

Quickest Change Detection using Time Inconsistent Anticipatory and Quantum Decision Modeling

Vikram Krishnamurthy, *Fellow IEEE*, Luke Snow, October 5, 2022

Abstract—This paper considers quickest detection scheme where the change in an underlying parameter influencing human decisions is to be detected by only observing the human decisions. Stemming from behavioral economics and mathematical psychology, we propose two generative models for the human decision maker. Namely, we consider an anticipatory decision making model and a quantum decision model. From a decision theoretic point of view, anticipatory models are time inconsistent, meaning that Bellman's principle of optimality does not hold. The appropriate formalism is thus the subgame Nash equilibrium. We show that the interaction between anticipatory agents and sequential quickest detection results in unusual (nonconvex) structure of the quickest change detection policy. In contrast the quantum decision model, despite its mathematical complexity, results in the typical convex quickest detection policy. The optimal quickest detection policy is shown to perform strictly worse than classical quickest detection for both models, via a Blackwell dominance argument. The model and structural results provided contribute to an understanding of the dynamics of human-sensor interfacing.

I. INTRODUCTION

In this paper we construct and analyze a sequential quickest detection framework which aims to detect a change in an underlying state by observing human decisions that are influenced by the state. The problem of 'quickest detection' [1] is fundamental to statistical signal processing and has applications in a wide variety of sectors [2]. Bayesian quickest detection utilizes a prior-posterior updating scheme and an assumed distribution for the change point. In this paper, we consider *Bayesian* quickest detection, in which the observed signals are *human decisions*. This problem of detecting a state change from the observation of human decisions is widespread, and includes contexts such as detecting a market shock by observing individual financial investment decisions, sentiment change through social media monitoring, or adversarial group strategy change through individual decision monitoring. We provide several structural results which characterize the optimal detection performance of the analyst who attempts to detect an underlying state change by observing human decisions, in the context of two specific decision making models: the anticipatory model of [3] and the quantum decision model of [4].

Vikram Krishnamurthy, School of Electrical and Computer Engineering, Cornell University. Email: vikramk@cornell.edu. This research was supported by the U.S. Army Research Office under grant W911NF-19-1-0365, U.S. Air Force Office of Scientific Research under grant FA9550-22-1-0016, and National Science Foundation under grant CCF-2112457.

A. Anticipatory Decision Making

Anticipatory decision making has applications in cyber-physical systems such as human-sensor, human-robot and command-control systems [5]. In behavioral economics, Caplin & Leahy [3] propose a model for anticipatory human decision making via a horizon-2 decision process: the first stage involves choosing an action to minimize an anticipatory psychological reward (involving the probabilities of choosing actions at stage 2), while at the second stage the agent realizes its actual reward. Such anticipatory models mimic important features of human decision making:

- (i) Extensive studies in psychology and neuroscience [6] show that humans are anticipation-driven, and even simple decisions involve sophisticated multi-stage planning.
- (ii) Anticipatory agents act to reduce anxiety. [7] presented experimental results where people chose a larger electric shock than waiting anxiously for a smaller shock.
- (iii) Anticipative agents often deliberately avoid information. [8] reports that giving patients more information before a stressful medical procedure raised their anxiety.

An important aspect of anticipatory decision making is *time-inconsistency*. The dependence of the current reward on future plans results in a deviation between planning and execution. This phenomenon is called time-inconsistency¹ [9] and Bellman's principle of optimality no longer holds. Time inconsistency results in the *planning fallacy* of Kahneman & Tversky [10]: people tend to underestimate the time required to complete a future task. Compared to rational agents, optimistic agents take higher risk of making the wrong decision but have higher anticipatory reward. [11] show that it is optimal for agents with anticipatory reward to take irrational beliefs (referred to as subjective beliefs) deliberately. As will be discussed below, the appropriate concept of optimality for time-inconsistent problems is the *subgame Nash equilibrium*.

B. Quantum Decision Theory

In psychology, Quantum Decision Theory ([12], [13] and references therein) has emerged as a new paradigm which is capable of generalizing current models and accounting for certain violations of axiomatic assumptions. For example, it has been empirically shown that humans routinely violate Savage's 'Sure-Thing Principle' [14], which is equivalent to violation of the law of total probability, and that human decision making is affected by the order

¹In game-theoretic terms, time-inconsistency arises when the optimal policy to the current multi-stage decision problem is sub-game imperfect.

of presentation of information [15] ("order effects"). These violations are natural motivators for treating the decision making agent's mental state as a quantum state in Hilbert Space; The mathematics of quantum probability was developed as an explanation of observed self-interfering and non-commutative behaviors of physical systems, directly analogous to the findings which Quantum Decision Theory (QDT) aims to treat. Indeed, the models of Quantum Decision Theory have been shown to reliably account for violations of the 'Sure-Thing Principle' and order effects [12], while models relying on classical probabilistic structures cannot.

Remark: Quantum Decision Theory does not claim that the human mind is acting in a quantum manner in any physical sense, but simply uses the mathematical structure of quantum probability as a parsimonious generative blackbox model for human decision making.

We utilize the quantum probabilistic model of [4], which provides a generalized decision making process that can account for certain empirically observed decision making phenomena and provides a quantitative way of reasoning about effects of cognitive biases and suboptimality, such as bounded rationality, through a free-parametrization. We have also recently utilized this model within a human-machine assisted decision making scheme, in which a machine provides input signals to a human to dynamically steer the human's decisions towards optimality [16].

C. Quickest Detection Framework

The quickest detection framework of this paper is schematically illustrated in Fig. 1. An underlying state changes at a geometrically distributed unknown time. At each time instant, a sensor obtains a noisy measurement of the underlying state, computes the posterior probability of the state, and provides this information to a human decision maker (e.g. as a recommendation). The human uses this information to choose an action at each time instant according to the anticipatory or quantum decision model. These human decisions are monitored by a Quickest Change Detector, which computes a belief in the underlying state by exploiting knowledge of the anticipatory model structure or quantum decision parameters. Based on the computed belief, the Quickest Detector then decides to continue or declares that a change has occurred, in which case the problem terminates. Note that the quickest detector assumes the decisions are output from *either* the anticipatory or quantum model, and this assumption is fixed for all time. Thus we characterize the structure of the detector's optimal policy under each assumption.

II. ANTICIPATORY DECISION MAKING

This section defines time inconsistent decision problems and reviews the influential behavioral economics model [3] for human decision making with anticipatory feelings. This model will be used in Sec.V to formulate our human sensor interactive quickest change detection problem.

A. Anticipatory Model of Caplin & Leahy [3]

We now review the time inconsistent model for anticipatory human decision making in Caplin & Leahy's paper [3];

see also [17]. A key step in the formulation below is the anticipatory state (2) at time 1 which depends on the probability of future actions (at time 2); this gives the model its anticipatory property.

1) *Anticipatory Model and Time Inconsistency:* The anticipatory decision model in [3] comprises two time steps indexed by $k = 1, 2$. The physical state $s_k \in \mathcal{S}$, $k = 1, 2$, where $\mathcal{S}$ denotes the state space, evolves with Markov transition kernel $p(s_2|s_1)$. Let $a_1 \in \mathcal{A}_1$ and $a_2 \in \mathcal{A}_2$ denote the actions taken by the agent (human) at time 1 and 2. These actions are determined by the non-randomized policies μ_1 and μ_2 where

$$a_1 = \mu_1(s_1), \quad a_2 = \mu_2(s_2, a_1). \quad (1)$$

The first key idea in Caplin & Leahy [3] is to define the anticipatory (psychological) state z_k , $k = 1, 2$:

$$\begin{aligned} z_1 &= \phi(s_1, a_1, \{p(a_2 = a|s_1, a_1, \mu_2), a \in \mathcal{A}_2\}), \\ z_2 &= (s_2, a_2, a_1), \end{aligned} \quad (2)$$

for some pre-defined function ϕ . Note μ_2 is a deterministic function that parametrizes $p(a_2 = a|s_1, a_1, \mu_2)$. In [3], z_k models the human decision maker's state of mind (anxiety). More generally, z_k can model any anticipatory plan, such as for example in situation awareness systems. Note that the anticipatory state z_1 depends on the set of conditional probabilities $\{p(a_2 = a|s_1, a_1, \mu_2), a \in \mathcal{A}_2\}$. These conditional probabilities model anticipation (anxiety)² of the decision maker at time 1 about possible actions it can make at time 2. The anticipation is resolved at time 2 when physical state s_2 is observed and all uncertainty is resolved; hence the anticipatory state z_2 only contains physical state s_2 and realized action a_2 .

The next key idea in [3] is that the anticipatory agent makes decisions by maximizing the 2-stage anticipatory utility

$$\sup_{\mu_1, \mu_2} J(\mu_1, \mu_2) = \mathbb{E}_{\mu_1, \mu_2} \{r_1(z_1) + r_2(z_2)\} \quad (3)$$

Here $r_k(z_k) \in \mathbb{R}$ denote the reward functions. The 2-stage anticipatory utility, called psychological utility in [3], (3) looks just like a standard time separable utility except for the presence of the anxiety term $\{p(a_2 = a|s_1, a_1, \mu_2), a \in \mathcal{A}_2\}$ in $r_1(z_1)$. This μ_2 dependency gives rise to time inconsistency in decision making. As in [3], we assume that the agent knows all the parameters in the above anticipatory model. The key point is that the reward at time 1 depends on the psychological (anticipatory) state which in turn depends on the probability of future actions and states.

²As discussed in [3], introducing anticipatory emotions explains why changing an outcome from zero to a small positive number can have a large effect on anticipation. Human decision makers are sensitive to the possibility rather than probability of negative outcomes [?]. A terrorist attack (unlikely event) worries people a lot more than a car crash (high probability event).

2) *Subgame Perfect Nash Equilibrium*: Caplin & Leahy [3] ‘solve’ the time inconsistent decision problem (3) using the extended Bellman equation, such that the optimal policy at time 2 is defined as:

$$\mu_2^*(s_2, a_1) = \underset{a_2}{\operatorname{argmax}} r_2(s_2, a_2, a_1) \quad (4)$$

Note that by definition (1), μ_2^* depends on a_1 and s_2 .

To specify the optimal policy at time 1, we first introduce the following compact notation. Define

$$\lambda_a \stackrel{\text{defn}}{=} \int_{\mathcal{S}} I(s_2 : \mu_2^*(s_2, a_1) = a) p(s_2|s_1) ds_2, \quad (5)$$

$$\lambda = \{\lambda_a, a \in \mathcal{A}_2\}$$

At time 1, due to time inconsistency, the agent chooses a time consistent policy μ_1^* based on extended Bellman equation [3]:

$$\mu_1^*(s_1) = \underset{a_1}{\operatorname{argmax}} J_1(s_1, a_1, \mu_2^*), \quad (6)$$

$$V_1(s_1) = \max_{a_1} J_1(s_1, a_1, \mu_2^*),$$

$$J_1(s_1, a_1, \mu_2^*) = r_1(\phi(s_1, a_1, \lambda)) + \mathbb{E}\{r_2(s_2, a_2, a_1)|s_1, a_1, \mu_2^*\}$$

$$= r_1(\phi(s_1, a_1, \lambda)) + \int_{\mathcal{S}} r_2(s_2, \mu_2^*(s_2, a_1), a_1) p(s_2|s_1) ds_2$$

Recall $p(s_2|s_1)$ is the transition kernel of the physical state.

(ii) The anticipatory (psychological) state z_1 in (2) consisted of the set of conditional probabilities $\{p(a_2 = a|s_1, a_1, \mu_2), a \in \mathcal{A}_2\}$. More generally, one can formulate the anticipatory state with these conditional probabilities replaced by

$$\{\mathbb{E}\{\Psi(a_2 = a, s_2)|s_1, a_1, \mu_2\}, a \in \mathcal{A}_2\} \quad (7)$$

for some pre-defined function Ψ .

(iii) We mentioned previously that the subgame Nash equilibrium approach to time inconsistency disregards the fact that μ_2^* is no longer optimal at time 1. Another insightful way of viewing this is that since the estimated anticipatory reward $r_1(\phi(s_1, a_1, \lambda))$ requires the agent to extrapolate what might happen at the second stage, plans are not optimal once an action is taken. As an example, people tend to assign higher future workload than what they will actually take on.

B. Characterizing the Nash Equilibrium Policy of Anticipatory Decision Maker

The previous section gave a general setup of the anticipatory decision making model and associated subgame Nash equilibrium policy. However, the Nash equilibrium (6) is the solution of the extended Bellman equation (integral equation) and is difficult to compute in general. In this section, we make specific assumptions on the anticipatory model to give a useful characterization of the Nash equilibrium. Specifically, these assumptions result in a bang-bang and threshold structure for the subgame Nash equilibrium policy (Theorem 1 below).

Bayesian parametrization of transition kernel and reward: Recall r_2 is the reward at time 2; see (2), (3). In the rest of the paper, we will parametrize r_2 and the transition kernel $p(s_2|s_1)$ by a Bayesian parameter. The parametrized reward and transition kernel are constructed as follows: Define the reward $r_2(s_2, a_2, a_1, x)$ and transition kernel $p(s_2|s_1, x)$ which now also depends on a state of nature (ground truth) x . The process $x \in \mathcal{X} = \{1, 2, \dots, m\}$ will be formally defined in Sec.V to model change in quickest detection. Then define the parametrized reward $r_{\eta,2}$ and transition kernel $p_{\eta}(s_2|s_1)$ as

$$r_{\eta,2}(s_2, a_2, a_1) = \sum_{x \in \mathcal{X}} r_2(s_2, a_2, a_1, x) \eta(x)$$

$$p_{\eta}(s_2|s_1) = \sum_{x \in \mathcal{X}} p(s_2|s_1, x) \eta(x) \quad (8)$$

Here η is an m -dimensional Bayesian belief (posterior) vector that lies in the unit $m - 1$ dimensional simplex Π of probability mass functions: $\eta = [\eta(1), \dots, \eta(m)]' \in \Pi$, where

$$\Pi = \{\eta : \eta(i) \in [0, 1], \sum_{i=1}^m \eta(i) = 1\} \quad (9)$$

The posterior η will be formally defined in (18) and appears naturally in the quickest change detection formulation in Sec.V (where the underlying state of nature x jump changes). In this section, η is simply a fixed probability vector in the two-stage anticipatory decision model discussed above.

1) Structural Characterization of Nash equilibrium:

With $r_{\eta,2}$ defined in (8), for notational convenience, define

$$\Delta_{\eta}(s_2, a_1) = r_{\eta,2}(s_2, 2, a_1) - r_{\eta,2}(s_2, 1, a_1) \quad (10)$$

We make the following assumptions on the anticipatory decision model of Sec.II-A:

- (A1) The action spaces are $\mathcal{A}_1 = [0, 1]$, $\mathcal{A}_2 = \{1, 2\}$. Recall the actions $a_1 \in \mathcal{A}_1$ and $a_2 \in \mathcal{A}_2$. The state space is $\mathcal{S} = [0, 1]$. Recall $s_1, s_2 \in \mathcal{S}$.
- (A2) $r_{\eta,2}(s_2, a_2, a_1)$ is convex in a_1 .
- (A3) $\Delta_{\eta}(s_2, a_1)$ defined in (10) is increasing in s_2 . Equivalently, $r_{\eta,2}(s_2, a_2, a_1)$ is supermodular in (s_2, a_2) .
- (A4) The solution $s_2^*(a_1)$ of $\Delta_{\eta}(s_2, a_1) = 0$ exists for $a_1 \in (0, 1)$ and is continuously differentiable on $(0, 1)$.
- (A5) $\frac{\partial \Delta_{\eta}}{\partial a_1} \frac{\partial^2 \Delta_{\eta}}{\partial s_2 \partial a_1} - \frac{\partial \Delta_{\eta}}{\partial s_2} \frac{\partial^2 \Delta_{\eta}}{\partial a_1^2} \geq 0$
- (A6) The anticipatory reward is $r_1(z_1) = \beta z_1$ where $\beta > 0$ and the psychological state (see (7)) is

$$z_1 = \max\{\mathbb{E}\{\Psi(a_2 = a, s_2)|s_1, a_1, \mu_2\}, a \in \mathcal{A}_2\}$$

- (A7) $\Psi(a_2 = 1, s_2) p_{\eta}(s_2|s_1)$ is increasing in s_2
 $\Psi(a_2 = 2, s_2) p_{\eta}(s_2|s_1)$ is decreasing in s_2 .
- (A8) The observation likelihoods $B_{x,y}$ (19) are TP2 (totally positive of order 2); that is, $B_{\bar{x},y} B_{x,\bar{y}} \leq B_{x,y} B_{\bar{x},\bar{y}}$, $\bar{x} > x, \bar{y} > y$.
- (A9) $r_2(s_2, a_2, a_1, x)$ (see (8)) is supermodular in (x, a_2) , i.e., $r_2(s_2, a_2, a_1, \bar{x}) - r_2(s_2, a_2, a_1, x)$ is increasing in a_2 .

The following structural result characterizes the structure of the subgame Nash equilibrium. For subsequent

reference, we will denote the explicit dependence of μ_1^* and μ_2^* on Bayesian parameter η (see (9)) as $\mu_{1,\eta}^*$ and $\mu_{2,\eta}^*$.

Theorem 1. *Consider the anticipatory decision model of Sec.II-A with action and state spaces specified by (A1). Then*

- 1) *Under (A3), (A4), the subgame perfect Nash equilibrium policy μ_2^* specified by (4) has a threshold structure:*

$$\mu_{2,\eta}^*(s_2, a_1) = \begin{cases} 1 & \text{if } s_2 \leq s_{2,\eta}^*(a_1) \\ 2 & \text{if } s_2 > s_{2,\eta}^*(a_1) \end{cases} \quad (11)$$

for some threshold state $s_{2,\eta}^(a_1) \in [0, 1]$ which depends on the Bayesian parameter η .*

- 2) *Under (A4), (A5), threshold state $s_{2,\eta}^*(a_1)$ is convex in a_1 .*
- 3) *Under (A2)-(A7), the utility-to-go $J_1(s, a_1, \mu_2^*)$ defined in (6) is convex in a_1 . Therefore, the subgame Nash equilibrium policy μ_1^* has the following bang-bang³ structure:*

$$\mu_{1,\eta}^*(s_1) = \begin{cases} 1 & \text{if } \beta > \beta^* \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

for some positive constant β^ . (β is defined in (A6).)*

Proof. See Appendix of arXiv paper [18]. $\square$

III. QUANTUM MODEL OF MARTINEZ ET AL [4]

This section presents the open-quantum system model that we will use to represent the decision preference evolution of the human decision maker. We define the evolution of the density operator of the decision maker using the open-system Quantum Lindbladian Equation, proposed in [4] and implemented in [19]. For our purposes we abstract away from the time-evolution such that decisions are made from the *steady-state distribution*, the existence of which is proved in [4]. This steady-state represents the ceasing of any deliberation.

A. Modeling Psychological State via Quantum Probability

Suppose there are n underlying states in the state space $\mathcal{X}$, and A actions in the action space $\mathcal{A}$. For each state $i \in \{1, \dots, n\}$ construct a corresponding unit complex vector $\mathcal{E}_i \in \mathbb{C}^n$ such that $\{\mathcal{E}_i\}_{i=1}^n$ are orthonormal. For each action $i \in \{1, \dots, A\}$, construct a complex vector $a_i \in \mathbb{C}^A$ such that $\{a_i\}_{i=1}^A$ are orthonormal. Denote $\mathcal{H}_{\mathcal{X}} = \text{span}\{\mathcal{E}_1, \dots, \mathcal{E}_n\}$, $\mathcal{H}_{\mathcal{A}} = \text{span}\{a_1, \dots, a_A\}$, and form the tensor product Hilbert space $\mathcal{H} = \mathcal{H}_{\mathcal{X}} \otimes \mathcal{H}_{\mathcal{A}}$. The agent's psychological state is represented by a density operator ρ_t which acts on the Hilbert space $\rho_t : \mathcal{H} \rightarrow \mathcal{H}$. Specifically, $\rho_t = \sum_j p_j |\psi_j\rangle \langle \psi_j|$ with $\sum_j p_j = 1$, $|\psi_j\rangle \in \mathcal{H} \forall j$. The

psychological state ρ_t evolves according to the Lindbladian operator $\mathcal{L}_{(\alpha, \lambda, \phi)}$ by the ordinary differential equation⁴

$$\frac{d\rho_t}{dt} = \mathcal{L}_{(\alpha, \lambda, \phi)} \rho_t \quad (13)$$

where (α, λ, ϕ) are free parameters which govern the evolution, each having a psychological interpretation, see [4]. Implicit in $\mathcal{L}_{(\alpha, \lambda, \phi)}$ is a belief $\eta(x)$ in the underlying state $x \in \{1, \dots, n\}$ and a utility function $u : \mathcal{A} \times \mathcal{X} \rightarrow \mathbb{R}$. The psychological state ρ_t encodes a time dependent probability distribution $\Gamma(a, t)$ over actions $a \in \mathcal{A}$ in the following way. Let P_i be the projector on to the subspace spanned by action vector $a_i \in \mathcal{H}$, then $\Gamma(a_i, t) = \text{Tr}(P_i \rho_t P_i^\dagger)$, where $P_i^\dagger$ is the adjoint of P_i .

For the sake of brevity, we refer the reader's to [4] for the construction of the Lindbladian operator $\mathcal{L}_{(\alpha, \lambda, \phi)}$.

B. Practicality in Human Decision Making

The above quantum model for human decision-making accounts for violations of the sure-thing principle (STP), which we now describe. Suppose there exists an action a and two states $\mathcal{E}_1, \mathcal{E}_2$. Suppose Γ is a non-degenerate posterior belief (strictly in the interior of the unit simplex) of the underlying state. The violation of the sure-thing principle occurs when $P(a|\Gamma)$ is not a convex combination of $P(a|\mathcal{E}_1)$ and $P(a|\mathcal{E}_2)$, i.e.

$$P(a|\Gamma) \neq \epsilon P(a|\mathcal{E}_1) + (1 - \epsilon) P(a|\mathcal{E}_2) \quad \forall \epsilon \in (0, 1)$$

Pothos and Busemeyer [20] (see also [14]) review empirical evidence for the violation of STP and show how quantum models can account for it by introducing quantum interference in the probability evolution. Note that this violation cannot be accounted for by traditional models which rely on classical probability, as the sure-thing principle follows directly as a consequence of the law of total probability.

C. Decision making protocol

Each quantum decision maker (human) in the sequential decision process behaves as follows. The agent has initial psychological state ρ_0 and utility $u : \mathcal{A} \times \mathcal{X} \rightarrow \mathbb{R}$. An underlying state distribution $\eta(x)$ is provided by a Bayesian inference machine (Sensor). u and $\eta(x)$ parameterize $\mathcal{L}_{(\alpha, \lambda, \phi)}$, along with psychological parameters α, λ, ϕ . The psychological state at time t , ρ_t , evolves according to (13) and induces a distribution $\Gamma^\eta(a, t)$ over the action space as

$$\Gamma^\eta(a, t) = \text{Tr}(P_a \rho_t P_a^\dagger) \quad (14)$$

By [4], we are guaranteed the existence of a *steady-state distribution*

$$\Gamma^\eta(a) = \lim_{t \rightarrow \infty} \Gamma^\eta(a, t)$$

We assume action a_n is taken *probabilistically* according to the steady-state distribution $\Gamma^\eta(a)$ which is independent

³The phrase ‘‘bang-bang controller’’ comes from classical optimal control theory. It characterizes a control policy with continuous-valued actions that switches between two extremes.

⁴The reader may be familiar with the Schrödinger equation which governs the time evolution of *closed* quantum systems. The Lindbladian equation is a generalization which governs the time evolution of *open* quantum systems (i.e. those that interact with an external dissipative environment). The recent literature in psychology uses the Lindbladian framework to model human decision making.

from the initial state ρ_0 . This represents the action choice occurring after deliberation has ended, and the steady state is typically reached relatively quickly⁵. We can then abstract away from the time dependence to get the map

$$\mathcal{L}_{(\alpha,\lambda,\phi)} : (\eta(x), u(x, a)) \rightarrow \Gamma^\eta(a) \quad (15)$$

At each discrete time point of the quickest detection protocol, the agent:

- consists of initial psychological state ρ_0 , utility $u : \mathcal{A} \times \mathcal{X} \rightarrow \mathbb{R}$, and parametrization (α, λ, ϕ) . Note that these quantities are time independent and thus constant for all discrete time steps.
- is provided state information in the form of a Bayesian posterior $\eta_n(x)$ by the Sensor.
- deliberates until reaching a steady-state action distribution $\Gamma^{\eta_n}(a)$, from map (15).
- takes action a_n probabilistically from $\Gamma^{\eta_n}(a)$

IV. QUICKEST CHANGE DETECTION WITH ANTICIPATORY AND QUANTUM DECISION MAKERS

We now introduce the quickest change detection protocol and the formulation of an optimal policy for such a protocol. The aim of quickest detection is to determine the jump time τ^0 of the state of nature $\{x_n\}$ i.e., evaluate the optimal stationary policy μ^* of the global decision maker that minimizes the Kolmogorov-Shiryaev criterion for detection of disorder:

$$\begin{aligned} J_{\mu^*}(\pi) &= \inf_{\mu} J_{\mu}(\pi), \\ J_{\mu}(\pi) &= d\mathbb{E}_{\mu}[(\tau - \tau^0)^+] + f\mathbb{P}_{\mu}(\tau < \tau^0) \end{aligned} \quad (16)$$

where $\tau = \inf\{n : u_n = 1\}$ is the time at which the global decision maker announces the change. The parameters d and f specify the delay penalty and false alarm penalty, respectively.

The optimal policy $\mu^*(\pi)$ (16) can be formulated as the solution of a stochastic dynamic programming equation. The quickest detection problem (16) is an example of a stopping-time partially observed Markov decision process (POMDP) with a stationary optimal policy.

We now introduce some notation, then describe the protocol in detail.

- The state of nature $\{x_n \in \{1, 2\}, n \geq 0\}$ models the change event which we aim to detect. x_n starts in state 2 and jumps to state 1 at a geometrically distributed random time τ^0 with $\mathbb{E}[\tau^0] = \frac{1}{1-p}$ for some $p \in [0, 1)$. So, $\{x_n\}$ is a 2-state Markov chain with absorbing transition matrix and initial probability

$$P = \begin{bmatrix} 1 & 0 \\ 1-p & p \end{bmatrix}, \quad \pi_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (17)$$

with change time $\tau^0 = \inf\{n : x_n = 1\}$.

- The (anticipatory / quantum) decision agents act sequentially. A sensor observes the state of nature x_n in noise and computes a Bayesian posterior distribution $\eta(x)$ of the underlying state. This is given to the

human, who then makes a local decision a_n according to

- Anticipatory*: the subgame Nash equilibrium of Theorem 1
- Quantum*: the steady-state action distribution $\Gamma^\eta(a_i)$ induced by the Lindbladian operator $\mathcal{L}_{(\alpha,\lambda,\phi,\eta)}$ and map (15).
- Based on the history of local actions $a_1, \dots, a_n$, the global decision maker chooses action

$$u_n = \{1(\text{stop and announce change}), 2(\text{continue})\}$$

- Define the public belief π_n and private belief η_n at time n as the posterior distributions initialized with $\eta_0 = \pi_0 = [0, 1]'$:

$$\begin{aligned} \pi_n(x) &= \mathbb{P}(x_n = x | a_1, \dots, a_n), x = 1, 2 \\ \eta_n(x) &= \mathbb{P}(x_n = x | a_1, \dots, a_{n-1}, y_n), \end{aligned} \quad (18)$$

where y_n is the private observation recorded by agent n . We have $\pi_n(x), \eta_n(x) \in \Pi$, the unit one-simplex.

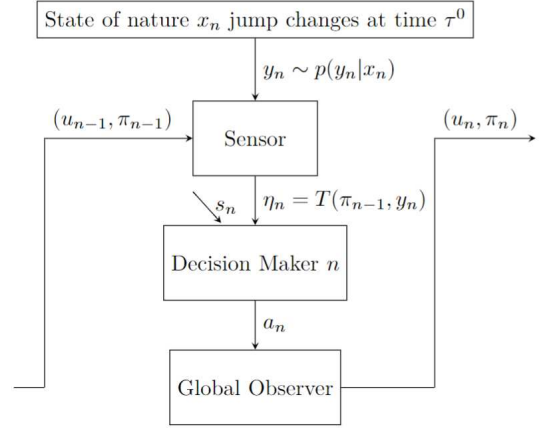


Figure 1: Sequential Quickest Change Detection with decision making agents. The underlying state of nature x_n jump changes at time $\tau^0 \sim \text{Geo}(1-p)$, where p is known. At each time n a sensor observes $y_n \sim P(y|x)$, and the public belief signal π_{n-1} from the previous time point. The sensor outputs a private belief η_n (obtained via Bayesian update) in the underlying state to the decision maker. The decision maker chooses action a_n either by the Anticipatory subgame Nash Equilibrium of Theorem 1 or probabilistically from the distribution $\Gamma^\eta(a)$ induced by the Quantum map (15). The quickest detector sees a_n and outputs its public belief π_n and signal u_n according to (22) and (25). Note that we assume the quickest detector has knowledge of the decision maker's process (either anticipatory or quantum).

A. Change Detection Protocol

We now detail the multi-agent quickest change detection protocol; see also [21].

- Sensor obtains public belief π_{n-1} and signal u_{n-1} from global decision maker. The process only continues if $u_{n-1} = 2$.
- Let $\mathcal{Y}$ denote the observation space. The sensor records noisy observation $y_n \in \mathcal{Y}$ of state of nature x_n with conditional density

$$B_{x,y} = p(y_n = y | x_n = x) \quad (19)$$

- Private Belief*. The sensor evaluates the Bayesian private belief

⁵See [4] for a proof of the steady state and a discussion of relaxation times of this evolution

$$\eta_n = T(\pi_{n-1}, y_n), \quad T(\pi, y) = \frac{B_y P' \pi}{\sigma(\pi, y)}, \quad (20)$$

$$\sigma(\pi, y) = \mathbf{1}' B_y P' \pi, \quad B_y = \text{diag}(B_{1,y} B_{2,y}) \quad (21)$$

and feeds this to the human agent.

- iv) *Local decision.* State posterior η_n and physical state s_n are observed by the agent.
- a) *Anticipatory:* The agent's private belief η_n affects its reward and transition kernel of physical state process $\{s_n, n \geq 1\}$ as in (8):

$$\begin{aligned} r_{\eta,2}(s_2, a_2, a_1) &= \sum_{x \in \mathcal{X}} r_2(s_2, a_2, a_1, x) \eta(x) \\ p_{\eta}(s_2|s_1) &= \sum_{x \in \mathcal{X}} p(s_2|s_1, x) \eta(x) \end{aligned} \quad (8 \text{ repeated})$$

The agent uses η_n, s_n to make anticipatory decisions $a_n = (a_{n,1}, a_{n,2})$ via (12), (11) in Theorem 1.

- b) *Quantum:* The agent's private belief η_n parameterizes the Lindbladian operator $\mathcal{L}_{(\alpha, \lambda, \phi)}$. This induces a steady-state action probability distribution $\Gamma^n(a_i)$ via the map (15), and the action a_n is taken probabilistically according to Γ^n .
- v) *Quickest Detector.* Based on the decisions a_n of local decision maker n , the quickest detector:
 - i) Updates the public belief from π_{n-1} to π_n as

$$\pi_n = \bar{T}(\pi_{n-1}, a_n, s_n) \quad (22)$$

$$\bar{T}(\pi, a, s) = \frac{R_a^\pi(s) P' \pi}{\bar{\sigma}(\pi, a, s)}, \quad \bar{\sigma}(\pi, a, s) = \mathbf{1}' R_a^\pi(s) P' \pi$$

$$\text{where } R_a^\pi(s) = \text{diag}(R_{1,a}^\pi(s), R_{2,a}^\pi(s)),$$

$$R_{x,a_n}^\pi(s) = \mathbb{P}(a_n = a | x_n = x, \pi_{n-1}, s_n = s) \quad (23)$$

The action probabilities $R_{x,a}^\pi$ are computed as

$$R_{x,a}^\pi(s) = \int_{\mathcal{Y}} P(a|\pi, s, y) B_{x,y} dy \quad (24)$$

where $P(a|\pi, s, y) = I(\mu_{2,T(\pi,y)}^*(s, a_{n,1}) = a_{n,2})$ when the anticipatory decision model is assumed, and $P(a|\pi, s, y) = \bar{\Gamma}_y^\pi(a)$ when the quantum decision model is assumed. Recall $a_n = (a_{n,1}, a_{n,2})$ and $\mu_{2,\eta}^*$ is the anticipatory decision maker's subgame Nash equilibrium policy (11), and $\bar{\Gamma}_y^\pi$ is the quantum decision maker's induced action distribution (15) given public belief π and observation y . Specifically, $\bar{\Gamma}_y^\pi$ is the output of the map (15), with input $\eta(x) = T(\pi, y), u(x, a)$ and parametrization (α, λ, ϕ) . Note that in the quantum decision-making protocol, the action a_n is independent from the state s_n , so likewise the value of $R_{x,a}^\pi(s)$, and therefore $\bar{T}(\pi, a, s)$, does not depend on s .

- ii) Chooses global action u_n using optimal policy μ^* :

$$u_n = \mu^*(\pi_n) \in \{1(\text{stop}), 2(\text{continue})\}. \quad (25)$$

- iii) Is $u_n = 2$, then set n to $n + 1$ and go to step 1. If $u_n = 1$, then stop and announce change.

B. Quickest Detector Optimal Policy [17]

Considering the aim of quickest detection, characterized by (16), we now outline the details of the optimal policy stochastic dynamic programming formulation.

- 1) *Costs:* To present the dynamic programming equation we first formulate the false alarm and delay costs (16) incurred by the global decision maker in terms of the public belief.

- i) *False alarm penalty:* If global decision $u_n = 1$ (stop) is chosen before the change point τ^0 , then a false alarm penalty is incurred. The false alarm event $\{x_n = 2, u_n = 1\}$ represents the event that a change is announced before the change happens at time τ^0 . Recall (17) the jump change occurs at time τ^0 from state 2 to state 1. Then recalling $f \geq 0$ is the false alarm penalty in (16), the expected false alarm penalty is

$$\begin{aligned} f \mathbb{P}_\mu(\tau < \tau^0) &= f \mathbb{E}_\mu\{\mathbb{E}[I(x_n = 2, u_n = 1) | \mathcal{G}_n]\} \\ \mathcal{G}_n &= \sigma\text{-algebra generated by } (a_1, \dots, a_n) \end{aligned} \quad (26)$$

Clearly $\mathbb{E}[I(x_n = 2, u_n = 1) | \mathcal{G}_n]$ can be expressed in terms of the public belief $\pi_n(2) = P(x_n = 2 | a_1, \dots, a_n)$ as

$$C(\pi_n, u_n = 1) = f e_2' \pi_n, \quad \text{where } e_2 = [0, 1]' \quad (27)$$

- ii) *Delay cost of continuing:* If global decision $u_n = 2$ is taken then Protocol 1 continues to the next time. A delay cost is incurred when the event $\{x_n = 1, u_n = 2\}$ occurs, i.e. no change is declared at time n . The expected delay cost is $d \mathbb{E}[I(X_n = 1, u_n = 2) | \mathcal{G}_n]$ where $d > 0$ denotes the delay cost. In terms of the public belief, the delay cost is

$$C(\pi_n, u_n = 2) = d e_1' \pi_n, \quad \text{where } e_1 = [1, 0]' \quad (28)$$

We can re-express Kolmogorov-Shiryaev criterion (16) as

$$J_\mu = \mathbb{E}_\mu\left\{\sum_{n=0}^{\tau-1} C(\pi_n, 2) + C(\pi_\tau, 1)\right\} \quad (29)$$

where $\tau = \inf\{n : u_n = 1\}$ is adapted to the σ -algebra $\mathcal{G}_n$. Since $C(\pi, 1), C(\pi, 2)$ are non-negative and bounded for all $\pi \in \Pi$, stopping is guaranteed in finite time.

- 2) *Bellman's equation for Quickest Detection Policy* Consider the costs (27), (28) defined in terms of the public belief π . Then the optimal stationary policy $\phi^*(\pi, s)$ defined in (25), (16). and associated value function $V(\pi, s)$ are the solution of Bellman's dynamic programming functional equation [21]

$$\begin{aligned} Q(\pi, s, 1) &\stackrel{\text{defn}}{=} C(\pi, 1), \\ Q(\pi, s, 2) &\stackrel{\text{defn}}{=} C(\pi, 2) \\ &+ \int_{\mathcal{S}} \sum_{a \in \mathcal{A}_1 \times \mathcal{A}_2} \mathcal{V}(\bar{T}(\pi, a, \bar{s}), \bar{s}) \bar{\sigma}(\pi, a, \bar{s}) p(\bar{s} | s) d\bar{s} \\ \phi^*(\pi, s) &= \arg \min\{Q(\pi, s, 1), Q(\pi, s, 2)\}, \\ \mathcal{V}(\pi, s) &= \min\{Q(\pi, s, 1), Q(\pi, s, 2)\} = J_\phi^*(\pi, s) \end{aligned} \quad (30)$$

The public belief update $\bar{T}$ and normalization measure $\bar{\sigma}$ were defined in (22). Recall (25) that $u_n = \phi^*(\pi_n, s_n)$ is the global decision maker's action whether to continue or stop. Note that in the case of the quantum decision protocol, since $\bar{T}(\pi, a, s)$ is independent from s , the value function $V(\pi, s)$ and associated terms in (30) have no dependence on s . The goal of the global decision-maker is to solve for the optimal quickest change policy ϕ^* in (30) or equivalently, determine the optimal stopping set $\mathcal{S}$

$$\mathcal{S} = \{\pi, s : \phi^*(\pi, s) = 1\} = \{\pi, s : Q(\pi, s, 1) \leq Q(\pi, s, 2)\} \quad (31)$$

3) Value Iteration Algorithm

The optimal policy $\phi^*(\pi, s)$ and value function $\mathcal{V}(\pi, s)$ can be constructed as the solution of a fixed point iteration of Bellman's equation (30) – the resulting algorithm is called the value iteration algorithm. The value iteration algorithm proceeds as follows: Initialize $\mathcal{V}_0(\pi, s) = 0$ and for iterations $k = 1, 2, \dots$

$$\begin{aligned} \mathcal{V}_{k+1}(\pi, s) &= \min_{u \in \mathcal{U}} Q_{k+1}(\pi, s, u), \\ \phi_{k+1}^*(\pi, s) &= \arg\min_{u \in \mathcal{U}} Q_{k+1}(\pi, s, u) \quad \pi \in \Pi, \\ Q_{k+1}(\pi, s, 1) &= C(\pi, 1), \quad Q_{k+1}(\pi, s, 2) = C(\pi, 2) \\ &+ \int_{\mathcal{S}} \sum_{a \in \mathcal{A}_1 \times \mathcal{A}_2} \mathcal{V}_k(\bar{T}(\pi, \bar{s}, a), \bar{s}) \bar{\sigma}(\pi, \bar{s}, a) p(\bar{s}|s) d\bar{s}, \end{aligned} \quad (32)$$

Let $\mathcal{B}$ denote the set of bounded real-valued functions on Π . For any $\mathcal{V}, \tilde{\mathcal{V}} \in \mathcal{B}$ and $\pi \in \Pi$, define the sup-norm metric $\sup \|\mathcal{V}(\pi, s) - \tilde{\mathcal{V}}(\pi, s)\|$, $s \in \mathcal{S}$. Since $C(\pi, 1)$, $C(\pi, 2)$, $\pi \in \Pi$, are bounded, the value iteration algorithm (32) generates a sequence of lower semi-continuous value functions $\{\mathcal{V}_k\} \subset \mathcal{B}$ that converge pointwise as $k \rightarrow \infty$ to $\mathcal{V}(\pi, s) \in \mathcal{B}$, the solution of Bellman's equation [22].

V. CHARACTERIZING THE STRUCTURE OF THE QUICKEST DETECTOR

In this section we analyze several structural properties of the quickest detection protocol detailed in Sec. IV. Our results in this section are structured as follows: In Section V-A we prove that the optimal policy (30) has a single threshold structure for the quantum decision model and that the structure for the anticipatory model may exhibit multiple thresholds. In Section V-B we provide a lower bound on the optimal cost incurred by the quickest detector via the policy of Sec. IV. Specifically, this lower bound is given by the optimal cost incurred within the classical quickest change detection protocol, i.e. without intermediate human decisions.

A. Threshold Structure of Optimal Policy

1) *Anticipatory Decision Model:* We assume in this section that the observation space and action space of the anticipatory agent are $\mathcal{Y} = \{1, \dots, Y\}$, $\mathcal{A}_2 = \{1, 2\}$. The purpose of this section is to show that even though the

public belief $\pi \in \Pi$ is continuum, there are only $Y + 1$ possible distinct action likelihood probability matrices.

Specifically, define the following Y points in the one-dimensional simplex Π :

$$\pi_y^* = \{\pi : (r_1 - r_2)' B_y P' \pi = 0\}, \quad y = 1, \dots, Y$$

Note that $\pi_y^* = [1 - \pi_y^*(2), \pi_y^*(2)]'$ depends on a_1, s .

Theorem 2. Under (A8), (A9), it follows that

$$\pi_1^*(2) \leq \pi_2^*(2) \leq \dots \leq \pi_Y^*(2) \quad (33)$$

Thus the belief space $\Pi = [0, 1]$ can be partitioned into at most $Y + 1$ non empty intervals denoted $\mathcal{P}_1, \dots, \mathcal{P}_{Y+1}$ where

$$\mathcal{P}_1 = [0, \pi_1^*(2)], \mathcal{P}_2 = (\pi_1^*(2), \pi_2^*(2)], \dots, \mathcal{P}_{Y+1} = (\pi_Y^*(2), 1] \quad (34)$$

On each such interval, the action likelihood R^π (23) is a constant with respect to belief π . Specifically, for fixed a_1, s

$$R^\pi(s) = \begin{bmatrix} \sum_{i=0}^{l-1} B_{1i} & \sum_{i=l}^Y B_{1i} \\ \sum_{i=0}^{l-1} B_{2i} & \sum_{i=l}^Y B_{2i} \end{bmatrix}, \quad \pi \in \mathcal{P}_l \quad (35)$$

Proof. See Appendix of arXiv paper [18]. $\square$

The following Theorem uses this result to show that the optimal detection policy given the anticipatory decision model need not consist of only a single threshold. This is similar to the multi-threshold policies in [23].

Theorem 3. Given the anticipatory decision making quickest change detection protocol detailed in Section II-A, it is not guaranteed that the quickest detector's optimal policy μ^* (16) may not exhibit a single threshold state π' such that

$$\mu^*(\pi) = \begin{cases} 2, & \pi < \pi' \\ 1, & \pi \geq \pi' \end{cases}$$

Proof. See Appendix of arXiv paper [18]. $\square$

2) *Quantum Decision Model:* We will show that, given the quantum decision making quickest change detection protocol detailed in Section IV-A, the quickest detector's optimal policy (16) exhibits a single-threshold behavior. In contrast, the optimal policy induced by anticipatory decision making can exhibit multiple thresholds

Theorem 4. Given the quantum decision making quickest change detection protocol detailed in Section III-C, the quickest detector's optimal policy μ^* (16) exhibits a single threshold state π' such that

$$\mu^*(\pi) = \begin{cases} 2, & \pi < \pi' \\ 1, & \pi \geq \pi' \end{cases}$$

Proof. See Appendix of arXiv paper [18]. $\square$

B. Lower Bound on Performance

First define the optimal policy and cost in classical quickest change detection. Similar to (30), the optimal policy $\underline{\phi}^*(\pi)$ and cost $\underline{\mathcal{V}}(\pi)$ incurred in classical quickest detection, satisfy the following stochastic dynamic programming equation:

$$\underline{\phi}^*(\pi) = \arg \min_{u \in \mathcal{U}} \underline{Q}(\pi, u), \quad \underline{\mathcal{V}}(\pi) = \min_{u \in \mathcal{U}} \underline{Q}(\pi, u), \quad (36)$$

$$\text{where } \underline{Q}(\pi, 2) = C(\pi, 2) + \sum_{y \in \mathcal{Y}} \underline{\mathcal{V}}(T(\pi, y)) \sigma(\pi, y),$$

$$\underline{Q}(\pi, 1) = C(\pi, 1), \quad \underline{J}_{\mu^*}(\pi) = \underline{\mathcal{V}}(\pi).$$

Here $T(\pi, y)$ is the Bayesian filter update defined in (20) and $\underline{J}_{\mu^*}(\pi)$ is the cumulative cost of the optimal policy starting with initial belief π . Note that unlike Protocol 1, in classical quickest detection, there is no public belief update (22) or interaction between the public and private beliefs.

The following theorem says that for any initial belief π , the optimal detection policy with either anticipative or quantum-decision agents acting sequentially (Protocol 1) incurs a higher cumulative cost than that of classical quickest detection.

Theorem 5. *Consider the quickest change detection problem involving either anticipatory or quantum-decision agents described in Protocol 1 and associated value function $\mathcal{V}(\pi, s)$ in (30). Consider also the classical quickest detection problem with value function $\underline{\mathcal{V}}(\pi)$ in (36). Then for any initial belief $\pi \in \Pi$, the optimal cost incurred by classical quickest detection is smaller than that of quickest detection with anticipatory agents. That is, $\underline{\mathcal{V}}(\pi) \leq \mathcal{V}(\pi, s)$ for all $\pi \in \Pi, s \in \mathcal{S}$.*

Proof. See Appendix of arXiv paper [18]. $\square$

VI. CONCLUSION

In this work we have presented the framework of quickest change detection using human decisions, incorporating a time-inconsistent anticipatory model and a quantum decision model. The anticipatory model accounts for a decision-maker's tendency to plan for future events while making current decisions, and the quantum model utilizes the generalized structure of quantum probability to account for a wider range of human decision-making phenomena. Structural Results for the detector's optimal policy were obtained, which reveal that the anticipatory model can result in a counter-intuitive non-convex optimal stopping region. In contrast, while the quantum decision model is mathematically more complex, it results in the standard convex (single-threshold) optimal stopping region. We also show that the optimal detection performance (given by the value function) with both models is lower bounded by that of classical quickest detection, indicating that the intermediate human decisions strictly hinder detection performance.

REFERENCES

- [1] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.
- [2] H. V. Poor and O. Hadjiladis, *Quickest Detection*. Cambridge University Press, 2008.
- [3] A. Caplin and J. Leahy, "Psychological expected utility theory and anticipatory feelings," *The Quarterly Journal of Economics*, vol. 116, no. 1, pp. 55–79, 2001.
- [4] I. Martínez-Martínez and E. Sánchez-Burillo, "Quantum stochastic walks on networks for decision-making," *Scientific reports*, vol. 6, no. 1, pp. 1–13, 2016.
- [5] R. Rosen, "Anticipatory systems," in *Anticipatory systems*. Springer, 2012, pp. 313–370.
- [6] R. Bénabou and J. Tirole, "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, vol. 30, no. 3, pp. 141–64, 2016.
- [7] J. O. Cook and L. W. Barnes Jr, "Choice of delay of inevitable shock," *The Journal of Abnormal and Social Psychology*, vol. 68, no. 6, p. 669, 1964.
- [8] S. M. Miller and C. E. Mangan, "Interacting effects of information and coping style in adapting to gynecologic stress: should the doctor tell all?" *Journal of personality and social psychology*, vol. 45, no. 1, p. 223, 1983.
- [9] T. Björk and A. Murgoci, "A theory of markovian time-inconsistent stochastic control in discrete time," *Finance and Stochastics*, vol. 18, no. 3, pp. 545–592, 2014.
- [10] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 2013, pp. 99–127.
- [11] M. K. Brunnermeier and J. A. Parker, "Optimal expectations," *American Economic Review*, vol. 95, no. 4, pp. 1092–1118, 2005.
- [12] J. Busemeyer and P. Bruza, *Quantum Models of Cognition and Decision*, ser. Quantum Models of Cognition and Decision. Cambridge University Press, 2012. [Online]. Available: https://books.google.com/books?id=0vxvhTG_ZLAC
- [13] V. I. Yukalov and D. Sornette, "Mathematical structure of quantum decision theory," *Advances in Complex Systems*, vol. 13, no. 05, pp. 659–698, 2010.
- [14] A. Y. Khrennikov and E. Haven, "Quantum mechanics and violations of the sure-thing principle: The use of probability interference and other concepts," *Journal of Mathematical Psychology*, vol. 53, no. 5, pp. 378–388, 2009.
- [15] J. S. Trueblood and J. R. Busemeyer, "A quantum probability account of order effects in inference," *Cognitive science*, vol. 35, no. 8, pp. 1518–1552, 2011.
- [16] L. Snow, S. Jain, and V. Krishnamurthy, "Lyapunov based stochastic stability of a quantum decision system for human-machine interaction," 2022.
- [17] V. Krishnamurthy, "Quickest change detection of time inconsistent anticipatory agents. human-sensor and cyber-physical systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1054–1069, 2021.
- [18] V. Krishnamurthy and L. Snow, "Quickest change detection using time inconsistent anticipatory and quantum decision modeling," 2022.
- [19] J. Busemeyer, Q. Zhang, S. Balakrishnan, and Z. Wang, "Application of quantum—Markov open system models to human cognition and decision," *Entropy*, vol. 22, no. 9, p. 990, 2020.
- [20] E. M. Pothos and J. R. Busemeyer, "A quantum probability explanation for violations of 'rational' decision theory," *Proceedings of the Royal Society B: Biological Sciences*, vol. 276, no. 1665, pp. 2171–2178, 2009.
- [21] V. Krishnamurthy, *Partially observed Markov decision processes*. Cambridge University Press, 2016.
- [22] O. Hernández-Lerma and J. B. Lasserre, *Discrete-time Markov control processes: basic optimality criteria*. Springer Science & Business Media, 2012, vol. 30.
- [23] V. Krishnamurthy, "Quickest detection POMDPs with social learning: Interaction of local and global decision makers," *IEEE Transactions Information Theory*, vol. 58, no. 8, pp. 5563–5587, 2012.