Quickest Change Detection of Time Inconsistent Anticipatory Agents. Human-Sensor and Cyber-Physical Systems

Vikram Krishnamurthy, *Fellow, IEEE*

***Abstract*—In behavioral economics, human decision makers are modeled as anticipatory agents that make decisions by taking into account the probability of future decisions (plans). We consider cyber-physical systems involving the interaction between anticipatory agents and statistical detection. A sensing device records the decisions of an anticipatory agent. Given these decisions, how can the sensing device achieve quickest detection of a change in the anticipatory system? 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 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. Our methodology yields a useful framework for situation awareness systems and anticipatory human decision makers interacting with sequential detectors.**

***Index Terms*—Time inconsistency, anticipatory decision making, subgame Nash equilibrium, quickest change detection, change blindness, Blackwell dominance, multi-threshold policy.**

NOMENCLATURE
GLOSSARY OF SYMBOLS

Anticipatory agent. Sec. II-B and III

s_1, s_2	physical state
z_1, z_2	psychological state (5), (11)
a_1, a_2	actions (4)
μ_1^*, μ_2^*	Nash equilibrium policy (9), (8)
$V_1(\cdot), V_2(\cdot)$	value function

Quickest detection. Sec. IV

n	discrete time n (also agent n)
x_n	jump state (for quickest detection)
P	transition matrix of $\{x_n, n \geq 0\}$ (22)
f, d	false alarm and delay penalty parameters

Anticipatory agents acting sequentially. Sec. IV

s_n	physical state
z_n	psychological state
a_{n_1}, a_{n_2}	local decision maker's actions
η_n	private belief of local decision maker n (23)
$\mu_{n,1}^*, \mu_{n,2}^*$	Nash equilibrium policy (9), (8)
y_n	private observation of x_n at time n
B_{x_n,y_n}	observation likelihood $p(y_n\|x_n)$ (24)
$T(\pi, y)$	private belief update (25)
$\sigma(\pi, y)$	normalization measure for private belief

Global Decision maker. Sec. IV and Sec. V

u_n	action at time $n \in \{1(\text{stop}), 2(\text{cont})\}$
$\phi^*(\pi, s)$	optimal policy for quickest detection
π_n	public belief at n (23)
$R_{x,a}^{\pi}(s)$	action likelihood $p(a\|x, \pi, s)$ (27), (28)
$\bar{T}(\pi, a, s)$	public belief update (26)
$\bar{\sigma}(\pi, a, s)$	normalization measure for public belief
$\mathcal{V}(\pi, s)$	value function
$C(\pi, u)$	costs incurred in quickest detection

Manuscript received May 14, 2020; revised October 13, 2020 and December 7, 2020; accepted January 1, 2021. Date of publication January 12, 2021; date of current version February 9, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Vasanthan Raghavan. This work was supported by the U.S. Army Research Office under Grant W911NF-19-1-0365, in part by Air Force Office of Scientific Researcher under Grant FA9550-18-1-0007, and in part by National Science Foundation under Grant CCF-1714180. A short version of some of the results in this paper has been submitted to ICASSP 2021.

The author is with the School of Electrical, and Computer Engineering, Cornell University, Ithaca, New York 14853-2201 USA (e-mail: vikramk@cornell.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TSP.2021.3050977, provided by the authors.

Digital Object Identifier 10.1109/TSP.2021.3050977

I. INTRODUCTION

'COGNITIVE SENSING' is widely used in signal processing, but lacks the important property of anticipatory decision making. *An anticipatory agent makes decisions by taking to account the probability of future decisions*. This crucial property is studied in behavioral economics involving human decision makers and yields remarkable behavior such as time inconsistency as discussed below.

This paper is an early step in understanding the interaction between statistical detection and behavioral economics models. Signal processing and behavioral economics are mature areas; yet their intersection, namely cyber-physical systems involving interaction of human decision makers with sensing based detection is relatively unexplored. The main question we address is: *If multiple anticipatory decision makers interact sequentially (or a single anticipatory agent acts repeatedly), how can a global decision maker use these anticipatory decisions to achieve optimal sequential change detection?* Fig. 1 shows our schematic setup. Anticipatory agents can mimic either strategic human decision makers [1] or an automated command-control decision

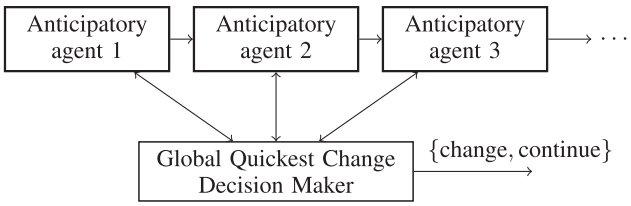


Fig. 1. Quickest Change Detection Problem involving a single anticipatory agent acting repeatedly (or multiple anticipatory agents acting sequentially) and a global decision maker. The anticipatory model for individual decision makers is discussed in Sec. II-B and Sec. III and results in time inconsistent decision making. The interaction of the agents with a global decision maker to achieve quickest detection is detailed in Sec. IV.

system [2]. The anticipatory agents act sequentially and are affected by the decisions of previous agents. A global decision maker monitors the decisions of these anticipatory agents. *How can the global decision maker use the local decisions from these anticipatory agents to perform quickest change detection?*

A. Anticipatory Decision Making

Anticipatory decision making has applications in cyber-physical systems such as human-sensor, human-robot and command-control systems [2]. Here are two applications.

1) *Human decision makers:* In behavioral economics, Caplin & Leahy [1] propose a remarkable 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, neuroscience [4], [5] show that humans are anticipation-driven, and even simple decisions involve sophisticated multi-stage planning.

ii) Anticipatory agents act to reduce anxiety. [6] presented experimental results where people chose a larger electric shock than waiting anxiously for a smaller shock.

iii) Anticipative agents often deliberately avoid information. [7] reports that giving patients more information before a stressful medical procedure raised their anxiety.

2) *Level 3 Situation Awareness:* In defense command-control systems, Level 3 Situation Awareness (SA) [8] involves the ability to project implications of future actions (plans). Level 3 SA [9] is achieved through knowledge of Levels 1 and 2 SA, and then extrapolating this information forward in time (as an anticipatory reward involving probabilities of future actions) to determine how it will affect future decisions/plans [8]. Prediction is concerned with guessing future states based on extensive training; in contrast, anticipatory decision making [10] involves preparing to respond to previously unseen scenarios. [11] shows that many command and control systems overestimate their ability to react.

B. Anticipatory Decision Making Yields Time Inconsistency

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-inconsistency[1] [12] and Bellman's principle of optimality no longer holds. Time inconsistency results in the *planning fallacy* of Kahneman & Tversky [13]: 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. [14] show that it is optimal for agents with anticipatory reward to take irrational beliefs (referred to as subjective beliefs) deliberately. This explains the optimistic planning fallacy, in which people tend to overestimate future rewards. As will be discussed below, the appropriate concept of optimality for time-inconsistent problems is the *subgame Nash equilibrium*.

C. Quickest Detection With Anticipatory Agents

Having motivated anticipatory decision making, we turn to the second main idea of the paper, namely, *Bayesian quickest change detection by a global decision maker which uses the decisions of anticipatory agents* (local decision makers); see Fig. 1. In Bayesian quickest detection, the change time is specified by a prior [15], [16]. Classical quickest detection deals with detecting a change in an underlying state given noisy observations. This paper considers a substantial generalization: *How to detect a change in the strategy of anticipatory agents when they interact sequentially?* Unlike classical quickest detection, we only have access to the actions of the agents from their sub-game Nash equilibrium. As a result, the Bayesian belief (posterior) update structure is much more complex than classical quickest detection. This causes remarkably counter-intuitive behavior as we will investigate below.

We start by outlining important applications that motivate the quickest detection problem with anticipatory agents.

The first class of examples involve social media based accommodation systems such as Airbnb. Individuals with anticipatory feelings make decisions whether to rent a property; these decisions are affected by the reviews (decisions) of previous agents. A global decision maker (e.g. Airbnb) monitors these local decisions. *How can the global decision maker detect if there is a sudden change in the demand for a specific accommodation due to the presence of a new competitor?* The supplementary document discusses this example in detail.

A related example arises in the measurement of the adoption of a new product using a micro-blogging platform like Twitter. The adoption of the technology diffuses through the market but its effects can only be observed through the tweets of select individuals of the population. These selected individuals interact and learn from the decisions (tweeted sentiments) of other members. Suppose the state of nature suddenly changes due to a sudden market shock or presence of a new competitor. The goal for a market analyst is to detect this change.

The second class of examples involves anticipatory situation awareness (SA) in a *team* setting [17]. For example, [18] introduced a situational adapting system to assess team SA for fighter pilots based on information fusion. Suppose individual SA systems monitor an enemy target or enemy radar (state). Given noisy measurements of the state, each SA system (equipped with

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

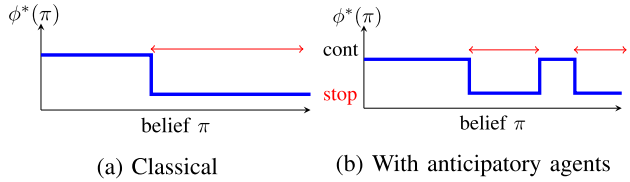
a Bayesian tracker) makes decisions about the threat and relays these decisions to subsequent SA systems in the team. A global decision maker (supervisory system) monitors these decisions to assess overall threat level. How can the global decision maker detect a sudden change in the threat? Such a change is reflective of the enemy target making purposeful maneuvers; or the enemy radar switching modes between search, acquisition or track.

The third example involves human-sensor interface systems, where anticipatory human decision makers are equipped with sensing/computing devices. The sensing device observes the state in noise. The computing device evaluates the posterior distribution and provides the agent with these probabilities. The agent (human) then makes anticipatory decisions. The aim is to devise a change detection algorithm that compensates for the anticipatory human decision maker. Such schemes are studied extensively in situation assessment of pilots [19] and validated based on simulations involving pilots performing a landing approach into an airport. Other examples include assistive care for the dementia [20] where a machine monitors human decisions (activities) for changes in routine behavior indicating sudden onset of memory impairment.

D. Main Results

Sec. II reviews time inconsistent sequential decision problems and the framework for anticipatory decision making as a 2-stage stochastic optimization problem. Due to the time inconsistency of the decision problem, the appropriate notion of optimality is the subgame Nash equilibrium policy. In Sec. III, our main contribution is to introduce sufficient conditions on the anticipatory model so that the Nash equilibrium has a useful structure; see Theorem 1. This structure reveals several interesting features about anticipatory decision making. Sec. IV formulates the quickest change detection protocol involving multiple anticipatory agents where a global decision maker uses the decisions of the anticipatory decision makers to decide if a state has changed. The optimal policy that minimizes the Kolmogorov-Shiryaev criterion is formulated as the solution of a stochastic dynamic programming problem. Then Sec. V characterizes the structure of the Bayesian belief updates and achievable cost of the quickest detector without brute force computations. It derives important structural properties of the Bayesian updates of the local and global decision makers (Theorem 2 and Theorem 3), constructs a lower bound for the optimal cost incurred using Blackwell dominance (Theorem 4), and presents numerical examples of the unusual structure of the optimal quickest change policy (non-convex stopping region) and non-concave value function.

In classical quickest detection [3], [15], the optimal policy has a threshold structure: when the posterior probability of change exceeds a threshold, it is optimal to declare a change; see Fig. 2. The stopping set (set of posteriors probabilities where it is optimal to declare "change") is convex. In quickest detection involving a global decision maker interacting with anticipatory agents (this paper), the remarkable feature is that the stopping set is disconnected, see Fig. 2. One sees the counter-intuitive property: the optimal detection policy switches from announce "change" to announce "no change" as the posterior probability of a change increases! Thus making a global decision as to whether a change has occurred based on local decisions of interacting agents is non-trivial.



(a) Classical (b) With anticipatory agents

Fig. 2. Optimal Quickest Change Detection Policy ϕ^* as a function of Bayesian belief π. In classical quickest detection, the stopping set is convex (connected). In comparison, for quickest detection with anticipatory agents (this paper), the stopping set is nonconvex (disconnected) as indicated in red.

E. Perspective on Main Results

1) Social Learning. The anticipatory model used in this paper is from [1]; see also [14], [21]. This generalizes classical social learning models that have been studied extensively in sociology, economics and signal processing [22]–[25]. Classical social learning assumes that agents make one-shot (myopic) decisions to maximize their expected utility. The behavioral economics models considered here are useful generalizations of social learning since they involve multi-stage planning; as mentioned earlier, even simple human decisions involve multi-stage planning with time-inconsistency.

Our sequential framework of multiple decision makers is similar to team decision theory [26], [27]; the key difference being time inconsistency.

This paper differs from [23] where quickest detection was considered with myopic social learning based local decisions. Motivated by behavioral economics [1], we consider a 2-stage decision framework for each local decision maker that is more general than myopic social learning. This 2-stage framework captures several salient features of human decision-making including anticipation, time inconsistency and deliberate avoidance of information. Also, in our quickest detection formulation, the jump change affects both the rewards of the agents and the transition kernel of the physical state (in the myopic case [23], there is no transition mechanism).

2) How un-informed local decision makers affect global decision making? In order to optimize its change detection policy, the global decision maker must interpret decisions of the local decision makers, knowing that the local decision makers are anticipatory and that they use decisions from previous agents. A well known characteristic of this sequential multi-agent framework is that agents herd [22] - they ignore their own observations and parrot decisions of previous agents. The multi-threshold structure of the global decision maker's optimal policy (Fig. 2) can be interpreted as saying that the global decision maker acts in a non-trivial manner to compensate for the poorly informed local decision makers. In comparison, the classical threshold policy (Fig. 2) results when the local decision makers are well informed (exchange their posterior distributions rather than anticipatory actions); see [23] for discussion in terms of Bayesian social learning.

3) Change Blindness. The multi-threshold change detection policy in Fig. 2 can be interpreted as a form of *change blindness*, namely, people fail to detect surprisingly large changes to

scenes [28]. Even though the posterior probability of a change is higher than a change threshold, the optimal behavior indicated is to detect no change.

4) Deliberate Avoidance of Information. Theorem 1 in Sec. III shows that the subgame Nash equilibrium at time 1 has a bang-bang structure. It justifies the observation [1] that agents with anticipatory emotions may choose to deliberately avoid information. As mentioned earlier, [7] reports that giving some patients more information before a stressful medical procedure raised their anxiety. [5] shows that humans selectively treat the opportunity to gain knowledge about future favorable outcomes, but not unfavorable outcomes.

Finally, we emphasize that humans likely do not solve time inconsistent problems to make decisions. The time inconsistent behavioral economics models in [1], [14], [21] are widely used because they provide generative models for the peculiarities of anticipatory human decision making.

F. Organization

The paper is organized into three inter-related parts:

1) **Part 1** deals with anticipatory models for a single decision maker and characterizes the Nash equilibrium.
2) **Part 2** of the paper deals with quickest change detection with a team of anticipatory decision makers.
3) **Supplementary Material** contains proofs of theorems and a detailed tutorial example of anticipatory decision making in social media.

Part 1. Anticipatory Models, Nash Equilibrium

Sec. II formulates anticipatory decision making. Sec. III characterizes the structure of the Nash equilibrium with examples.

II. Anticipatory Decision Making

This section defines time inconsistent decision problems and reviews the influential behavioral economics model [1] for human decision making with anticipatory feelings. This model will be used in Sec. IV to formulate our human sensor interactive quickest change detection problem.

A. Time Inconsistent Sequential Decision Problems

We start with a brief discussion of time inconsistent decision problems; see [12] for an exposition. Let $\{s_k, k = 1, \dots, N\}$ denote a controlled Markov chain evolving on a finite time horizon size N. The initial distribution for s_1 is denoted as π_1. Let μ_k denote a (possibly randomized) decision policy that maps the state s_k to an action a_k at time k. For $n = 1, 2, \dots, N$, define the expected utility-to-go

$$J_n(s_n, \boldsymbol{\mu}_{n:N}) = \mathbb{E}_{\boldsymbol{\mu}_{n:N}} \left\{ \sum_{k=n}^{N} r_{n,k}\left(s_n, s_k, \mu_k(s_k)\right) \right\} \quad (1)$$

The aim is to compute the policy sequence $\operatorname{argmax}_{\boldsymbol{\mu}} J_n(s_n, \boldsymbol{\mu}_{n:N})$. As the reward $r_{n,k}$ depends on n and k, and also s_n, s_k, this optimization problem is *time inconsistent* since the principle of optimality (Bellman's dynamic programming equation) does not hold; see [12].

1) Subgame Perfect Nash Equilibria: As discussed in [12], an appropriate method of "solving" a time inconsistent problem is in game-theoretic terms.[2]

1) Given state $s_N = s$, player N chooses policy
$$\mu_N^*(s) = \operatorname*{argmax}_{a_N} J_N(s, a_N) \quad (2)$$
This yields the value function $V_N = J_N(s, \mu_N^*)$.
2) Given $s_{N-1} = s$, and that player N is using policy μ_N^*, player $N-1$ chooses policy
$$\mu_{N-1}^*(s) = \operatorname*{argmax}_{a_{N-1}} J_{N-1}(s, a_{N-1}, \mu_N^*)$$
$$V_{N-1}(s) = J_{N-1}(s, \mu_{N-1}^*, \mu_N^*) \quad (3)$$
3) Proceed by backward induction to compute policies $\mu_{N-2}^*, \dots, \mu_1^*$.

The above procedure is called the *extended* Bellman equation in [12]. The sequence of policies $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_N^*)$, constitutes a subgame perfect Nash equilibrium; see [12] for details.

2) Remarks:

i) As might be expected, for the time consistent case where $r_{n,k}(s_n, s_k, a_k) = r_k(s_k, a_k)$ in (1), the extended Bellman's equation becomes the standard Bellman's dynamic programming equation.

ii) For the time inconsistent case, neither the Nash equilibrium $\boldsymbol{\mu}^*$ nor its value $J_n(\boldsymbol{\mu}^*)$ are unique. This is in contrast to time consistent dynamic programming where the optimal policy may not be unique but the optimal value is always unique.

B. Anticipatory Model of Caplin & Leahy [1]

We now review the time inconsistent model for anticipatory human decision making in Caplin & Leahy's paper [1]. Their model uses the terminology of temporal lotteries in dynamic choice theory [29]. We translate their model to a more familiar Markov decision process framework. While the messy notation below is unavoidable, the reader should keep in mind that the final outcome is a time inconsistent problem of the form (1) with horizon $N = 2$. A key step in the formulation below is the anticipatory state (5) 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 [1] 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

$$a_1 = \mu_1(s_1), \quad a_2 = \mu_2(s_2, a_1). \quad (4)$$

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

$$z_1 = \phi\left(s_1, a_1, \{p(a_2 = a|s_1, a_1, \mu_2), a \in \mathcal{A}_2\}\right),$$
$$z_2 = (s_2, a_2, a_1), \quad (5)$$

[2]The following intuitive argument from [12] is helpful: Looking to maximize $J_n(s, \boldsymbol{\mu}_{n:N})$ over the class of policies restricted to $[n : N]$, a player at time n would like in principle to maximize $J_n(s, \boldsymbol{\mu}_{n:N})$ over $\mu_n, \dots, \mu_N$. But the player at time n can only choose the policy μ_n - so the maximization is not possible. Instead of looking for optimal feedback laws, in a time inconsistent problem one considers the subgame perfect Nash equilibrium.

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 [1], 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)[3] 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 [1] is that the anticipatory agent makes decisions by maximizing the 2-stage anticipatory utility

$$\sup_{\mu_1,\mu_2} J(\mu_1, \mu_2) = \mathbb{E}_{\mu_1,\mu_2}\{r_1(z_1) + r_2(z_2)\} \tag{6}$$

Here $r_k(z_k) \in \mathbb{R}$ denote the reward functions. The 2-stage anticipatory utility, called psychological utility in [1]. (6) 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. Indeed (6) is a special case of the general time inconsistent formulation (1) with

$$\begin{aligned} &r_{2,2} = r_2(s_2, a_2, a_1),\ r_{1,1} = 0, \\ &r_{1,2} = r_1\left(\phi(s_1, a_1, \{p(a_2 = a|s_1, a_1, \mu_2), a \in \mathcal{A}_2\})\right) + r_{2,2} \end{aligned} \tag{7}$$

As in [1], we assume 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.

2) Subgame Perfect Nash Equilibrium: Caplin & Leahy [1] 'solve' the time inconsistent decision problem (6) using the extended Bellman equation described in Sec. II-A. Indeed, the optimal policy at time 2 simply follows from (2) with $N = 2$:

$$\mu_2^*(s_2, a_1) = \operatorname*{argmax}_{a_2} r_2(s_2, a_2, a_1) \tag{8}$$

Note that by definition (4), μ_2^* depends on a_1 and s_2.

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

$$\begin{aligned} \mu_1^*(s_1) &= \operatorname*{argmax}_{a_1} J_1(s_1, a_1, \mu_2^*), \\ V_1(s_1) &= \max_{a_1} J_1(s_1, a_1, \mu_2^*), \\ J_1(s_1, a_1, \mu_2^*) &= r_1\left(\phi(s_1, a_1, \{p(a_2 = a|s_1, a_1, \mu_2^*), a \in \mathcal{A}_2\})\right) \\ &\quad + \mathbb{E}\{r_2(s_2, a_2, a_1)|s_1, a_1,\ \mu_2^*\} \end{aligned} \tag{9}$$

Note that the second term $\mathbb{E}\{r_2(s_2, a_2, a_1)|s_1, a_1,\ \mu_2^*\}$ is

$$\begin{aligned} &\int_{\mathcal{S}} r_2\left(s_2, \mu_2^*(s_2, a_1), a_1\right)\ p(s_2|s_1)\, ds_2 \\ &= \int_{\mathcal{A}_2}\int_{\mathcal{S}} r_2\left(s_2, a, a_1\right)\ I\left((a = \mu_2^*(s_2, a_1))\right)\ p(s_2|s_1) ds_2 da \end{aligned} \tag{10}$$

[3] As discussed in [1] 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 [30]. A terrorist attack (unlikely event) worries people a lot more than a car crash (high probability event).

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

Remarks:

1) Eqs.(9), (10) are identical to the master equation [1, Eq.2].

2) The anticipatory (psychological) state z_1 in (5) 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

$$\{\mathbb{E}\{\Psi(a_2 = a, s_2)|s_1, a_1, \mu_2\}, a \in \mathcal{A}_2\} \tag{11}$$

for some pre-defined function Ψ. As an example (which is elaborated on in the supplementary material)

$$z_1 = \max\{p(a_2 = 1|s_1, a_1, \mu_2), \mathbb{E}\{s_2 I(a_2 = 2)|s_1, a_1, \mu_2)\}$$

3) 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 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.

Summary. The key point in anticipatory decision making is the presence of probabilities of choosing future actions in the current reward, as depicted in the anticipatory state (5). As a result, maximizing the 2-stage anticipatory utility (6) is a time inconsistent problem. The anticipatory decision maker chooses actions a_1, a_2 according to policies μ_1^* in (9) and μ_2^* in (8); these policies constitute a subgame perfect Nash equilibrium. Indeed (9), (10) corresponds to the key equation (2) in [1]. The paper [1] has received significant attention in behavioral economics [4], neuroscience and psychology [5].

C. Example 1. Financial Investment and Anticipatory Betting

The following example (based on [1]) presents anticipatory decision making in a simplified setting to illustrate rapidly the key ideas. The problem is time inconsistent since the utility at time 1 depends on the expected physical state at time 2.

There are two periods. An investor makes two decisions denoted a and $\bar{a}$ in period 1 (this simplifies the problem).

1) The decision $a \in \{\texttt{stock}, \texttt{bond}\}$ is whether to invest in short term stock or long term bonds. If $a = \texttt{stock}$, then the agent chooses $\bar{a} \in [0, W_0]$, namely how many units to invest in stock, where W_0 denotes the initial wealth.
2) The physical state s_1 denotes the probability that stock yields a return good. For simplicity, assume $s_1 = 1/2$.
3) At time 2, the physical state $s_2 \in \{\texttt{good}, \texttt{bad}\}$ denotes whether the return on stock is satisfactory or not.
4) If the investor chooses $a = \texttt{stock}$, invests $\bar{a}$, and the return s_2 is good, then it earns $2\bar{a}$; so its wealth at the end of period 2 is $w = W_0 + \bar{a}$. If the return is bad, the investor loses $\bar{a}$ and its wealth at the end of period 2 is $w = W_0 - \bar{a}$.
5) If the investor chooses $a = \texttt{bond}$, then it invests the entire W_0 and obtains a return of $W_0 + \iota$, where ι denotes the interest payment.

Given final wealth w, assume the agent's utility at time 2 is

$$r_2(s_2, \bar{a}, a) = w - \beta\, w^2 \tag{12}$$

This utility models a risk averse agent with quadratic penalty loss (which is used widely in behavioral economics).

We assume that the agent's anticipatory utility at time 1 is

$$J_1(s_1, a = \texttt{stock}, \bar{a}) = \alpha(u_A + \bar{a} - \beta\bar{a}^2) + \mathbb{E}\{r_2(s_2, \bar{a}, a)|s_1\}$$

$$J_1(s_1, a = \texttt{bond}, \bar{a}) = g + \mathbb{E}\{r_2(s_2, \bar{a}, a)|s_1\} \quad (13)$$

where u_A, α, g, β are positive constants. This decision problem is time inconsistent since the utility at time 1 depends on the expected physical state s_2. Recall that decisions $a, \bar{a}$ are made at time 1 only (so there is no μ_2^* in (9)). The term $\alpha(u_A + \bar{a} - \beta\bar{a}^2)$ is the excitement (suspense) of investing $\bar{a}$; the term $-\beta\bar{a}^2$ models the risk averseness of the agent.

Let us work out J_1 in (13) explicitly. Since the probability of stock returning good is 1/2, clearly

$$\mathbb{E}\{r_2(s_2, \bar{a}, a = \texttt{stock})|s_1\} = W_0 - \beta(W_0^2 + \bar{a}^2)$$

$$\mathbb{E}\{r_2(s_2, \bar{a}, a = \texttt{bond})|s_1\} = W_0 + \iota - \beta\,(W_0 + \iota)^2 \quad (14)$$

Therefore the optimal investment $\bar{a}$ is zero if only the second period expected utility is considered. The utility J_1 in (13) captures the tradeoff between the excitement and future anticipatory gain/loss, leading to a time inconsistent problem.

The time consistent optimal policy at time 1 using (9) is:

$$\mu_1^*(s_1) = (a^*, \bar{a}^*)$$

$$\bar{a}^* = \operatorname*{argmax}_{\bar{a}\geq 0} \alpha u_A + \alpha\bar{a} - (1+\alpha)\beta\bar{a}^2 = \frac{\alpha}{2(1+\alpha)\beta}$$

$$a^* = \begin{cases} \texttt{stock} & \text{if } \iota(1 - 2\beta W_0) - \beta\iota^2 + g < \alpha(u_A + \frac{\alpha}{4(1+\alpha)\beta}) \\ \texttt{bond} & \text{otherwise} \end{cases} \quad (15)$$

Anticipatory Betting/Gambling [1]

We now describe an example involving anticipatory betting/gambling [1]. The setup is a special case of above. An agent chooses action $a \in \{\texttt{bet}, \overline{\texttt{bet}}\}$. $\bar{a} \in [0, W_0]$ denotes how much money is bet. The physical state $s_1 = P(\texttt{win}) = 1/2$, namely, anticipated probability of win at stage 1, and $s_2 \in \{\texttt{win}, \overline{\texttt{lose}}\}$ denotes the actual outcome at stage 2.

1) If the agent chooses $a = \texttt{bet}$ then the final wealth is $W_0 + \bar{a}$ if the bet is won ($s_2 = \texttt{win}$) or $W_0 - \bar{a}$ if the bet is lost ($s_2 = \overline{\texttt{lose}}$).
2) If the agent chooses $a = \overline{\texttt{bet}}$, then the final wealth remains W_0 (instead of $W_0 + \iota$, i.e., interest $\iota = 0$).
3) The risk averse utility at stages 2 and 1 are (12), (13) with $\iota = 0$, bet replacing stock and $\overline{\texttt{bet}}$ replacing bond.

Then (14) holds with $\iota = 0$. The Nash equilibrium policy $\mu_1^*(s_1)$ is (15) with $\iota = 0$, bet replacing stock, $\overline{\texttt{bet}}$ replacing bond. Sec. V-E illustrates this model in quickest detection.

Implications of Anticipatory Investment/Betting

The agent chooses stock (or bet) even though it loses in terms of the risk averse final utility (14), yet individuals gamble because it heightens the action suspense (anticipation) prior to the resolution of uncertainty in the second stage. This illustrates the time inconsistency of the problem: in the final period it is not useful to invest in stock (or bet). Yet the investment is made at stage 1 with anticipatory feelings rather than the ultimate outcome; see [1] for implications in gambling/betting. To quote Samuelson [31]: "I am satisfied that a large fraction of the sociology of gambling and of risk taking will never significantly be discernible in terms of the money prizes alone, as distinct from elements of suspense...."

III. Characterizing the Nash Equilibrium Policy of Anticipatory Decision Maker and Examples

The previous section gave a general setup of the anticipatory decision making model and associated subgame Nash equilibrium policy. However, the Nash equilibrium (9), (10) is the solution of the extended Bellman equation (integral equation) and is difficult to compute in general. In this section, our main contribution is to 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). This structural result illustrates the optimality of simple decision-making rules and will be illustrated by an example involving situation awareness.

Bayesian parametrization of transition kernel and reward

Recall r_2 is the reward at time 2; see (5), (6). 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 parameterized 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 $x \in \mathcal{X} = \{1, 2, \ldots, m\}$ will be formally defined in Sec. IV to model change in quickest detection. Then define the parametrized reward $r_{\eta,2}$ and transition kernel $p_\eta(s_2|s_1)$ as

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

$$p_\eta(s_2|s_1) = \sum_{x\in\mathcal{X}} p(s_2|s_1, x)\,\eta(x) \quad (16)$$

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

$$\Pi = \{\eta : \eta(i) \in [0,1], \quad \sum_{i=1}^{m} \eta(i) = 1\} \quad (17)$$

The posterior η will be formally defined in (23) and appears naturally in the quickest change detection formulation in Sec. IV (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.

A. Structural Characterization of Nash Equilibrium

With $r_{\eta,2}$ defined in (16), for notational convenience, define

$$\Delta_\eta(s_2, a_1) = r_{\eta,2}(s_2, 2, a_1) - r_{\eta,2}(s_2, 1, a_1) \quad (18)$$

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

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 (18) 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 (11)) 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.

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 (17)) as $\mu_{1,\eta}^*$ and $\mu_{2,\eta}^*$.

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

1) Under (A3), (A4), the subgame perfect Nash equilibrium policy μ_2^* specified by (8) has a threshold structure:

$$\mu_{2,\eta}^*(s_2, a_1) = \begin{cases} 1 & \text{if } s_2 \leq s_{2,\eta}^*(a_1) \\ 2 & s_2 > s_{2,\eta}^*(a_1) \end{cases} \tag{19}$$

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 (9) is convex in a_1. Therefore, the subgame Nash equilibrium policy μ_1^* has the following bang-bang[4] structure:

$$\mu_{1,\eta}^*(s_1) = \begin{cases} 1 & \text{if } \beta > \beta^* \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

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

The proof is in the supplementary document.

Deliberate Avoidance of Information. The structure of the Nash policy in Theorem 1 yields interesting consequences. Suppose a_1 denotes a non-refundable financial deposit made by the agent at time 1 in anticipation of choosing action $a_2 = 1$ at time 2. Due to the bang-bang structure of (20) the agent makes a full deposit $a_1 = 1$ if $\beta > \beta^*$. Yet this full non-refundable deposit does not guarantee that the agent will choose $a_2 = 1$ since if $s_2 > s_2^*(a_1)$, then the agent will choose $a_2 = 2$. Thus the agent would like to avoid observing s_2. There is an elegant interpretation of this in [1], namely, the agent might deliberately choose not to observe the state s_2 in order not to lose the deposit. "In this manner, anticipatory emotions may rationalize the deliberate avoidance of information" [1].

B. Discussion of Assumptions

Assumptions (A1)–(A7) are generalizations of (and therefore weaker than) the assumptions in [1], where an example of anticipatory decision making for choosing a holiday destination is discussed. Note that (A2) to (A5) are assumptions on $r_{\eta,2}$, while (A6),(A7) are assumptions on r_1.

A1) In [1] and also the social media accommodation example (supplementary material), $\mathcal{A}_1 = [0,1]$ denotes the feasible set of deposits made to secure an accommodation, while $\mathcal{A}_2 = \{1,2\}$ denotes the choices of accommodation.

[4]The phrase "bang-bang controller" comes from classical optimal control theory. It characterizes a control policy with continuous-valued actions that switches between two extremes.

A2) In [1] and also the accommodation example, the reward $r_{\eta,2}$ is chosen as linear in a_1. This is because a_1 is a deposit made at time 1; so the reward at time 2 is the net wealth minus the deposit at time 1.

Assuming the reward $r_{\eta,2}$ to be convex in a_1 is more general and still yields the same structural result.

A3) is a supermodularity assumption and implies that a_2 and s_2 satisfy Edgeworth complementarity [32]. This means that increasing s_2 increases the marginal value of choosing $a_2 = 2$ compared to $a_2 = 1$. This is intuitive: For the accommodation example, a higher review of N gives more incentive to choose accommodation N. Supermodularity is widely used to characterize the structure of policies in stochastic control and game-theory. By Topkis' famous theorem [32], supermodularity (A3) implies Nash policy $\mu_2^*(s_2, a_1)$ is non-decreasing in s_2 for fixed a_1. This together with (A4) implies that μ_2^* has a threshold structure (19) wrt s_2 (see proof). In [1] and the social media accommodation example, (A3) holds trivially since $r_{\eta,2}(s_2, a_2 = K, a_1)$ is independent of s_2;

To motivate the remaining assumptions, we first note that Assumptions (A2)–(A7) imply that the anticipatory state z_1 is convex in a_1 (as shown in the proof). Since a convex function is maximized at its end points of $\mathcal{A}_1 = [0,1]$, namely 0 and 1, the bang-bang structure (20) for μ_1^* holds. We now dive deeper into (A4)–(A7).

A4): (A4) is simply an assumption on the well-posedness of the setup; namely, that there exists a threshold point $s_{2,\eta}^*(a_1)$; implying that the anticipatory agent makes simple intuitive decisions a_2 based on the state s_2.

A5) is a prescriptive assumption on the rewards $r_{\eta,2}$. From a risk averse point of view, it is natural that a higher deposit a_1 should result in requiring a substantially higher review x_2 in order for a_2 to forfeit the deposit on K and switch to N. This is captured by requiring that the threshold point $s_{2,\eta}^*(a_1)$ in (19) is convex in a_1. The natural question then is: What assumptions on the reward guarantee this convexity? Statement 2 of Theorem 1 is equivalent to showing convexity in a_1 of the solution $s_2^*(a_1)$ of the algebraic equation $\Delta_\eta(s_2, a_1) = 0$. It is here that (A5) is used. (A5) and (A4) are sufficient for the implicit solution to an algebraic equation involving two variables to be convex wrt the other variable. Showing convexity of the implicit solution to an algebraic equation dates back to [33] where (A5) is used. (A4) can be relaxed based on the classical implicit function theorem [34]; see supplementary document for details. In the accommodation example, (A4), (A5) hold trivially since Δ_η is linear in s_2, a_1.

A6) states that the anticipatory reward is linear in the psychological state. Therefore β denotes the importance of the anticipatory reward relative to the reward at time 2. This assumption is identical to that in [1].

A7) is also a prescriptive assumption on the system behavior to ensures that the psychological state z_1 is convex in action a_1. Actually in [1] and the accommodation example, the psychological state z_1 is linear and increasing in action a_1. From a behavioral point of view, convexity of the psychological state in a_1 is natural since it yields the bang-bang structure (20) of the Nash equilibrium which motivates the "deliberate avoidance of information behavior" discussed above.

The convexity of rewards (A2) and assumption (A7) together with Statement 2 imply that anticipatory (psychological) state

z_1 is convex in a_1. Specifically in the accommodation example and also [1], $p_\eta(s_2|s_1)$ is uniformly distributed in s_2,

$$\Psi(a_2 = 1, s_2) = s_2\, I\left(x_2 \in \left[\frac{2+a_1}{3\eta}, 1\right]\right)$$

$$\Psi(a_2 = 2, s_2) = I\left(x_2 \in \left[0, \frac{2+a_1}{3\eta}\right]\right)$$

which clearly satisfy (A7); see the examples for details.

C. Example 2. Anticipatory Situation Awareness (SA)

We now discuss an anticipatory decision making example involving Level 3 SA. The example will be developed further in the context of quickest time change detection in Sec. IV.

1) Model: The physical states s_1 and s_2 denote the probability that the threat level of a target (or group of targets) is `low threat` or `high threat`, at stages 1 and 2.

Regarding the actions, at the first stage the SA system chooses action $a_1 \in [0, 1]$ which denotes fraction of resources devoted to tracking a specific target. At the second stage, the SA makes the final choice of whether to take active measures (e.g. intercept the target) or choose passive measures (continue to track it), i.e., $a_2 \in \mathcal{A}_2 = \{\texttt{active}, \texttt{passive}\}$.

Next we model the anticipatory decision making of the SA system. We choose the anticipatory reward to reflect beliefs about threat levels that will be derived in choosing respectively, `active` and `passive`. We choose the anticipatory state z_1 at time 1 as the conditional probabilities (see (5))

$$z_1 = \max\{6\, p(a_2 = \texttt{active}, s_2 = \texttt{high threat}|a_1, \mu_2), \\ 4\, p(a_2 = \texttt{passive}|a_1, \mu_2)\} \quad (21)$$

(We allocate numerical values to make the example more readable.) So the anticipatory reward increases with the SA's plan to use an active measure if the threat is high.

We now construct the rewards r_1, r_2 defined in (6).

1) Choosing action a_1 expends $2\,a_1$ resources on planning for `active` measures at time 2. If `passive` is chosen at time 2, then the resources of $2\,a_1$ are wasted (lost).
2) The reward accrued by choosing `active` when the threat level is s_2 is $6s_2\eta$; the reward for choosing `passive` is fixed at 4. Here[5] $\eta \in [0, 1]$ is the posterior probability that the threat level with action `active` is high given information from sensing functionalities.

Based on the above description, the rewards are

$$r_1 = \beta z_1, \quad r_{\eta,2}(s_2, a_2 = \texttt{active}, a_1) = 6\, s_2\eta - 2\, a_1,$$

$$r_{\eta,2}(s_2, a_2 = \texttt{passive}, a_1) = 4$$

2) Structure of Nash Equilibrium: For simplicity, assume s_2 is uniformly distributed in [0,1]. For the above example, we can verify Assumptions (A1)–(A7) hold and therefore Theorem 1 holds. Specifically, (A1) holds by formulation; (A2) holds trivially since $r_{\eta,2}$ is linear in a_1; (A3) holds since $r_{\eta,2}(s_2, a_2 = \texttt{passive}, a_1)$ is independent of s_2; (A4) and (A5) hold trivially since Δ_η is linear in s_2 and a_1; (A6) holds by construction since it is easily shown that for optimal policy μ_2^*, $z_1 = 4\, p(a_2 = \texttt{passive}|a_1, \mu_2^*)$. Finally, (A7) holds since $p(s_2)$ is the uniform density by assumption.

[5]We assume $\eta = [\eta(1), \eta(2)]'$ is a 2-dimensional probability vector, i.e., $m = 2$ in (17). For notational convenience, we refer to $\eta(2)$ as η.

3) Consequences: Theorem 1 implies μ_2^* has a threshold structure (19), and μ_1^* has a bang-bang structure (20). The bang-bang structure (20), represents a dilemma to the SA system. The SA system fully utilizes its resources, $a_1 = 1$ towards plan `active` if $\beta > \beta^*$. Yet this does not guarantee that the SA system will choose $a_2 = \texttt{active}$ since if $s_2 > s_2^*(a_1)$, then the agent will choose $a_2 = \texttt{passive}$. Thus a human-in the-loop in the SA system might deliberately choose not to observe the state s_2 in order not to lose the effort invested at stage 1.

D. Other Examples

Example 3. Airbnb example of Social media accommodation: The supplementary document gives a detailed example in social media accommodation with a similar dilemma due to the bang-bang Nash equilibrium structure: avoid information at stage 2 so as not to lose the full deposit made at stage 1.

Example 4. Asset Prices and Anxiety: [1] presents a two stage model for portfolio choice and the anxiety of holding risky assets. The anxiety encountered at time 1 depends on the expected reward and variance of the reward at time 2.

Part 2. Quickest Change Detection for Team Anticipatory Decision Makers

Part 1 of the paper described how a single anticipatory agent makes decisions over a two-period time horizon. In Part 2, we consider a *team* of anticipatory agents (or equivalently, a single agent that acts multiple times). These anticipatory agents interact with each other sequentially. Each anticipatory agent observes the state of nature (Markov chain) in noise and makes local decisions as described in Sec. II-B. A global decision maker observes these decisions. *How can a global decision maker use these local decisions to detect a change in the state of nature*? Specifically the aim is to achieve quickest change detection by minimizing the Kolmogorov-Shiryaev criterion (defined in (30) below) which involves the false alarm and delay penalties.

Examples of Team-Based Quickest Detection

Before proceeding with the quickest change detection formulation, it is helpful to keep the following examples in mind:

i) *Change in Quality of Social Media Accommodation*. Suppose individual anticipatory agents choose between reserving accommodation in two places. By monitoring these decisions, how can a global decision maker (e.g. Airbnb) detect if there is a sudden change in the demand for a specific accommodation due to the presence of a new competitor (or change on quality in the accommodation)? This example is discussed in the supplementary material as a detailed tutorial.

ii) *Supervisory SA*. Sec. III-C discussed the importance of anticipatory situation assessment. Suppose a supervisory situation assessment (SA) system monitors the decisions of individual SA systems. Individual SA systems are anticipatory (as discussed in Sec. III-C) and monitor an enemy target or radar state. How can the supervisory decision maker detect if there is a sudden change in the enemy target (due to a purposeful maneuver)? This example is discussed in Sec. IV-C.

iii) *Detecting change in betting strategy*. How to detect a sudden change in the betting strategy of individuals that act sequentially? Sec. V-E discusses a numerical example which builds on the anticipatory betting model of Sec. II-C.

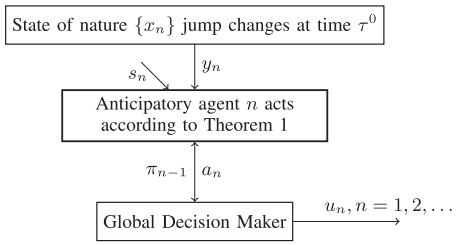


Fig. 3. Quickest Change Detection Problem involving multiple anticipatory local decision makers and a global decision maker. (i) Anticipatory agent $n = 1, 2 \ldots$ observes state of nature x_n in noise as y_n and receives public belief π_{n-1} from previous agent. It then makes anticipatory decision a_n as described in Theorem 1. (ii) The global decision maker uses local decision a_n to update the public belief π_n and makes decision $u_n \in \{1(\text{stop and declare change}), 2(\text{continue})\}$.

iv) *Detecting Market Shocks*. Suppose individual anticipatory investors make decisions based on their observation of the underlying value of an asset as in Sec. II-C, where the decisions of previous investors affect the individual's belief. How can an analyst detect sudden market shocks? See [35] for examples in high frequency financial trading.

IV. Anticipatory Quickest Change Detection

Notation: Since we consider the sequential interaction of multiple anticipatory agents, we adapt the notation of Sec. III:

1) Each anticipatory agent acts in a predetermined order indexed by $n = 1, 2, \ldots$.
2) The physical states s_1, s_2 (defined in Sec. II-B) encountered by agent n are now denoted by $s_{n,1}, s_{n,2}$.
3) Anticipatory decisions a_1, a_2 (characterized in Theorem 1) of agent n are denoted as $a_n \stackrel{\text{defn}}{=} [a_{n,1}, a_{n,2}]$.
4) The Bayesian belief parameter η (16) of agent n is η_n.
5) Due to the bang-bang structure ((20) in Theorem 1) of the Nash equilibrium policy, $a_{n,1}$ is independent of $s_{n,1}$. Also from (19), $a_{n,2}$ depends on $s_{n,2}$ and not $s_{n,1}$. So for convenience we denote $s_{n,2}$ as s_n.
6) The physical state process $\{s_n, n \geq 1\}$ on state space $\mathcal{S}$ is Markovian with transition density $p(s_{n+1}|s_n, x_n)$, see (16). Here x_n is the state of nature process (defined below) that models the jump change we aim to detect.

Jump Change Model. The state of nature $\{x_n \in \{1, 2\}, n \geq 0\}$ models the change event we aim to detect. It starts state 2 at time 0 and jumps to state 1 at a geometrically distributed random time τ^0 with mean $1/(1-p)$, for some prespecified $p \in [0, 1)$. Equivalently, $\{x_n\}$ is a 2-state Markov chain with absorbing transition matrix and initial probability

$$P = \begin{bmatrix} 1 & 0 \\ 1-p & p \end{bmatrix}, \quad \pi_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{22}$$

with change time $\tau^0 = \inf\{n : x_n = 1\}$. Clearly the transition matrix P implies that $\mathbb{E}\{\tau^0\} = 1/(1-p)$.

A. Multi-Agent Quickest Detection Protocol

Quickest detection involves detecting change time τ^0 with minimal cost. The multi-agent formulation considered here comprises of interacting local decision makers (anticipatory agents) and a global decision maker (see Fig. 3):

1) The jump change process (state-of-nature) $\{x_n, n \geq 0\}$ affects the transition kernel and reward of the physical state process $\{s_n, n \geq 1\}$; see (16).
2) Each anticipatory agent n acts sequentially indexed by $n = 1, 2, \ldots$. Agent n observes state of nature x_n in noise and makes a local decisions $a_n = (a_{n,1}, a_{n,2})$ corresponding to actions a_1, a_2 in Sec. III.
3) Based on the history of local actions $a_1, \ldots, a_n$, the global decision maker chooses action $u_n \in \{1 \text{ (stop and announce change)}, 2 \text{ (continue)}\}$

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

$$\begin{aligned} \pi_n(x) &= \mathbb{P}(x_n = x | a_1, \ldots, a_n), \quad x = 1, 2. \\ \eta_n(x) &= \mathbb{P}(x_n = x | a_1, \ldots, a_{n-1}, y_n), \end{aligned} \tag{23}$$

Protocol 1: Multi-Agent Bayesian Quickest Detection.

1) *Local anticipatory decision maker* n
   a) Obtains public belief π_{n-1} from global decision maker.
   b) The agent records private noisy observation $y_n \in \mathcal{Y}$ of state of nature x_n with conditional density
   $$B_{x,y} = p(y_n = y | x_n = x) \tag{24}$$
   c) *Private Belief*. The agent evaluates the private belief
   $$\eta_n = T(\pi_{n-1}, y_n) \text{ where, } T(\pi, y) = \frac{B_y P' \pi}{\sigma(\pi, y)},$$
   $$\sigma(\pi, y) = \mathbf{1}' B_y P' \pi, \quad B_y = \text{diag}(B_{1,y}, B_{2,y}) \tag{25}$$
   d) *Change Event & Local decision*. The agent's private belief η_n affects its reward and transition kernel of physical state process $\{s_n, n \geq 1\}$ as in (16):
   $$r_{\eta,2}(s_2, a_2, a_1) = \sum_{x \in \mathcal{X}} r_2(s_2, a_2, a_1, x)\, \eta(x)$$
   $$p_\eta(s_2 | s_1) = \sum_{x \in \mathcal{X}} p(s_2 | s_1, x) \eta(x) \tag{16 repeated}$$
   The agent uses η_n, s_n to make anticipatory decisions $a_n = (a_{n,1}, a_{n,2})$ via (20), (19) in Theorem 1.
2) *Global decision maker*. Based on the decisions a_n of local decision maker n, the global decision maker:
   a) Updates the public belief from π_{n-1} to π_n as
   $$\pi_n = \bar{T}(\pi_{n-1}, a_n, s_n) \tag{26}$$
   $$\bar{T}(\pi, a, s) = \frac{R_a^\pi(s)\, P' \pi}{\bar{\sigma}(\pi, a, s)}, \quad \bar{\sigma}(\pi, a, s) = \mathbf{1}' R_a^\pi(s) P' \pi$$
   $$\text{where } R_a^\pi(s) = \text{diag}(R_{1,a}^\pi(s), R_{2,a}^\pi(s)),$$
   $$R_{x,a_n}^\pi(s) = \mathbb{P}(a_n = a | x_n = x, \pi_{n-1}, s_n = s) \tag{27}$$
   The action probabilities $R_{x,a}^\pi$ are computed as
   $$R_{x,a}^\pi(s) = \int_{\mathcal{Y}} I(\mu_{2,T(\pi,y)}^*(s, a_{n,1}) = a_{n,2}) B_{x,y} dy \tag{28}$$
   Recall $a_n = (a_{n,1}, a_{n,2})$ and $\mu_{2,\eta}^*$ is the local decision maker's subgame Nash equilibrium policy (19).
   b) Chooses global action u_n using optimal policy ϕ^*:
   $$u_n = \phi^*(\pi_n, s_n) \in \{1 \text{ (stop)}, 2 \text{ (continue)}\}. \tag{29}$$
   c) If $u_n = 2$, then set n to $n+1$ and go to Step 1.
   If $u_n = 1$, then stop and announce change.

where y_n is the private observation recorded by agent n (see (24). Note $\eta = [1-\eta(2), \eta(2)]'$ and $\pi = [1-\pi(2), \pi(2)]'$; they lie in the one dimensional simplex $\Pi = [0,1]$.

We are now ready to describe the multi-agent quickest detection protocol, see also Fig. 3 for a schematic setup.

Remark: Note that there are two states in our formulation: the state of nature x (that jump changes) which is observed in noise by anticipatory agents, and the physical state s which determines the agent's anticipation. As specified in Step 1 d, the state of nature x affects the transition kernel of s and reward of each anticipatory agent. The global decision maker's quickest detection aim is to detect the jump change in x.

B. Quickest Detection Objective of Global Decision Maker

We assume the global decision maker knows P (22), physical state s_n, agent's action a_n, and agent's policy $\mu^*_{2,\eta}$. The global decision maker does not know y_n (agent's observation/perception) or the agent's private belief η_n in Step 1. For simplicity, we assume all agents have the same anticipatory model parameters; otherwise the optimal quickest detection strategy is non-stationary. We emphasize that the transition probabilities of the physical state and utility of each agent depends on its private belief η_n of x_n, see (16) in Protocol 1. 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 ϕ^* of the global decision maker that minimizes the Kolmogorov–Shiryaev criterion for detection of disorder [3]:

$$J_{\phi^*}(\pi, s) = \inf_{\phi} J_{\phi}(\pi, s),$$

$$J_{\phi}(\pi, s) = d\,\mathbb{E}_{\phi}\{(\tau - \tau^0)^+\} + f\,\mathbb{P}_{\phi}(\tau < \tau^0). \qquad (30)$$

Here $\tau = \inf\{n : u_n = 1\}$ 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. So waiting too long to announce a change incurs a delay penalty d at each time instant after the system has changed, while declaring a change before it happens, incurs a false alarm penalty f. $\mathbb{P}_{\phi}$ and $\mathbb{E}_{\phi}$ are the probability measure and expectation of the evolution of the local decisions, observations and Markov state which are strategy dependent. In (30), π denotes the initial distribution of the Markov chain x and s is the initial state of the physical state process.

Anticipatory	**Classical**
$x_n \sim P$ (change state)	$x_n \sim P$ (change state)
Local Anticipatory Decision:	Decision maker:
$y_n \sim B_{x_n,y}$ (observation)	$y_n \sim B_{x_n,y}$
$\eta_n = T(\pi_{n-1}, y_n)$	$\eta_n = T(\eta_{n-1}, y_n)$
$a_n = \mu^*_{2,\eta_n}(s_{n,2}, a_{n,1})$	$u_n = \phi^*(\eta_n) \in \{1,2\}$
Global Decision maker:	
$\pi_n = \bar{T}(\pi_{n-1}, a_n, s_n)$	
$u_n = \phi^*(\pi_n) \in \{1,2\}$	

(31)

Remark: Comparison with Classical Quickest Detection. Quickest detection with anticipatory agents (Protocol 1) is substantially more general than classical quickest detection.

As shown in (31), in classical quickest detection the decision maker has access to observations $\{y_n\}$ which are noisy measurements of $\{x_n\}$, and then computes belief η_n. In comparison, in our framework the global decision maker only has access to the local decisions $\{a_n\}$ of the anticipatory agents; these local decisions depend on y_n via the dynamics in Steps 1c and 1 d in Protocol 1. In particular, the public belief π_n in (27) depends on the action likelihoods; whereas in classical quickest detection the belief depends on the observation likelihoods. The objective of classical quickest detection is exactly (30) except that the belief is the classical Bayesian posterior $p(x_n|y_{1:n})$ instead of π_n defined in (23).

C. Example. Change Detection in Team Situation Awareness

Sec. III-C described an individual anticipatory situation awareness (SA) system. In complex environments individual SA is no longer adequate. We consider here *team-level* SA [17]. For example, [18] introduced a situational adapting system to assess team SA for fighter pilots based on information fusion. To achieve team SA, individual pilots need to develop and retain their own SA while performing the task, share their SA and notice relevant activities of other members in the team. In the simplest sequential framework of Team SA, we have the setup of Protocol 1 where:

1) The underlying state of nature x_n denotes the enemy target or radar state that is monitored by the SA system.
2) y_n are measurements of the enemy's state x_n.
3) π_{n-1} is the enemy's belief $p(x_{n-1}|a_1, \ldots, a_{n-1})$ obtained from a Bayesian tracking algorithm.
4) The physical state s_n is the probability of threat. Its transition kernel is modulated by ground truth x_n (16).
5) Individual agents in the team SA agent make decisions $a_{n,1}, a_{n,2}$ according to Protocol 1 and relay them to subsequent SA systems in the team.

Then quickest detection is motivated as follows: by monitoring the decisions $\{a_n\}$ of the individual SA systems, how can a supervisory system detect if there is a sudden change in the state $\{x_n\}$? Such a change is reflective of the enemy target making purposeful maneuvers; or the enemy radar switching modes between search, acquisition or track. Since it operates at a higher level of abstraction, the supervisory system does not have access to the observations y_n of individual SA systems.

A similar framework in social media accommodation is discussed in the supplementary document. Given the sequence of decisions $\{a_n\}$, the global decision maker (e.g. Airbnb) wishes to detect if there is a sudden appearance of competition or sudden change in quality of the accommodation x_n. The physical state s_n is the probability of a good review (review histogram) and its kernel depends on the ground truth x_n.

D. Discussion of Protocol 1

1) Sensor-human Interface. Suppose each anticipatory human decision maker is equipped with a sensing/computing device that performs Steps 2 to 4. Specifically, the noisy observation y_n in Step 3 is obtained by a sensor/computing device which then uses Bayes rule to evaluate the private belief η_n in Step 4 according to (25). The sensing functionality then provides η_n to the anticipatory decision maker. Recall that η_n enters the parametrized rewards of the anticipatory decision maker as discussed in (17). Finally, the anticipatory decision maker chooses action a_n in Step 5 according to the framework in

Sec. III. Thus Step 1 preserves the simplicity of the anticipatory human decision making model in [1].

2) Global decision maker. Step 6 details the decision making framework of the global decision maker. The global decision maker has access to the physical state s_n and the actions $a_{n,1}, a_{n,2}$ of the local decision maker. These are used by the global decision maker in Step 7 to update the public belief in (26). The action likelihoods in (28) follow from (25) and

$$R^{\pi}_{x,a}(s) = \int I(\mu^*_{2,\eta}(s, a_{n,1}) = a_{n,2})\, p(\eta|\pi_{n-1}, y) B_{x,y}\, d\eta\, dy$$

Finally in Step 8, the global decision maker applies the optimal policy ϕ^* to the updated public belief π_n, to choose whether to continue or stop (announce change).

3) Information Structure. Protocol 1 depicts three types of interactions. Local decision makers learn from previous local decision makers. Second, the local decisions a_n determine global decisions u_n. Finally, if the global decision maker chooses $u_n = 2$, then the protocol continues to the next time; otherwise a change is detected and the process stops.

4) Comparison With Bayesian Social Learning. Protocol 1 generalizes classical Bayesian social learning [22] in two ways. First, the public belief update (26) is a generalization of the Bayesian social learning filter [36], where the local decision maker is a myopic optimizer (in comparison, we now have a two-stage anticipatory local decision maker). Second, the local decision makers operate in closed loop; they are controlled by the global decision maker.

E. Stochastic Dynamic Programming Formulation

The aim of this section is to formulate the global decision maker's quickest change detection policy $\phi^*(\pi, s)$ (defined in (30)) as the solution of a stochastic dynamic programming equation. The quickest detection problem (30) is an example of a stopping-time partially observed Markov decision process (POMDP) problem with a stationary optimal policy [36].

1) Costs: To present the dynamic programming equation, as is standard, we first formulate the false alarm and delay costs (30) incurred by the global decision maker in terms of the public belief (also called the information state), see [36].

i) *False alarm penalty*: If global decision $u_n = 1$ (stop) is chosen at time n, then the Protocol 1 terminates. If $u_n = 1$ 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 (22) 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 (30), the expected false alarm penalty is

$$f\, \mathbb{P}_\phi(\tau < \tau^0) = f\, \mathbb{E}_\phi\{\mathbb{E}\, I(x_n = 2, u_n = 1)|\mathcal{G}_n\}\}$$
$$\mathcal{G}_n = \sigma\text{-algebra generated by } (a_1, \ldots, a_n) \quad (32)$$

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

$$C(\pi_n, u_n = 1) = f\, e_2' \pi_n, \quad \text{where } e_2 = [0 \quad 1]'. \quad (33)$$

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, even though the state has changed 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

$$C(\pi_n, u_n = 2) = d e_1' \pi_n, \quad \text{where } e_1 = [1 \quad 0]'. \quad (34)$$

We can re-express Kolmogorov-Shiryaev criterion (30) as[6]

$$J_\phi(\pi, s) = \mathbb{E}_\phi \left\{ \sum_{n=0}^{\tau-1} C(\pi_n, 2) + C(\pi_\tau, 1) \right\} \quad (35)$$

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 $\pi \in \Pi$, stopping is guaranteed in finite time.

2) Bellman's Equation for Quickest Detection Policy: Consider the costs (33), (34) defined in terms of the public belief π. Then the optimal stationary policy $\phi^*(\pi, s)$ defined in (29), (30). and associated value function $V(\pi, s)$ are the solution of Bellman's dynamic programming functional equation [36]

$$\begin{aligned}
&Q(\pi, s, 1) \stackrel{\text{defn}}{=} C(\pi, 1),\\
&Q(\pi, s, 2) \stackrel{\text{defn}}{=} C(\pi, 2)\\
&\quad + \int_{\mathcal{S}} \sum_{a \in \mathcal{A}_1 \times \mathcal{A}_2} \mathcal{V}\left(\bar{T}(\pi, a, \bar{s}), \bar{s}\right) \bar{\sigma}(\pi, a, \bar{s})\, p(\bar{s}|s)\} \, d\bar{s}\\
&\phi^*(\pi, s) = \arg\min\{Q(\pi, s, 1), Q(\pi, s, 2)\},\\
&\mathcal{V}(\pi, s) = \min\{Q(\pi, s, 1), Q(\pi, s, 2)\} = J^*_\phi(\pi, s) \quad (36)
\end{aligned}$$

The public belief update $\bar{T}$ and normalization measure $\bar{\sigma}$ were defined in (26). Recall (29) that $u_n = \phi^*(\pi_n, s_n)$ is the global decision maker's action whether to continue or stop.

The goal of the global decision-maker is to solve for the optimal quickest change policy ϕ^* in (36) or equivalently, determine the optimal stopping set $\mathcal{S}$

$$\mathcal{S} = \{\pi, s : \phi^*(\pi, s) = 1\} = \{\pi, s : Q(\pi, s, 1) \leq Q(\pi, s, 2)\} \quad (37)$$

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 (36) – 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 $k = 1, 2, \ldots$

$$\begin{aligned}
&\mathcal{V}_{k+1}(\pi, s) = \min_{u \in \mathcal{U}} Q_{k+1}(\pi, s, u),\\
&\phi^*_{k+1}(\pi, s) = \operatorname*{argmin}_{u \in \mathcal{U}} Q_{k+1}(\pi, s, u) \quad \pi \in \Pi,\\
&Q_{k+1}(\pi, s, 1) = C(\pi, 1), \quad Q_{k+1}(\pi, s, 2) = C(\pi, 2)\\
&\quad + \int_{\mathcal{S}} \sum_{a \in \mathcal{A}_1 \times \mathcal{A}_2} \mathcal{V}_k\left(\bar{T}(\pi, \bar{s}, a), \bar{s}\right) \bar{\sigma}(\pi, \bar{s}, a)\, p(\bar{s}|s) d\bar{s}, \quad (38)
\end{aligned}$$

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 (38) generates a

[6]The formal construction is as follows. Let $(\Omega, \mathcal{F})$ denote the underlying measurable space where $\Omega = (\mathcal{X} \times \mathcal{U} \times \mathcal{Y} \times \mathcal{S})^\infty$ is the product space endowed the with product topology, and $\mathcal{F}$ is the corresponding σ-algebra. Then for any $\pi \in \Pi$, $s \in \mathcal{S}$ and policy stationary policy ϕ, there exists a unique probability measure $\mathbb{P}_\phi$ on $(\Omega, \mathcal{F})$, see [37]. In (30) and (35), $\mathbb{E}_\phi$ denotes the expectation wrt measure $\mathbb{P}_\phi$.

sequence of lower semi-continuous value functions $\{\mathcal{V}_k\} \subset \mathcal{B}$ that converge pointwise as $k \to \infty$ to $\mathcal{V}(\pi, s) \in \mathcal{B}$, the solution of Bellman's equation [37].

Summary. Protocol 1 describes the quickest detection protocol involving anticipative agents acting sequentially. Each local decision maker (agent) $n = 1, 2, \ldots$ makes anticipatory decisions $a_{n,1}, a_{n,2}$ according to the framework in Sec. III. The global decision maker uses these actions to make decision $u_n = \phi^*(\pi_n, s_n) \in \{1, 2\}$. The optimal detection policy ϕ^* of the global decision maker satisfies Bellman's equation (36) and can be constructed by value iteration algorithm (38).

Classical quickest detection is a special case of (36), (37) with $Q(\pi, s, u)$ independent of s, $p(\bar{s}|s)\} = I(\bar{s} = s)$, and belief π replaced by classical Bayesian update (31). In classical quickest detection the optimal policy has a threshold structure and the stopping region $\mathcal{S}$ is convex; however, these properties do not hold for the multi-agent case considered here.

V. Structural Results for Quickest Detection With Anticipatory Agents

The previous section formulated Bellman's dynamic programming equation for the quickest detection policy of the global decision maker. However, since the belief space Π in (17) is a unit simplex (space of probability vectors), the value iteration algorithm (38) does not directly yield a practical solution for computing stopping set $\mathcal{S}$ since $\mathcal{V}_k(\pi)$ needs to be evaluated on the continuum $\pi \in \Pi$. Specifically, in quickest detection, since $x_k \in \{1, 2\}$, the belief space Π is a 1-dimensional simplex comprising 2-dimensional beliefs of the form $\pi = [1 - \pi(2), \pi(2)]'$. The value iteration algorithm (38) can be solved numerically by one-dimensional grid discretization of Π.

The aim of this section is to characterize mathematically the structure of the belief updates and achievable optimal cost in quickest detection without brute force computations.

Specifically we discuss 5 important structural results below:

1) The private belief update of individual anticipatory agents follows simple rules justifying human decision-making.
2) Even though the public belief update depends on the action probabilities R^π (27) where $\pi \in \Pi$ is continuum, there are only a finite number of such action probabilities.
3) In stark contrast to classical quickest detection, the value function (36) in Bellman's equation for quickest detection with anticipative agents is not necessarily concave.
4) We give numerical examples of the optimal quickest detection policy to highlight the unusual structure of non-concave value function and non-convex stopping regions. Our numerical examples illustrate change-blindness and detecting a change in betting strategy.
5) Finally, by using Blackwell dominance, we show that the cumulative cost incurred is always larger than classical quickest change detection.

A. Private Belief Update Follows Simple Monotone Rules

As discussed at the beginning of Sec. IV, the agent either uses a sensing/computing device to evaluate its private Bayesian belief or constructs an approximation to the private belief in order to make an anticipative decision. Below we show that the Bayesian update for the private belief is monotone in the observation and prior; thus it follows simple rules and is a useful idealization of human decision making.

Recall Theorem 1 asserted monotonicity of the anticipatory decision maker's policy $\mu^*_{2,\eta}(s_2, a_1)$ wrt physical state s_2. Here we show monotonicity wrt the Bayesian parameter π (recall π is the prior for η in the Bayesian update (25)) and observation y. We make the following assumptions

A8) The observation likelihoods $B_{x,y}$ (24) 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 (16)) 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.

A8) is widely studied in monotone decision making; see the classical paper [38]; numerous examples of noise distributions are TP2. As described in [39], observation $\bar{y}$ is said to be more "favorable news" than observation y if (A8) holds. (A9) is a supermodularity condition on the rewards; see (A3).

In the theorem below recall that $\mu^*_{2,T(\pi,y)}$ is the subgame Nash equilibrium of the local anticipatory decision maker.

Theorem 2: The following properties hold for the anticipatory action $a_{n,2} = \mu^*_{2,T(\pi,y)}(s, a_{n,1})$ in (19) made by agent n:

1) Under (A8) and (A9), $a_{n,2}$ is increasing and ordinal in observation y. That is for any monotone function ϕ, it follows that $\phi(a_{n,2})$ is also increasing in y.
2) Under (A8), $\mu^*_{2,T(\pi,y)}(s, a_{n,1})$ is increasing in belief π with respect to the monotone likelihood ratio (MLR) stochastic order[7] for any observation y_n. □

We can interpret Theorem 2 as follows. If anticipative agent n makes recommendations that are monotone and ordinal in the observations and monotone in the prior, then they mimic the Bayesian social learning model. Even if the agent does not exactly follow a Bayesian social learning model, its monotone ordinal behavior implies that such a Bayesian model is a useful idealization. Humans typically make *monotone* decisions - the more favorable the private observation, the higher the recommendation. Humans make *ordinal* decisions[8] since humans tend to think in symbolic ordinal terms.

We now discuss assumption (A9). Denote the reward vector $r_a \stackrel{\text{defn}}{=} [r_2(s_2, a_2 = a, a_1, x = 1), \ldots, r_2(s_2, a_2 = a, a_1, x = m)]'$ Then (A9) is a stronger version of the following more general single-crossing condition [32]: For $\bar{y} > y$

$$(r_{a+1} - r_a)' B_{\bar{y}} \pi \leq 0 \Rightarrow (r_{a+1} - r_a)' B_y \pi \leq 0. \qquad (39)$$

This single crossing condition is ordinal, since for any monotone function ϕ, it is equivalent to

$$\phi((r_{a+1} - r_a)' B_{\bar{y}} \pi) \leq 0 \Rightarrow \phi((r_{a+1} - r_a)' B_y \pi) \leq 0.$$

B. Structure of Public Belief Update

We assume in this section that the observation space and action space of the anticipatory agent are $\mathcal{Y} = \{1, \ldots, Y\}$,

[7]Given probability mass functions $\{p_i\}$ and $\{q_i\}$, $i = 1, \ldots, X$ then p MLR dominates q if $\log p_i - \log p_{i+1} \leq \log q_i - \log q_{i+1}$.

[8]Humans typically convert numerical attributes to ordinal scales before making decisions. For example, it does not matter if the cost of a meal at a restaurant is \$200 or \$205; an individual would classify this cost as "high". Also credit rating agencies use ordinal symbols such as AAA, AA, A.

$\mathcal{A}_2 = \{1, 2\}$. 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:

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

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

Theorem 3: Under (A8), (A9), it follows that

$$\pi_1^*(2) \leq \pi_2^*(2) \cdots \leq \pi_Y^*(2) \tag{40}$$

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

$$\mathcal{P}_1 = [0, \pi_1^*(2)], \mathcal{P}_2 = (\pi_1^*(2), \pi_2^*(2)], \ldots, \mathcal{P}_{y+1} = (\pi_Y^*(2), 1] \tag{41}$$

On each such interval, the action likelihood R^π (27) is a constant with respect to belief π. Specifically, for fixed a_1, s

$$R^\pi(s) = \begin{bmatrix} \sum_{i=0}^{l-1} B_{1i} & \sum_{i=l}^{Y} B_{1i} \\ \sum_{i=0}^{l-1} B_{1i} & \sum_{i=l}^{Y} B_{1i} \end{bmatrix}, \quad \pi \in \mathcal{P}_l \tag{42}$$

Example. For $Y = 3$, the 4 possible action likelihood matrices R^π are

$$R^1(s) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix},\ R^2(s) = \begin{bmatrix} B_{11} & B_{12} + B_{13} \\ B_{21} & B_{22} + B_{23} \end{bmatrix},$$

$$R^3(s) = \begin{bmatrix} B_{11} + B_{12} & B_{13} \\ B_{21} + B_{22} & B_{23} \end{bmatrix},\ R^4(s) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}. \tag{43}$$

Although tangential to this paper, agents deploying Protocol 1 can exhibit herding behavior. i.e., agents choose actions independent of their private observations; see [22], [23] for the distinction between herds and information cascades.

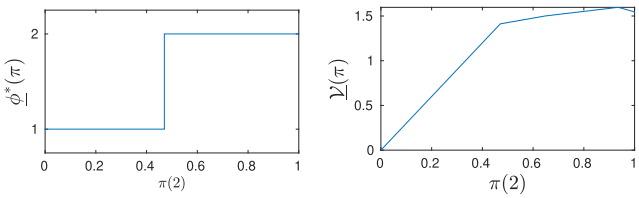
C. Quickest Detection With Anticipatory Agents is Non-Trivial

In classical quickest change detection, the value function is always concave and the optimal stopping region is convex, see [36] for a partially observed Markov decision formulation and proof of this. The aim of this section is to show that due to the interaction of local and global decision makers, quickest detection with anticipatory agents exhibits non-trivial behavior: the value function is not necessarily concave and the stopping region is not necessarily a convex set.

Consider the value iteration algorithm (38) which is used as a basis for mathematical induction to prove properties associated with Bellman's equation (36). Note that from (38), $\mathcal{V}_k(\pi, s)$ is positively homogeneous, that is, for any $\alpha > 0$, $\mathcal{V}_k(\alpha\pi, s) = \alpha\mathcal{V}_k(\pi, s)$. So choosing $\alpha = \sigma(\pi, a)$ yields

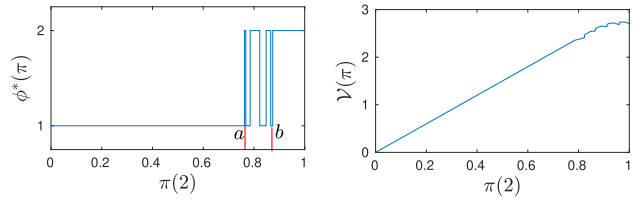
$$\mathcal{V}_{k+1}(\pi, s) = \min \Bigg\{ C(\pi, 1) + \int_{\mathcal{S}} \sum_a \sum_{l=1}^{Y+1} \mathcal{V}_k(R_a^l(s) P' \pi, s) I(\pi \in \mathcal{P}_l) p(\bar{s}|s) d\bar{s}, C(\pi, 2) \Bigg\} \tag{44}$$

Recall $C(\pi, 1)$ and $C(\pi, 2)$ are linear in π. However, it is clear from (44) that if $\mathcal{V}_k(\pi, s)$ is assumed to be concave on Π, $\mathcal{V}_{k+1}(\pi, s)$ is not necessarily concave on Π; since patching together convex functions on different intervals does not necessarily yield a convex function. The key point is that the action likelihoods R^π (47) are explicit and discontinuous functions of π. This results in a possibly non-concave value function $V(\pi)$ making the stopping set $\mathcal{S}$ non-convex.



(a) Optimal Policy (b) Value Function

Fig. 4. Classical Quickest Detection. The optimal policy $\underline{\phi}^*(\pi)$ has a threshold structure. So the optimal stopping set $\mathcal{S} = \{\pi : \underline{\phi}^*(\pi) = 2\}$ is convex. The value function $\underline{\mathcal{V}}(\pi)$ is concave.



(a) Optimal Policy (b) Value Function

Fig. 5. Quickest Detection with Multiple Agents. The optimal policy $\phi^*(\pi)$ has a multi-threshold structure implying that the optimal stopping set $\mathcal{S} = \{\pi : \phi^*(\pi) = 2\}$ is not convex (comprises of disconnected regions). The global decision maker exhibits change blindness. As the posterior probability of change $\pi(1) = 1 - \pi(2)$ increases, the global decision maker declares there is no change in several regions (between the red lines). Note that the value function $\mathcal{V}(\pi)$ is not concave.

D. Numerical Example of Multi-Threshold Quickest Detection Policy: Change-Blindness

The non-concave value function in quickest detection with anticipatory agents leads to unusual multi-threshold behavior in the optimal policy, as we now illustrate.

1) Setup: Consider quickest detection where the state of nature $\{x_n, n \geq 0\}$ jumps according to transition matrix

$$P = \begin{bmatrix} 1 & 0 \\ 0.05 & 0.95 \end{bmatrix}. \tag{45}$$

The global decision maker's delay and false alarm penalties are $d = 1.05, f = 3$; these specify the costs (33), (34) in Bellman's equation (36).

The local anticipative decision maker's reward matrix is $(r_2(x, a_2), x \in \{1, 2\}, a \in \{1, 2\}) = \begin{bmatrix} 5 & 4 \\ 6.5 & 9 \end{bmatrix}$ Also its observation likelihood matrix is $B = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$.

2) Nonconvex Stopping Time and Value Function: The local and global decision makers operate according to Protocol 1. Fig. 4 displays the value function and optimal policy for classical quickest detection. Fig. 5 displays the value function and optimal policy for quickest detection with anticipatory agents. The policy and value function were obtained by running the value iteration algorithm for 1000 iterations with $\Pi = [0, 1]$ grid quantized uniformly to 1000 values.

For classical quickest detection, Fig. 4 shows that, as expected, the value function is concave and the optimal policy is a threshold. So the stopping region $\{\pi : \underline{\phi}^*(\pi) = 1\}$ is the interval $\pi(2) \in [0, 0.466]$.

In contrast, for quickest detection involving anticipatory agents, Fig. 5 shows the value function is not concave. Also the optimal policy has an unusual multi-threshold structure: if it is optimal to declare a change for a particular posterior probability, it may not be optimal to declare a change when the posterior probability of change is larger! (Recall $1 - \pi(2)$ is the posterior probability of change). In this sense, Fig. 5 depicts two forms of change-blindness. First, a human global decision maker might choose to ignore the optimal policy $\phi^*(\pi)$ and simply use the classical quickest detection policy $\underline{\phi}^*(\pi)$. A second, and more interesting form of change-blindness occurs when the human global decision maker chooses the "simple" stopping set as $\pi(2) \in [0, a]$ and ignores the important regions between $[a, b]$ where it is optimal to stop.

E. Numerical Example. Change Detection of Betting Strategy

Spot fixing is a form of illegal match fixing where players deliberately under-perform in specific segments of a team sport. Identifying spot fixing in cricket and soccer is important with the advent of live betting. Sudden increases in the betting rate, heavy underdog bets and wide swings in the quality of play can prompt monitors to take a closer look at a match. Quickest change detection of these parameters based on monitoring real time betting is relevant for detecting illegal spot-fixing for example in T20 cricket; see also [40]–[42]. Here we consider a highly simplified formulation where the aim is detect a sudden change in the intrinsic value (state of nature x_n) of the bet possibly due to spot fixing.

1) Model: Suppose each anticipatory betting agent n acts according to Sec. II-C and makes decisions $a_n \in \{\texttt{bet}, \overline{\texttt{bet}}\}$, $\bar{a}_n \in [0, W_0]$. The agents $n = 1, 2, \ldots$ act sequentially according to Protocol 1. Each agent n has access to whether the previous agents placed bets, i.e., agent n knows the actions $\{a_l \in \{\texttt{bet}, \overline{\texttt{bet}}\}, l = 1, \ldots n-1\}$. The state of nature x_n is the underlying value of the bet. Each agent n obtains a noisy value y_n of x_n; this determines its private belief η_n of x_n. As in (12), we assume that each agent is risk averse and we choose its risk averse parameter $\beta = \eta_n(1)$, i.e., the risk averse parameter of agent n is its belief of the underlying value of the bet. The current score in the game (physical state s_n) also affects β; but for simplicity we omit this.

An analyst monitors the betting decisions $\{a_n\}$. Due to privacy constraints, the amounts bet $\{\bar{a}_n\}$ are not known to the analyst. How can the analyst detect a sudden change in the intrinsic value of the bet x_n indicating spot fixing?

We chose the anticipatory model parameters $u_A = 10$, $g = 15$, $\iota = 0$ (recall notation in (15)). The transition probability P for the jump change and observation probabilities B are as Sec. V-D1. The quickest detection penalties are $d = 1$, $f = 10$. The system operates according to Protocol 1.

2) Non-Concave Non-Monotone Value Function: Fig. 6 displays the quickest detection value function $\mathcal{V}(\pi)$ and optimal policy ϕ^* (36). Unlike classical quickest detection the value function is non-concave and not increasing, but the optimal policy (not shown) still has a threshold structure. Even this simplistic example shows a rich variation of the value function as α is adjusted: if $\alpha = 0.5$, the $\mathcal{V}(\cdot)$ is concave; if $\alpha = 1.3$, $\mathcal{V}(\cdot)$.is non-concave with multiple discontinuities.

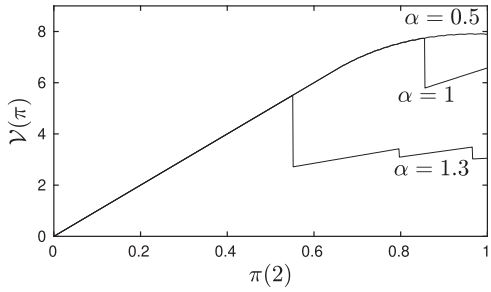
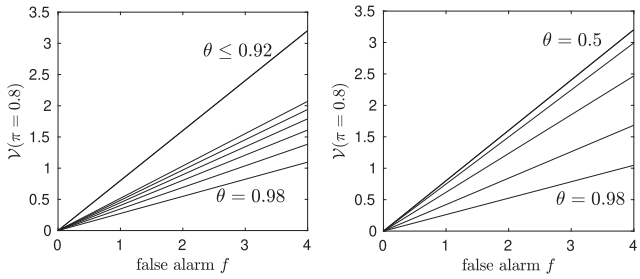


Fig. 6. Non-concave value function for quickest change detection of betting strategy amongst anticipatory agents. The parameters are specified in Sec. V-E.



(a) Multi-agent $\theta \in \{0.92, 0.93, , \ldots, 0.98\}$

(b) Classical $\theta \in \{0.5, 0.6, \ldots, 0.9, 0.98\}$

Fig. 7. Comparison of Optimal Expected Cost for Quickest Detection with Anticipatory Agents vs Classical case

3) Performance of Quickest Detectors: Fig. 7 compares the performance of the quickest detector with anticipatory agents vs the classical quickest detector using the same parameters as above with $\alpha = 1$. The observation probability matrix is $B = \begin{bmatrix} \theta & 1-\theta \\ 1-\theta & \theta \end{bmatrix}$ where parameter θ is varied. The delay penalty is fixed at $d = 1$ while the false alarm $f \in [0.2, 4]$. The optimal expected cost $\mathcal{V}$ was obtained by solving Bellman's equation (36) by quantizing the beliefs to a uniform grid of 1000 points and running 1000 iterations of the value iteration algorithm (38). We chose $\pi = [0.2,\ 0.8]'$ in the plot since the value function $\mathcal{V}(\pi)$ has a discontinuity just after $\pi(2) = 0.8$.

Fig. 7(a) shows that for quickest detection with multiple anticipatory agents, the optimal expected cost increases as θ (accuracy of observations) decreases. Interestingly, the optimal cost remains unchanged for $\theta \in [0.5, 0.92]$, i.e., no matter how accurate the observation probabilities θ are in this range, there is no decrease in cost (improvement in performance) of the optimal detector. For each θ, Fig. 7(b) shows that classical quickest detection has a lower optimal expected cost. For example, all the performance curves (lines) for $\theta = 0.5, 0.6, \ldots .0.9$ in the classical case (Fig. 7(b)) lie below the curve $\theta = 0.92$ in the multi-agent case (Fig. 7(a)). This property will be justified in Theorem 4 below via Blackwell dominance.

F. Blackwell Dominance Implications for Optimal Cost

In this section we show that quickest detection with anticipative agents (Protocol 1) results in a cumulative Kolmogorov Shiryaev cost $J_{\phi^*}(\pi, s)$ (defined in (30) or equivalently (35)) that is always larger than that of classical quickest detection. In Protocol 1, agents have access to the public belief (which depends on local decisions of previous agents) instead of the actual observations. One expects that this information loss results in less efficient quickest time change detection compared to classical quickest detection. Here we confirm this intuition. The main idea involves Blackwell dominance of observation measures. The result is useful because even though explicit computation of the optimal policy for the setup in Protocol 1 is difficult, we can lower bound the optimal achievable cost by that of classical quickest detection.

First define the optimal policy and cost in classical quickest change detection. Similar to (36), the optimal policy $\underline{\phi}^*(\pi)$ and cost $\underline{\mathcal{V}}(\pi)$ incurred in classical quickest detection, satisfy the following stochastic dynamic programming equation:

$$\underline{\phi}^*(\pi) = \arg\min_{u\in\mathcal{U}} \underline{Q}(\pi, u),\ \underline{\mathcal{V}}(\pi) = \min_{u\in\mathcal{U}} \underline{Q}(\pi, u),$$
$$\text{where } \underline{Q}(\pi, 2) = C(\pi, 2) + \sum_{y\in\mathcal{Y}} \underline{\mathcal{V}}\left(T(\pi, y)\right)\sigma(\pi, y),$$
$$\underline{Q}(\pi, 1) = C(\pi, 1), \qquad \underline{J}_{\mu^*}(\pi) = \underline{\mathcal{V}}(\pi). \tag{46}$$

Here $T(\pi, y)$ is the Bayesian filter update defined in (25) 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 (26) or interaction between the public and private beliefs.

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

Theorem 4: Consider the quickest change detection problem involving anticipatory agents described in Protocol 1 and associated value function $\mathcal{V}(\pi, s)$ in (36). Consider also the classical quickest detection problem with value function $\underline{\mathcal{V}}(\pi)$ in (46). 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 $\pi \in \Pi, s \in \mathcal{S}$.

The proof is in the supplementary document. The intuition behind the proof is as follows. From (28)

$$R^{\pi}_{x,a}(s) = \int_{\mathcal{Y}} B_{x,y} M^{\pi}_{y,a,s} dy,$$
$$\text{where } M^{\pi}_{y,a,s} \stackrel{\text{defn}}{=} I(\mu^*_{2,T(\pi,y)}(s, a_{n,1}) = a) \tag{47}$$

where B and M^{π} are stochastic kernels. Thus observation y with conditional distribution specified by B is said to be more informative than (Blackwell dominates) observation a with conditional distribution R^{π}, see [36]. The main idea in the proof is that under the assumptions of Theorem 4, the value function $\underline{\mathcal{V}}(\pi)$ is concave for $\pi \in \Pi$. Then the result is established using Jensen's inequality together with Blackwell dominance on Bellman's equation (36).

A useful consequence of Theorem 4 is that performance analysis of standard quickest detection [16] applies as a lower bound for quickest detection with anticipatory agents.

VI. Discussion

This paper is an early step in addressing sequential detection problems with behavioral economics constraints. Although both signal processing and behavioral economics are mature areas, insights gained by construction of generative anticipatory models, estimation algorithms, along with careful analysis is crucial in designing human-sensor cyber-physical systems.

They main results of the paper are:

1) Formulation of the two stage decision making model of [1] for individual decision makers involving the anticipatory state. The key idea is that the anticipatory state involves the probabilities of future actions thereby leading to time inconsistency in decision making.

2) Characterizing the structure of the subgame Nash equilibrium as a bang-bang controller in the first time stage, and a threshold policy at the second time stage (Theorem 1). The bang-bang structure justifies the observation in [1] that agents with anticipatory emotions may choose to avoid information.

3) Formulation of the multi-agent quickest detection problem where the anticipatory agents interact with a global decision maker. We gave several examples to motivate this problem including change detection in social media accommodation, detecting spot fixing in sports and team-situation awareness.

4) Structural characterization of the unusual structure of the optimal change detection policy (compared to classical quickest detection). Sec. V characterized the structure of the Bayesian belief updates and achievable cost of the quickest detector without brute force computations. We derived important structural properties of the Bayesian updates of the local and global decision makers (Theorem 2 and Theorem 3), constructed a lower bound for the optimal cost incurred using Blackwell dominance (Theorem 4), and presented numerical examples of the unusual structure of the optimal quickest change policy (non-convex stopping region). The multi-threshold change detection policy was interpreted as *change blindness*, namely people fail to detect surprisingly large changes to scenes.

In future work we will generalize the anticipatory model using the subjective belief multi-horizon formulation of [21]. It is also worthwhile conducting a performance analysis of a multi-threshold detector; see [16] for performance analysis involving a single threshold detector. An important open question is: based on a dataset of actions of an agent, how to identify anticipatory behavior and if so, how to estimate the utility function of an anticipatory agent (inverse reinforcement learning)? For myopic Bayesian utility maximization, [43] give necessary and sufficient conditions for identifying optimal behavior; the utility functions then are feasible points of a set of convex constraints. In [44] we have used such methods to analyze user engagement in massive YouTube datasets.

Acknowledgment

The author acknowledges useful discussions with Andrew Caplin, Department of Economics, NYU.

REFERENCES

[1] A. Caplin and J. Leahy, "Psychological expected utility theory and anticipatory feelings," *Quart. J. Econ.*, vol. 116, no. 1, pp. 55–79, 2001.

[2] R. Rosen, "Anticipatory Systems," in *Anticipatory Systems*. Berlin, German: Springer, 2012, pp. 313–370.

[3] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probability Appl.*, vol. 8, no. 1, pp. 22–46, 1963.

[4] R. Bénabou and J. Tirole, "Mindful economics: The production, consumption, and value of beliefs," *J. Econ. Perspectives*, vol. 30, no. 3, pp. 141–64, 2016.

[5] C. J. Charpentier, E. S. Bromberg-Martin, and T. Sharot, "Valuation of knowledge and ignorance in mesolimbic reward circuitry," *Proc. Nat. Acad. Sci.*, vol. 115, no. 31, pp. E7255-E7264, 2018.

[6] J. O. Cook and L. W. Barnes Jr, "Choice of delay of inevitable shock." *J. Abnorm. Social Psychol.*, vol. 68, no. 6, pp. 669–672, 1964.

[7] S. M. Miller and C. E. Mangan, "Interacting effects of information and coping style in adapting to gynecologic stress: Should the doctor tell all?," *J. Pers. Social Psychol.*, vol. 45, no. 1, pp. 223–236, 1983.

[8] M. R. Endsley, *Designing for Situation Awareness: An Approach to User-Centered Design*. Boca Raton, FL, USA: CRC Press, 2016.

[9] E. Blasch, "Enhanced air operations using jview for an air-ground fused situation awareness UDOP," in *Proc. IEEE/AIAA 32nd Digit. Avionics Syst. Conf.*, 2013, pp. 5A 5–1.

[10] G. Klein, D. Snowden, and C. L. Pin, "Anticipatory Thinking," *Informed Knowledge: Expert Performance in Complex Situations*, K. L. Mosier and U. M. Fisher, Eds., New York, NY, USA: Taylor & Francis, pp. 235–245, 2011.

[11] Z. Lanir, "Fundamental Surprise," Eugene, OR: Decision Research, 1986. [Online]. Available: https://www.theisrm.org/documents/Lanir%20(1984)%20Fundamental%20Surprises.pdf

[12] T. Björk and A. Murgoci, "A theory of Markovian time-inconsistent stochastic control in discrete time," *Finance Stochastics*, vol. 18, no. 3, pp. 545–592, 2014.

[13] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.

[14] M. K. Brunnermeier and J. A. Parker, "Optimal expectations," *Amer. Econ. Rev.*, vol. 95, no. 4, pp. 1092–1118, 2005.

[15] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge University Press, 2008.

[16] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic bayesian theory of quickest change detection," *Theory Probability Appl.*, vol. 49, no. 3, pp. 458–497, 2005.

[17] J. C. Gorman, N. J. Cooke, and J. L. Winner, "Measuring team situation awareness in decentralized command and control environments," *Ergonomics*, vol. 49, no. 12-13, pp. 1312–1325, 2006.

[18] T. Erlandsson, T. Helldin, G. Falkman, and L. Niklasson, "Information fusion supporting team situation awareness for future fighting aircraft," in *Proc. 13th Int. Conf. Inf. Fusion*, 2010, pp. 1–8.

[19] B. L. Hooey *et al.*, "Modeling pilot situation awareness," in *Human Modelling in Assisted Transportation*. Berlin, Germany: Springer, 2011, pp. 207–213.

[20] J. Boger, P. Poupart, and J. Hoey, "A decision-theoretic approach to task assistance for persons with dementia," in *Proc. Int. Joint Conf. Artif. Intell.*, 2005, pp. 1293–1299.

[21] M. K. Brunnermeier, F. Papakonstantinou, and J. A. Parker, "Optimal time-inconsistent beliefs: Misplanning, procrastination, and commitment," *Manage. Sci.*, vol. 63, no. 5, pp. 1318–1340, 2017.

[22] C. Chamley, *Rational Herds: Economic Models of Social Learning*. Cambridge University Press, 2004.

[23] V. Krishnamurthy, "Quickest detection POMDPs with social learning: Interaction of local and global decision makers," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5563–5587, Aug. 2012.

[24] V. Krishnamurthy and H. V. Poor, "A tutorial on interactive sensing in social networks," *IEEE Trans. Computat. Social Syst.*, vol. 1, no. 1, pp. 3–21, Mar. 2014.

[25] V. Bordignon, V. Matta, and A. H. Sayed, "Adaptive social learning," 2020, *arXiv:2004.02494*.

[26] J. Tsitsiklis, "Decentralized detection," *Adv. Stat. Signal Process.*, vol. 2, pp. 297–344, 1993.

[27] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[28] D. J. Simons and R. A. Rensink, "Change blindness: Past, present, and future," *Trends Cogn. Sci.*, vol. 9, no. 1, pp. 16–20, 2005.

[29] D. M. Kreps and E. L. Porteus, "Temporal resolution of uncertainty and dynamic choice theory," *Econometrica*, vol. 46, no. 1, pp. 185–200, 1978.

[30] G. F. Loewenstein, E. U. Weber, C. K. Hsee, and N. Welch, "Risk as feelings." *Psychol. Bull.*, vol. 127, no. 2, pp. 267–286, 2001.

[31] P. A. Samuelson, "Probability, utility, and the independence axiom," *Econometrica: J. Econometric Soc.*, vol. 20, no. 4, pp. 670–678, 1952.

[32] D. M. Topkis, *Supermodularity and Complementarity*. Princeton University Press, 1998.

[33] W. A. Brock and R. G. Thompson, "Convex solutions of implicit relations," *Math. Mag.*, vol. 39, no. 4, pp. 208–211, 1966.

[34] T. M. Apostol, "Mathematical Analysis," Addison Wesley, 1974.

[35] M. Avellaneda and S. Stoikov, "High-frequency trading in a limit order book," *Quantitative Finance*, vol. 8, no. 3, pp. 217–224, Apr. 2008.

[36] V. Krishnamurthy, *Partially Observed Markov Decision Processes. From Filtering to Controlled Sensing*. Cambridge University Press, 2016.

[37] O. Hernández-Lerma and J. B. Laserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. New York, NY, USA: Springer-Verlag, 1996.

[38] S. Karlin and Y. Rinott, "Classes of orderings of measures and related correlation inequalities. I. multivariate totally positive distributions," *J. Multivariate Anal.*, vol. 10, no. 4, pp. 467–498, Dec. 1980.

[39] P. Milgrom, "Good news and bad news: Representation theorems and applications," *Bell J. Econ.*, vol. 12, no. 2, pp. 380–391, 1981.

[40] B. Abarbanel and M. R. Johnson, "Esports consumer perspectives on match-fixing: Implications for gambling awareness and game integrity," *Int. Gambling Stud.*, vol. 19, no. 2, pp. 296–311, 2019.

[41] D. Forrest and I. G. McHale, "Using statistics to detect match fixing in sport," *IMA J. Manage. Math.*, vol. 30, no. 4, pp. 431–449, 2019.

[42] H. Qureshi and A. Verma, "It is just not Cricket," in *Match-Fixing in International Sports*. Berlin, Germany: Springer, 2013, pp. 69–88.

[43] A. Caplin and M. Dean, "Revealed preference, rational inattention, and costly information acquisition," *Amer. Econ. Rev.*, vol. 105, no. 7, pp. 2183–2203, 2015.

[44] W. Hoiles, V. Krishnamurthy, and K. Pattanayak, "Rationally inattentive inverse reinforcement learning explains YouTube commenting behavior," *J. Mach. Learn. Res.*, vol. 21, no. 170, pp. 1–39, 2020.

Vikram Krishnamurthy (Fellow, IEEE) received the Ph.D. degree from the Australian National University in 1992. He is currently a Professor with the School of Electrical & Computer Engineering, Cornell University. From 2002–2016, he was a Professor and Canada Research Chair awith the University of British Columbia, Canada. His research interests include statistical signal processing and stochastic control in social networks and adaptive sensing. He was a Distinguished Lecturer for the IEEE Signal Processing Society and Editor-in-Chief of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. In 2013, he was awarded an Honorary Doctorate from KTH (Royal Institute of Technology), Sweden. He is author of the books *Partially Observed Markov Decision Processes* and *Dynamics of Engineered Artificial Membranes and Biosensors* published by Cambridge University Press in 2016 and 2018, respectively.