the result.

General applicability and extensions: Our approach can be used to study strategic information transmission whenever the receiver's problem is a classical optimal sequential problem (e.g., optimal search games or sequential resource allocation) and the sender has a preference over the resulting process or stopping time. Possible future extensions include moving to continuous-time problems like in Koh et al. [8], which would necessitate solving the HJB variational inequality, and exploring frameworks outside the strict Bayesian assumption.

ACKNOWLEDGMENTS

A. J. Budkuley acknowledges several important discussions and suggestions from Priyanshu Gautam and Subodh Kumar (both from IIT Kharagpur). The authors were partially supported by Dept. of Science and Technology, Govt. of India, through grant MTR/2023/001412.

REFERENCES

[1] Francesco Bacchiocchi, Francesco Emanuele Stradi, Matteo Castiglioni, Nicola Gatti, and Alberto Marchesi. 2024. Markov persuasion processes: How to persuade multiple agents from scratch. In *ICML 2024 Workshop: Aligning Reinforcement Learning Experimentalists and Theorists.*

[2] Matteo Castiglioni, Andrea Celli, Alberto Marchesi, and Nicola Gatti. 2020. Online Bayesian persuasion. *Advances in neural information processing systems* 33 (2020), 16188–16198.

[3] Jeffrey C Ely. 2017. Beeps. *American Economic Review* 107, 1 (2017), 31–53.

[4] Priyanshu Gautam, Kumar Subodh, and Amitalok J Budkuley. 2025. On interactive bayesian persuasion under sequential decision making. In *2025 IEEE International Symposium on Information Theory (ISIT).* IEEE, 1–6.

[5] Matthew Gentzkow and Emir Kamenica. 2017. Bayesian persuasion with multiple senders and rich signal spaces. *Games and Economic Behavior* 104 (2017), 411–429.

[6] Emir Kamenica. 2019. Bayesian persuasion and information design. *Annual Review of Economics* 11 (2019), 249–272.

[7] Emir Kamenica and Matthew Gentzkow. 2011. Bayesian persuasion. *American Economic Review* 101, 6 (2011), 2590–2615.

[8] Andrew Koh, Sivakorn Sanguanmoo, and Weijie Zhong. 2024. Persuasion and Optimal Stopping. *arXiv preprint arXiv:2406.12278* (2024).

[9] Maël Le Treust and Tristan Tomala. 2019. Persuasion with limited communication capacity. *Journal of Economic Theory* 184 (2019), 104940.

[10] Fei Li and Peter Norman. 2021. Sequential persuasion. *Theoretical Economics* 16, 2 (2021), 639–675.

[11] Shang Li, Xiaoou Li, Xiaodong Wang, and Jingchen Liu. 2019. Sequential hypothesis test with online usage-constrained sensor selection. *IEEE Transactions on information theory* 65, 7 (2019), 4392–4410.

[12] Naghshvar Mohammad and Tara Javidi. 2013. Active sequential hypothesis testing. *The Annals of Statistics* 41, 6 (2013), 2703–2738.

[13] Anh Nguyen and Teck Yong Tan. 2021. Bayesian persuasion with costly messages. *Journal of Economic Theory* 193 (2021), 105212.

[14] Adam Osękowski. 2025. Bellman: Lecture Notes on Optimal Control. https://www.mimuw.edu.pl/~ados/Bellman/Bellman_2025.pdf.

[15] Goran Peskir and Albert Shiryaev. 2006. *Optimal stopping and free-boundary problems.* Springer.

[16] Albert N Shiryaev. 2008. *Optimal stopping rules.* Springer.

[17] Kumar Subodh, Priyanshu Gautam, and Amitalok J Budkuley. 2025. Probing-based sequential decision making under cognitive limitations. In *2025 IEEE International Symposium on Information Theory (ISIT).* IEEE, 1–6.

[18] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. 2014. *Sequential analysis: Hypothesis testing and changepoint detection.* CRC press.

[19] Elias Tsakas and Nikolas Tsakas. 2021. Noisy persuasion. *Games and Economic Behavior* 130 (2021), 44–61.

[20] Abraham Wald. 1945. Sequential Tests of Statistical Hypotheses. *The Annals of Mathematical Statistics* 16, 2 (1945), 117–186.

[21] Abraham Wald and Jacob Wolfowitz. 1948. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics* (1948), 326–339.

[22] Yun Wang. 2013. Bayesian persuasion with multiple receivers. *Available at SSRN 2625399* (2013).

[23] Jibang Wu, Zixuan Zhang, Zhe Feng, Zhaoran Wang, Zhuoran Yang, Michael I Jordan, and Haifeng Xu. 2022. Sequential Information Design: Markov Persuasion Process and Its Efficient Reinforcement Learning. In *Proceedings of the 23rd ACM Conference on Economics and Computation.* 471–472.