

# Dynamic Information Design: A Simple Problem on Optimal Sequential Information Disclosure

Farzaneh Farhadi · Demosthenis  
Teneketzis

Received: date / Accepted: date

**Abstract** We study a dynamic information design problem in a finite-horizon setting consisting of two strategic and long-term optimizing agents, namely a principal (he) and a detector (she). The principal observes the evolution of a Markov chain that has two states, one “good” and one “bad” absorbing state, and has to decide how to sequentially disclose information to the detector. The detector’s only information consists of the messages she receives from the principal. The detector’s objective is to detect as accurately as possible the time of the jump from the good to the bad state. The principal’s objective is to delay the detector as much as possible from detecting the jump to the bad state. For this setting, we determine the optimal strategies of the principal and the detector. The detector’s optimal strategy is described by time-varying thresholds on her posterior belief of the good state. We prove that it is optimal for the principal to give no information to the detector before a time threshold, run a mixed strategy to confuse the detector at the threshold time, and reveal the true state afterward. We present an algorithm that determines both the optimal time threshold and the optimal mixed strategy that could be employed by the principal. We show, through numerical experiments, that this optimal sequential mechanism outperforms any other information disclosure strategy presented in the literature. We also show that our results can be extended to the infinite-horizon problem, to the problem where the matrix of transition probabilities of the Markov chain is time-varying, and to the case where the Markov chain has more than two states and one of the states is absorbing.

---

F. Farhadi  
University of Michigan, Ann Arbor, USA  
*Present address:* Imperial College London, London, UK  
E-mail: f.farhadi@imperial.ac.uk

D. Teneketzis  
University of Michigan, Ann Arbor, USA  
E-mail: teneket@umich.edu

## 1 Introduction

The decentralization of information is an inevitable facet of managing a large system. In modern technological systems, agents constantly face the challenge of making decisions under incomplete and asymmetric information. This challenge arises in many different applications, including transportation, cyber-security, communication, energy management, smart grids and E-commerce.

There is a large body of literature on the issues related to decision making in informationally decentralized systems when the agents are cooperative and jointly wish to maximize a social welfare function which is usually the sum of their utilities [1,2]. The problem is more challenging when the agents are selfish and not concerned about the society as a whole, but only aim at maximizing their own utility. Currently there exist two approaches to the design of efficient multi-agent systems with asymmetric information and selfish/strategic agents. In the following, we briefly describe these approaches.

1. *Mechanism Design*. In this approach, each of the strategic agents (e.g. power generating companies in an energy network) is assumed to possess some private information (e.g. their own valuation/utility functions). There is a coordinator (e.g. an Independent System Operator, ISO) who wishes to optimize a network performance metric (e.g. social welfare) which depends on the agents' private information. To elicit strategic agents' true information, the coordinator (he) needs to provide incentives to them (e.g. a plan for energy production along with monetary taxes or subsidies) so as to align their strategic objectives (e.g. maximization of a strategic agent's utility) with his own objective (e.g. maximization of social welfare). In this situation, the system's information structure (who knows what and when) is fixed, and the coordinator's goal is to design a mechanism that incentivizes the strategic agents to reveal their information truthfully [3].

2. *Information Design*. In this approach, the coordinator (e.g. a central control center in a metropolitan area) knows perfectly the evolution of the system's state (e.g. the evolution of traffic in the transportation network), but the decision making (e.g. route selection) is done by the strategic agents (e.g. drivers) who have incomplete/imperfect knowledge of the state (e.g. drivers only know the traffic in their immediate vicinity). To incentivize strategic agents to take actions that are desirable for the coordinator, he can provide, sequentially over time, information about the system's state to them (e.g. provide traffic updates to selfish drivers). The goal of information provision/disclosure is the alignment of each agent's objective (e.g. minimization of the average time to reach the destination) with the coordinator's objective (e.g. minimization of the traffic in the roads). In this situation, the game-form/mechanism is fixed but the system's information structure is not fixed. It has to be designed by the coordinator through the sequential disclosure/provision of information to the strategic agents so as to serve his goal [4, 5].

This paper addresses an information design problem. Information design problems are also referred to as *Bayesian persuasion* problems because the strategic agents are assumed to understand how information is generated/

manipulated and react to information in a rational (Bayesian) manner. Therefore, information design problems can be seen as persuading Bayesian agents to act in desirable ways. The coordinator who would like to lead other agents to act as he wants is usually referred to as the principal. The system state is the principal's private information which can be used to persuade others to serve his goal.

Information design problems in dynamic environments involving a principal and *long-term-optimizing* strategic agents are challenging (see our discussion in Section 3.2). Currently, very little is known about the design of optimal sequential information disclosure policies in dynamic environments with long-term-optimizing agents. Therefore, we focus on a very simple problem that allows us to highlight how sequential information provision/disclosure strategies can be designed in dynamic environments with long-term optimizing strategic agents.

We consider a version of the quickest detection problem [6] with a strategic principal (that we call “he”) and one strategic agent/detector (that we call “she”). These gender pronouns are arbitrarily assigned only for a clear referral to the agents. The detector wants to detect when a two-state discrete-time Markov chain jumps from the “good” state to the “bad” absorbing state. The detector cannot make direct observations of the Markov chain's state. Instead, she receives, at each time instant, a message from the principal about the Markov chain's state. The principal observes perfectly the evolution of the Markov chain; his objective is to delay, as much as possible, detection of the jump by the detector. At the beginning of the process, the principal commits to a sequential information disclosure strategy/policy which he announces to the detector. The detector's knowledge of the policy shapes her interpretation of the messages she receives. For each fixed sequential information disclosure policy of the principal, the detector is faced with a standard quickest detection problem with noisy observations. The principal's goal is to determine the sequential information disclosure policy that convinces the detector to wait as long as possible before declaring the jumps. A precise formulation of this problem is presented in Sections 2 and 3.

The key features of our problem are:

- The problem has a dynamic nature;
- The principal and the agent/detector are long-term optimizers;
- The principal's private information varies/evolves over time;
- The principal has the power of commitment.

These features are not all simultaneously present in any work available so far (see literature survey in Section 1.1).

In this paper, we discover an optimal sequential information disclosure strategy for the principal. We prove that it is optimal for the principal to give no information to the detector before a time threshold, run a mixed strategy to confuse the detector at the threshold time, and reveal the true state afterwards. We present an algorithm that determines both the optimal time threshold and the optimal mixed strategy that could be employed by the principal.

We compare the performance of the policy we discover with that of other sequential information disclosure policies (see Section 5). The comparison shows that the policy we discover outperforms the sequential information disclosure policies currently available in the literature. We extend our results to the case where the Markov chain has  $n > 2$  states, one of which is absorbing.

The contribution of this paper is three-fold. (i) The novelty of our model (see the features of our model listed above). (ii) The analytical approach to the solution of the dynamic information design problem. This approach is based on technical arguments that are different from the standard concavification method that is commonly used in the information design literature (see literature survey in Section 1.1). (iii) The extension of our results to  $n$ -state Markov chains containing one absorbing state.

## 1.1 Review of Related Works

Information disclosure mechanisms can be seen as a communication protocol between an information transmitter (he) and one or multiple receivers. A significant part of the existing works in this area have studied the nonstrategic case, where the information transmitter and receivers are cooperative and jointly wish to maximize the global utility [7–11].

The strategic case of information disclosure mechanisms (Bayesian Persuasion), where the transmitter (principal) and receiver have misaligned objectives, is in tradition of cheap talk [12–17], and signaling games [18, 19]. In signaling games, the transmitter’s utility depends not only on the receiver’s actions, but also on his type (private information). However, in information design problems neither the principal’s type nor his actions enter his utility directly; they only influence his utility through the effect they have on the receivers’ actions. This feature of the model enables us to investigate/analyze the pure effect of principal’s private information on the receivers’ behaviors.

The main difference between cheap talk and information design is the level of commitment power they give to the principal. In the cheap talk literature, the principal has *no commitment power*; he decides on the information message to be sent after seeing the realization of his private information. The cheap talk model induces a simultaneous game between the principal and the agents. Thus, the main goal of this strand of works is to characterize the (Nash) equilibria of the induced game. Most of the existing work has focused on the static setting [12–15].

The work of [12] shows that when the state of the system (principal’s private information) is one-dimensional, it has compact domain, and both the principal’s and the agent’s utilities have certain properties, stated explicitly in [12], then the principal’s equilibrium strategy employs quantization.

More general models of cheap talk, such as multidimensional sources [13], noisy communication [14] and multiple principals with misaligned objectives [15], have been studied in the literature for static settings. There are a few works that study the dynamic version of the cheap talk communication [16,

17]. These works show that allowing for dynamic information transmission improves the informativeness of communication.

In information design problems the transmitter is endowed with *full commitment power*.

In this model, the transmitter is allowed to send any distribution of messages as a (measurable) function of the history of system state and messages, but he should choose and announce his information disclosure policy before observing the system state and then stay committed to it forever. By committing to an information disclosure policy at the very beginning of the system's operation, the transmitter attempts to persuade all other agents to employ strategies that achieve his objective.

The fact that a player in a game can improve his outcome through commitment was first established in [20–22] (see also [23]). The full commitment assumption holds in many economic ([24–26]) and engineering ([4, 5]) multi-agent systems.

The literature on information design is generally divided into two main categories: *static* and *dynamic* information design problems.

**Static information design problems.** The static version of the problem, where the state of the system is fixed and no time is involved, has been studied extensively in the literature. The authors of [27] consider a problem of static information disclosure where the state of the system is Gaussian and the utilities are quadratic, and show that there exists a linear policy for the principal that leads to a Stackelberg equilibrium. In [28, 29] the authors propose a concavification method for deriving an optimal information provision mechanism in static settings. The concavification approach was first developed by a group of U.S. economists and game theorists, led by R. Aumann and M. Maschler, with the context of the negotiations for the Strategic Arms Limitation Treaty (SALT) between the U.S. and U.S.S.R. (see [30]). The approach was developed for problems that can be modeled as repeated games of incomplete information. For over half a century, this idea played an important role in the analysis of repeated games (see [31, 32] and references therein) but has not been applied to information design problems until 2011 [28].

Following [28], the information design problem has been studied for more general settings, such as costly communication [33], multi-dimensional state [34, 35], multiple principals [36, 37], multiple receivers [38, 39], and receivers with different prior beliefs [40]. A case of information design problem with transfers where the principal can provide both information and monetary incentives to the receiver is also studied in [41]. There is a group of works in static information disclosure which are more applied and aim to understand or improve real-world institutions via information design. Research in this strand includes applications to grading in schools [42], research procurement [43], medical testing [44], price discrimination [26], insurance [45], and routing software [4, 5, 46–48]. A thorough discussion of the literature on information design up until 2018 appears in [49].

**Dynamic information design problems.** The dynamic version of the information design problem, where the informed-player/principal can disclose

his superior information sequentially over time, has recently been attracting rapidly growing interest. In dynamic environments, agents' decisions at each instant of time affect their opponents' decisions, not only at present but also in the future. The problem becomes more tangled if the information transmitter has commitment power. In this case, the information disclosure policy that the principal commits to for the future, has a direct effect on the receivers' estimation of what they can gain in the future, and hence on their current decisions. The interdependency among the agents' decision making processes over time makes the dynamic information design problems very complex and challenging.

Most of the available works avoid this challenge by assuming that the agents are myopic, that is they only look at each instant at the immediate consequence of their actions, ignoring the subsequent (future) effects. In [50] and [51] both the information transmitter and receivers are assumed to be myopic. Under this simplifying assumption, [51] shows that the principal's optimal information disclosure policy is a set of linear functions of the current state.

To make the problem closer to reality, the authors of [52–59] consider the myopic assumption only for the information receivers (In [57–59] the information receiver is myopic as a result of the model and their objective.). This set of works studies the interactions between a long-term-optimizing principal and either a myopic receiver or a sequence of short-lived receivers. In the latter case, at each instant of time a new receiver enters the system, forms her belief about the sender's strategy by observing the history of the past messages, takes her action, and then exits the system. In such a case, since the receivers leave the system after taking only one action, they are not concerned about the subsequent effects of their decisions. Therefore, considering a sequence of short-lived receivers is exactly equivalent to assuming that the receiver is myopic. Under this assumption, [53] proposes a generalization of the concavification method for dynamic information design problems. The authors of [54] show that when the receiver is myopic, the greedy disclosure policy where the principal minimizes the amount of information being disclosed in each stage, under the constraint that it maximizes his current payoff, is optimal in many cases, but not always.

There are only a few papers which study the dynamic information design problem with both long-term-optimizing principal and long-term-optimizing receivers [5, 60–67]. In [60–62] principal is assumed to have no commitment power. This assumption simplifies the problem as in this case, the principal's policy at each time instant affects the receiver's decisions only at the future and not at the past. Although in some of these works communication is costly, they could be seen, due to lack of commitments, as dynamic versions of cheap-talk problem. The authors in [5, 63–67] study the dynamic interactions between a long-term-optimizing principal who has full commitment power and a long-term optimizing receiver. In [63–65] the private information of the principal is considered to be constant and not varying with time. The problem with time-varying private information for the principal is discussed in [5, 66, 67]

for dynamic two-stage settings. The authors of [67] also tackle the information design problem in an *infinite-horizon* setting. In this setting, they first simplify the problem by restricting attention to a special class of information disclosure mechanisms and then characterize a mechanism that improves the principal's utility, but is not always optimal.

There are also a few works that combine mechanism design and information design (See [64, 68–70]). In these papers, the principal chooses both the game form and the information structure so as to persuade the agent to follow the recommendations. The scope of the problems discussed in this line of research is distinctly different from that of our paper.

In our problem we consider information design in a dynamic setting with an arbitrary time horizon and long-term optimizing principal/transmitter and detector/receiver, when the principal has full commitment power and his private information evolves over time. Because of all these features our model and problem are distinctly different from those appearing in the literature on information design that we reviewed above.

## 1.2 Organization of the Paper

The rest of the paper is organized as follows. We present our dynamic information design problem with strategic and long-term-optimizing agents in Section 2. In Section 3, we formulate the principal's problem as a dynamic information design problem and discuss its main features. We describe the optimal sequential information disclosure mechanism we propose for the solution to this problem in Section 4. In Section 5, we show the superiority of our proposed mechanism by comparing it to non-optimal mechanisms that are used in real-world applications. We present extensions of our results to more general settings in Section 6. We conclude our paper in Section 7. The proofs of all the technical results appear in Appendices 1-10.

## 1.3 Notation

We use the following notation. Normal font letters  $x$ , bold font letters  $\mathbf{x}$ , and calligraphic font letters  $\mathcal{X}$  stand for scalars, vectors, and sets, respectively. Scalars with a sub-index, e.g.  $x_t$ , represent elements of the vector with the same name.  $x_{t_1:t_2}$ , with  $t_2 \geq t_1$ , represents a vector consisting of  $t_1$ -th through  $t_2$ -th elements of vector  $x$ , i.e.  $x_{t_1:t_2} = (x_{t_1}, x_{t_1+1}, \dots, x_{t_2})$ . The superscript asterisk (\*) indicates an optimal value of the corresponding variable. A vector to the power of an integer, e.g.,  $\mathbf{x}^n$ , indicates a vector containing  $n$  copies of the original vector  $\mathbf{x}$ . Using this notation, a vector that repeats the same value  $a$ ,  $n$  times, is denoted by  $(a)^n$ . Notation  $((a)^n, (b)^m)$  represents a vector with  $n$  first elements being  $a$  and the next  $m$  elements being  $b$ . A set  $\mathcal{X}$  raised to the power of an integer  $n$ , i.e.,  $\mathcal{X}^n$ , denotes the Cartesian product of set  $\mathcal{X}$  with itself  $n$  times.

The strategy of the principal is denoted by  $\boldsymbol{\rho}$ . For a fixed strategy  $\boldsymbol{\rho}$ , notations  $m^\rho$  and  $\mathbb{P}(A|\boldsymbol{\rho})$  represent the messages sent by the principal based on strategy  $\boldsymbol{\rho}$  and the conditional probability of an event  $A$  given  $\boldsymbol{\rho}$ , respectively. Notation  $\mathbb{E}_{t_1:t_2}^\rho\{f|A\}$  denotes the expected value of function  $f$  from time  $t_1$  to  $t_2$  when the principal commits to strategy  $\boldsymbol{\rho}$  and event  $A$  has occurred.  $1_{\{A\}}$  is the indicator function of an event  $A$ .

## 2 Model

Consider a Markov chain  $\{s_t, t \in \mathcal{T} = \{1, 2, \dots, T\}\}$  with state space  $\{g(\text{good}), b(\text{bad})\}$  and a one-step transition probability matrix

$$P = \begin{pmatrix} 1-q & q \\ 0 & 1 \end{pmatrix}$$

The chain starts in the good state with probability  $\mu$ , i.e.  $\mathbb{P}(s_1 = g) = \mu$ , and then at each instant of time  $t = 2, \dots, T$ , may switch from  $g$  to  $b$  with probability  $q$ . State  $b$  is an absorbing state, meaning that once the Markov chain goes to the bad state, it remains in that state forever. We denote the random time when Markov chain jumps from the good state to the bad state by  $\theta$ . The situation where the Markov chain starts in the bad state is captured by considering  $\theta = 1$ . If the Markov chain remains in the good state until the end of the time period  $T$ , we consider  $\theta = T + 1$ . Therefore, the distribution of the random variable  $\theta$  is as follows:

$$\mathbb{P}(\theta = \theta') = \begin{cases} 1 - \mu, & \text{if } \theta' = 1, \\ \mu(1-q)^{\theta'-2}q, & \text{if } 2 \leq \theta' \leq T, \\ \mu(1-q)^{T-1}, & \text{if } \theta' = T + 1, \end{cases} \quad (1)$$

There is a strategic detector (she) in the system who wants to detect the jump to the bad state as accurately as possible. Let  $\tau$  denote the (random) time the detector declares that the jump has occurred. The detector's cost associated with declaring the jump at time  $\tau$  is

$$J^D(\tau, \theta) = 1_{\{\tau < \theta\}} + 1_{\{\tau \geq \theta\}}c(\tau - \theta), \quad (2)$$

where  $1_{\{A\}}$  is the indicator function of an event  $A$ , which takes value one if  $A$  occurs, and zero otherwise. The detector pays one unit of cost if she declares the jump before it actually happens (i.e. false alarm), and pays  $c$  units of cost per unit of delayed detection. The goal of the detector is to choose a detection time  $\tau$  so as to minimize the expected value of the cost (2). The detector does not observe the Markov chain's state  $s_t$ , but she receives some information about it from another agent called the principal.

At each instant of time  $t$ , the principal (he) observes perfectly the Markov chain's state  $s_t$  and sends, according to some information transmission/ disclosure strategy  $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_T)$ , a message  $m_t$  to the detector. When the detector receives the message  $m_t$ , she updates her belief about the state of the



system in a Bayesian way (using  $\rho$ ), and based on her new belief she decides whether or not to declare that the jump has occurred. Let  $a_t \in \{k, d\}$  denote the detector's action at time  $t \in \mathcal{T}$ , where  $a_t = k$  indicates that the detector keeps silent at time  $t$  and does not declare a jump, and  $a_t = d$  indicates that she declares that the jump has occurred. For any fixed choice of the principal's strategy  $\rho$ , the detector has to solve a quickest detection problem [6] to find her best sequence of actions. Therefore, the detector's optimal strategy at each time instant  $t$  is described by a threshold; these thresholds are time-varying and depend on the choice of the principal's strategy  $\rho$ .

The principal's objective is to delay detection of the jump. Therefore, utilizing his superior information about the state of the Markov chain, the principal attempts to provide informational incentives to the detector so as to persuade her to keep silent. The principal's utility is

$$U^P(\tau) = \tau - 1. \quad (3)$$

**Assumption 1** *The parameters of the model, including transition probability matrix  $P$ , prior belief  $\mu$ , delay cost  $c$ , and action set  $\{k, d\}$  are common knowledge between the principal and the detector. The objectives of the principal and the detector are also common knowledge. The only private information in the model is the Markov chain's state which is perfectly observed only by the principal and not by the detector.*

The detector's decision depends on her belief about the evolution of the system's unknown state  $s_t$ ; this belief depends on the principal's strategy  $\rho$ . Therefore, the principal must design a dynamic (over time) information disclosure mechanism in order to influence the evolution of the detector's beliefs and therefore her sequence of actions.

The above-described model is similar to that of Ely [53] (see Remark 1 below for a comparison between Ely's model and our model). Thus, it is motivated by applications similar to those described in [53], specifically, by the manner the information technology department of an organization (principal) provides information to the organization's employees (detectors) over time so as to influence their productivity, and by the manner a bank (principal) that may be filing releases, over time, information/reports about its financial state to its customers (agents) so as to convince them not to withdraw their savings. The model captures fundamental issues arising in strategic information transmission problems, where one strategic agent (the principal) has superior information about the status of a system as compared to other agents. In these situations, through appropriate observable actions/signaling, the principal attempts to control the other agents' perception of the system's status so as to induce them to act in line with his own interest. Examples of such problems are Cyber-physical systems (power, transportation, security systems) that are vulnerable to adversarial attacks. Adversarial attacks can be viewed as the attacker-defender game where the defender is the principal and the attackers are the agents.

In the next section, we formulate the above-mentioned problem as a dynamic information design, and present a dynamic information disclosure mechanism that maximizes the principal's utility.

*Remark 1* Our model is similar to Ely's model [53] in that they both consider an uninformed agent (detector) who wants to detect the transition of a two-state Markov chain to its absorbing state. In both models, the goal of the information provider is to delay such detection. However, there is a fundamental difference between our model and that of Ely. In Ely's model, the detector uses a time-invariant decision rule that is characterized by a fixed threshold  $p^*$ . Specifically, the detector declares that the jump from the "good" state to the "bad" state occurs, at the first time instant  $\tau$  at which her (posterior) belief that the Markov chain is in the bad state exceeds  $p^*$ . Such a decision strategy is not optimal for either a finite-horizon or an infinite-horizon quickest detection problem. In the finite T-horizon quickest detection problem the detector's optimal decision rule is characterized by a sequence of time-varying thresholds that depend on the parameter  $c$  (See Eq. (2)) and on the functional form of her observations (i.e. the principal's information disclosure strategy). In the infinite-horizon quickest detection problem the detector's optimal decision rule is characterized by a time-invariant threshold as long as the functional form of her observations (i.e. the principal's information disclosure strategy) is time-invariant. It turns out that in our problem as well as in Ely's problem [53] the principal's (optimal) information disclosure strategy is not time-invariant, therefore, the detector's optimal decision strategy is not characterized by a time-invariant threshold.

In our problem the detector's decision rule is the optimal decision rule for the T-horizon quickest detection problem where the functional form of her observations is the principal's optimal information disclosure strategy.

### 3 The Dynamic Information Design Problem

#### 3.1 Problem Formulation

A dynamic information disclosure mechanism specifies the set of messages  $\mathcal{M}_t$  that the principal sends to the detector at each instant of time  $t \in \mathcal{T}$ , along with a distribution over  $\mathcal{M}_t$  given all the information available to the principal at time  $t$ . The principal's information at time  $t$  consists of 1) the history of evolution of the state (which the principal observes perfectly) up to time  $t$ , i.e.  $s_{1:t}$ , 2) the history of his past messages to the detector, i.e.  $m_{1:t-1}$ , and 3) the history of the detector's past actions, i.e.  $a_{1:t-1}$ . The set of dynamic information disclosure mechanisms is completely general, it includes the extremes of full information, no information, and all conceivable intermediate mechanisms. The full information mechanism is the rule that reveals perfectly the current state  $s_t$  : i.e.,  $\mathcal{M}_t = \{g, b\}$  and  $\mathbb{P}(m_t = s_t) = 1$ . A no information mechanism is obtained when the set of messages  $\mathcal{M}_t$  has only one element and the principal sends that single message irrespective of the system state.

As it is clear from the above examples, a dynamic information disclosure mechanism could be very complicated since there is no restriction on the set of messages  $\mathcal{M}_t$  or the probability distribution on  $\mathcal{M}_t$ ,  $t \in \mathcal{T}$ . However, it is shown in [28] that there is no loss of generality in restricting attention to *direct* dynamic information disclosure mechanisms that are *obedient*.

The intuition behind this result is that for any information disclosure mechanism  $\Gamma$ , we can construct an obedient direct information disclosure mechanism that achieves the same performance as  $\Gamma$ . Thus, in this paper, we concentrate on direct dynamic information disclosure mechanisms that are obedient.

In a direct information disclosure mechanism, at each instant of time, the principal directly recommends to the detector the action she should take. In our problem, the detector's possible actions at each time  $t$  are to either keep silent ( $a_t = k$ ) or declare the jump ( $a_t = d$ ). Therefore, the set of messages used by the principal at each time  $t$  is  $\mathcal{M}_t = \mathcal{M} = \{k, d\}$ , where  $k$  is a recommendation to keep silent, and  $d$  is a recommendation to declare a jump. As a result, the principal's behavior in a direct information disclosure mechanism can be described by a recommendation policy  $\boldsymbol{\rho} = (\rho_t^{s_{1:t}, m_{1:t-1}, a_{1:t-1}}, t \in \mathcal{T})$ , where  $\rho_t^{s_{1:t}, m_{1:t-1}, a_{1:t-1}}$  is the probability according to which the principal sends message  $k$  to the detector (i.e. recommends her to keep silent), when the sequence of the states he has observed up to time  $t$  is  $s_{1:t} = (s_1, \dots, s_t)$ , the history of the past messages he has sent is  $m_{1:t-1} = (m_1, \dots, m_{t-1})$ , and the history of the detector's past actions is  $a_{1:t-1} = (a_1, \dots, a_{t-1})$ . For each  $t \in \mathcal{T}$ ,  $s_{1:t} \in \{g, b\}^t$ , and  $m_{1:t-1}, a_{1:t-1} \in \{k, d\}^{t-1}$ , the principal sends message  $d$  to the detector with probability  $1 - \rho_t^{s_{1:t}, m_{1:t-1}, a_{1:t-1}}$ . There are two features in our problem that help us to simplify the principal's information.

1) At each instant of time, the detector has two actions one of which (i.e. declaring the jump) terminates the whole process. Therefore, making a recommendation at time  $t$  is meaningful for the principal only if the detector kept silent at all the previous times, i.e.  $a_{1:t-1} = (k)^{t-1}$ . Thus,  $a_{1:t-1}$  can be omitted from the principal's information.

2) The bad state of our Markov chain is an absorbing state. Therefore, the state evolution of the Markov chain until time  $t$  is of the form  $s_{1:t} = ((g)^{\theta_t-1}, (b)^{t-\theta_t+1})$ , where  $\theta_t$  could take any integer value between 1 and  $t+1$ . We define  $\theta_t$  as the earliest possible time for the jump based on the principal's information at time  $t$ ; i.e.  $\theta_t = \min(\theta, t+1)$ . The parameter  $\theta_t$  is equal to  $\theta$  if the jump has occurred up to time  $t$ , however it takes value  $t+1$  when the principal finds the Markov chain in the good state at time  $t$ . Because of this feature, we can represent the recommendation policy  $\rho_t^{s_{1:t}, m_{1:t-1}}$  of each time  $t$  by  $\rho_t^{\theta_t, m_{1:t-1}}$ , where  $\theta_t \in \{1, \dots, t+1\}$ .

When the principal designs an information disclosure mechanism, he announces the corresponding recommendation policy  $\boldsymbol{\rho}$  to the detector and commits to it.

The detector is strategic and long-term optimizing; she utilizes the information she receives from the principal to her own advantage and does not necessarily follow the actions recommended by the principal. Therefore, to achieve his goal, the principal must design an information disclosure mecha-

nism that possesses the obedience property, that is, it provides the detector with strong enough incentives to follow his recommendations.

At each time  $t \in \mathcal{T}$ , the long-term-optimizing detector obeys the principal if the recommended action  $m_t$  satisfies the following set inclusion:

$$m_t \in \arg \min_{a_t} \left[ \min_{\gamma_{t+1:T}} \mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}\} \right], \quad \forall t, m_{1:t}, \quad (4)$$

where  $\gamma_{t+1:T}$  denotes her decision strategy profile from time  $t+1$  up to  $T$ , and  $\min_{\gamma_{t+1:T}} \mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}\}$  is her minimum expected continuation cost conditional on her information  $m_{1:t}$  when the principal commits to the information disclosure strategy/mechanism  $\rho$ . Constraint (4) ensures that at each time  $t \in \mathcal{T}$ , following the principal's recommendation  $m_t$  minimizes the detector's expected continuation cost. Therefore, to check the obedience property of a direct information disclosure mechanism at any time  $t$ ,  $t = 1, 2, \dots, T$ , we need to solve the series/sequence of optimization problems

$$\min_{a_t} \left[ \min_{\gamma_{t+1:T}} \mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}\} \right], \quad (5)$$

for all  $m_{1:t}$ . Solving these optimization problems is a challenging and time-consuming task. However, the one-shot deviation principle allows us to derive the obedience constraints by assuming that the detector sticks to the obedient strategy in the future. Specifically, considering a fixed direct information disclosure mechanism, the whole process can be seen as a finite extensive-form game between the detector and nature (principal), where at each stage  $t$  nature sends a signal  $m_t$  to the detector according to  $\rho_t$  and then the detector takes an action  $a_t$ . With this interpretation, the set of obedience constraints (4) is a necessary and sufficient set of conditions for the strategy profile  $a_t = m_t$ , for  $t \in \mathcal{T}$ , to be a subgame-perfect Nash equilibrium (SPNE) [71]. According to the one-shot deviation principle a strategy profile  $\gamma$  is a SPNE if and only if no player can increase their payoffs by deviating from  $\gamma$  for one period and then reverting to the strategy [72]. Therefore, a simpler version of the obedience constraints is as follows:

$$m_t \in \arg \min_{a_t} [\mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}, a_{t+1:T} = m_{t+1:T}^{\rho}\}], \quad (6)$$

for all  $t$  and  $m_{1:t}$ , where the condition  $a_{t+1:T} = m_{t+1:T}^{\rho}$  of the expectation reflects the fact that the detector obeys the recommendations from time  $t+1$  onward, and the notation  $m_{t+1:T}^{\rho}$  indicates that the messages  $m_{t+1:T}$  are generated according to the policy  $\rho$ .

Equation (6) includes  $\sum_{t=1}^T 2^t = 2(2^T - 1)$  constraints. These constraints force the detector to obey the recommendations after each message history  $m_{1:t}$ . However, due to the nature of our problem, the first time the detector is recommended to declare a jump in an obedient mechanism, she will do so and the process will terminate. Therefore, the message histories with at least

one  $d$  in the past are not going to occur and need not be checked. This feature reduces the obedience constraints need to be considered to the following:

$$\begin{aligned} & \mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t} = (k)^t, a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\} \\ & \leq \mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t} = (k)^t, a_t = d, \forall t \in \mathcal{T}, \end{aligned} \quad (7)$$

$$\begin{aligned} & \mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t} = ((k)^{t-1}, d), a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\} \\ & \geq \mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t} = ((k)^{t-1}, d), a_t = d, \forall t \in \mathcal{T}, \end{aligned} \quad (8)$$

( $2T$  constraints).

Now that we derived the obedience constraints, we go through calculating the principal's utility. When the detector follows the recommendation policy  $\rho$ , the expected utility the principal gets is

$$\begin{aligned} \mathbb{E}^{\rho} \{U^P(\tau)\} &= \mathbb{E}^{\rho} \{\tau - 1\} = \mathbb{E}^{\rho} \left\{ \sum_{t=1}^T 1_{\{a_{1:t} = (k)^t\}} \right\} \\ &= \sum_{t=1}^T \mathbb{P}(a_{1:t} = (k)^t | \rho) = \sum_{t=1}^T \mathbb{P}(m_{1:t}^{\rho} = (k)^t) \\ &= \sum_{t=1}^T \sum_{\theta'=1}^{T+1} \mathbb{P}(\theta = \theta') \prod_{t'=1}^t \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}}. \end{aligned} \quad (9)$$

Therefore, we can formulate the information design problem (Problem **P**) for the principal as follows:

$$\begin{aligned} & \max_{\rho} \mathbb{E}^{\rho} \{U^P(\tau)\}, \\ & \text{subject to the obedience constraints (7)-(8),} \\ & 0 \leq \rho_t^{\theta_t, m_{1:t-1}} \leq 1, \forall t, \theta_t, m_{1:t-1}. \end{aligned} \quad (\mathbf{P})$$

That is, the principal wants to choose a feasible dynamic information disclosure mechanism  $\rho$  that satisfies the obedience constraints (7)-(8) and maximizes his expected utility given by (9).

### 3.2 Features of The Problem

Solving the optimization problem (**P**) is a formidable task as the optimization variables are strongly coupled with many non-convex constraints. This strong coupling can be seen by taking a closer look at the expectations appearing in the obedience constraints (7)-(8). According to (2), the detector's expected continuation cost from time  $t$  onward when she has received messages  $m_{1:t}$  and decides to declare the jump at time  $t$  is

$$\mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}, a_t = d\} = \mathbb{P}(s_t = g | m_{1:t}), \quad (10)$$

i.e., the probability of false alarm at  $t$ . If the detector decides to keep silent at time  $t$  and sticks to the obedient strategy at the future, her expected continuation cost is

$$\begin{aligned} & \mathbb{E}_{t:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}, a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\} \\ &= c(1 - \mathbb{P}(s_t = g | m_{1:t})) + \mathbb{E}_{t+1:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}, a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\}, \quad (11) \end{aligned}$$

where the first term on the right hand side is the expected delay cost at time  $t$ , and the second term denotes the expected value of all the future costs from time  $t + 1$  onward. Substituting (10)-(11) in (7)-(8) shows that at each time  $t$ , the detector's decision to obey or disobey the principal's recommendation depends on two factors:

(i) the belief the detector has about the good state of the Markov chain, i.e.  $\mathbb{P}(s_t = g | m_{1:t})$ ; this belief is constructed according to Bayes' rule from the past messages  $m_{1:t}$  she received, hence the past recommendation rules  $\rho_{t'}^{\theta_{t'}, m_{1:t'-1}}$  ( $t' \leq t$ ) that generate these messages;

(ii) the cost the detector expects to incur in the future if she remains silent at time  $t$  and follows the recommendations afterwards, i.e.  $\mathbb{E}_{t+1:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}, a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\}$ . This expected cost depends on the messages the detector expects to receive in the future. These messages depend on the future recommendation rules to which the principal commits, i.e.  $\rho_{t'}^{\theta_{t'}, m_{1:t'-1}}$ ,  $t' > t$ ,  $\forall \theta$ .

Therefore, the detector's decision at each time  $t$  depends not only on the recommendation policy for time  $t$ , but also on the recommendation policies for all times before and after  $t$ . The dependence of the detector's decision at each time  $t$  on the recommendation policy for the entire horizon makes the discovery of an optimal information disclosure mechanism that satisfies the obedience property very challenging. This is mainly because any change in the principal's recommendation policy at any time  $t$  affects the detector's decisions at all times. Therefore, the principal cannot optimize the recommendation policies of different time slots separately just by considering the obedience constraints at that time, instead he needs to optimize the recommendation policies of the whole horizon simultaneously.

#### 4 An Optimal Sequential Information Disclosure Mechanism

In this section we present the main result of the paper, namely, a dynamic information disclosure mechanism that solves the principal's problem, expressed by Problem (P). The mechanism is described in Theorem 1 below. To state Theorem 1 we need the following definitions.

**Definition 1** Define by  $\Gamma^M$  the class of information disclosure mechanisms  $\rho = (\rho_t^{s_t, m_{1:t-1}}, t \in \mathcal{T})$  where  $\rho_t^{s_t, m_{1:t-1}}$  depends on the message profile  $m_{1:t-1}$  received by the detector up to time  $t - 1$  and only the current state  $s_t$  of the Markov chain at time  $t$  (not on the state evolution  $s_{1:t}$ ).

---

**Algorithm 1:** An optimal time-based prioritized information disclosure mechanism

---

```

1 for  $n_p = 1, \dots, T$  do
2    $q_{n_p} \leftarrow \min_{t \leq n_p} \left\{ \frac{1}{\mathbb{P}(\theta \leq n_p)} \left( \frac{\mathbb{P}(\theta > t)}{c} - \sum_{l=t}^{n_p-1} \mathbb{P}(\theta \leq l) \right) \right\};$ 
3   if  $q_{n_p} < 1$  then
4     break;
5   end
6 end
7  $n_p^* \leftarrow n_p;$ 
8  $q_{n_p}^* \leftarrow q_{n_p}.$ 

```

---

**Definition 2** A mechanism  $\rho = (\rho_t^{s_t, m_{1:t-1}}, t \in \mathcal{T}) \in \Gamma^M$  is called *time-based prioritized* if:

- (i)  $\rho_t^{g, m_{1:t-1}} = 1$  for all  $t$  and all  $m_{1:t-1}$ ;
- (ii) there is no time  $t$ ,  $t \in \mathcal{T}$ , such that  $\rho_{t+1}^{b, (k)^t} > 0$  while  $\rho_t^{b, (k)^{t-1}} < 1$ , where  $(k)^t = (k, \dots, k)$  is a vector of length  $t$  with all components equal to  $k$ ; and
- (iii) for all  $t \in \mathcal{T}$ ,  $\rho_t^{b, m_{1:t-1}}$  is arbitrary,  $0 \leq \rho_t^{b, m_{1:t-1}} \leq 1$ , when  $m_{1:t-1} \neq (k)^t$ .

We denote by  $\Gamma^{MP}$  the class of time-based prioritized mechanisms. In information disclosure mechanism  $\rho \in \Gamma^{MP}$ , the priority of keeping the detector silent at each time  $t > \theta$  is higher than keeping her silent at  $t+1$ . Therefore, if  $s_t = b$  the principal does not put any effort in manipulating the detector's information at  $t+1$  unless there is no room for improving his own performance at time  $t$ . As a consequence of Definition 2, for each mechanism  $\rho \in \Gamma^{MP}$ , there is a threshold  $n_p$  such that

$$\rho_t^{b, (k)^{t-1}} = \begin{cases} 1, & t < n_p, \\ 0, & t > n_p, \\ q_{n_p}, & t = n_p, q_{n_p} \in [0, 1]. \end{cases} \quad (12)$$

Therefore, any  $\rho \in \Gamma^{MP}$  can be uniquely described by  $\rho = (n_p, q_{n_p})$ .

**Theorem 1** Without loss of optimality, in Problem (P) the principal can restrict attention to time-based prioritized mechanisms. Determining an optimal time-based prioritized mechanism  $\rho^*$  for Problem (P) is equivalent to finding  $n_p^* + q_{n_p}^* = \max\{n_p + q_{n_p}\}$  such that the mechanism  $(n_p, q_{n_p})$  satisfies all the obedience constraints. The optimal time-based prioritized information disclosure mechanism can be obtained by Algorithm 1 above.

The principal's expected utility at  $\rho^* = (n_p^*, q_{n_p}^*)$  is

$$\mathbb{E}^{\rho^*}\{U^P\} = n_p^* - 1 + \mathbb{P}(\theta \leq n_p^*)q_{n_p}^* + \sum_{t=n_p^*}^T \mathbb{P}(\theta > t). \quad (13)$$

### Outline of the proof of Theorem 1

Theorem 1 provides us with an optimal sequential information disclosure mechanism. We prove this theorem in three steps.

In the first step, we reduce the complexity of the optimization problem **(P)** by reducing/simplifying the domain of  $\boldsymbol{\rho}$ . To this end, we show that:

1) the recommendation policy the principal uses at any time  $t$  when he has advised the detector to declare the jump at least once before  $t$ , plays no role in problem **(P)**;

2) without loss of optimality, the principal can restrict attention to mechanisms  $\boldsymbol{\rho} = (\rho_t^{s_t, m_{1:t-1}}, t \in \mathcal{T}) \in \Gamma^M$ . In this class of mechanisms, the recommendation policy at any time  $t$  depends only on the state  $s_t$  of the Markov chain and the principal's previous messages  $m_{1:t-1}$  (not on the state evolution  $s_{1:t-1}$ ).

In the second step, we prove that  $\rho_t^{g, m_{1:t-1}} = 1$ , for  $t \in \mathcal{T}$  and all  $m_{1:t-1}$ . Thus, when the Markov chain is in the good state, the principal always recommends the detector to wait.

In the third step, we use the results of the first two steps to prove that restricting attention to the class of time-based prioritized mechanisms is without loss of optimality. Then, we determine an optimal solution for the dynamic information disclosure problem **(P)** in this class.

**Proof of Theorem 1:** We prove in the appendix the lemmas appearing in each step.

**Step 1.** As we discussed in Section 3, the principal's behavior in a direct dynamic information disclosure mechanism is described by a recommendation policy  $\boldsymbol{\rho} = (\rho_t^{\theta_t, m_{1:t-1}}, t \in \mathcal{T})$ , where  $\rho_t^{\theta_t, m_{1:t-1}}$  is the probability with which the principal recommends the detector to keep silent at time  $t$ , when  $\theta_t = \min(\theta, t+1)$  and the message profile that has been sent so far is  $m_{1:t-1}$ . For each time  $t$ ,  $\theta_t$  could take any integer value between 1 and  $t+1$ . Moreover, the principal's message may take two values at each time slot, so we can have  $2^{t-1}$  different message profiles at each time  $t$ . Therefore, to design an optimal direct dynamic information disclosure mechanism we need to determine the optimal values of  $\sum_{t=1}^T 2^{t-1}(t+1) = 2^T T$  different variables. This number grows exponentially with the horizon length  $T$ . Furthermore, these variables are coupled with one another through the obedience constraints (7)-(8), hence their optimal values must be determined simultaneously. Since such a determination is a difficult problem, our goal in this step is to reduce the number of design variables, without any loss of optimality. We achieve our goal via the results of Lemmas 1 and 2. Before stating Lemmas 1 and 2, we define the following function that counts the number of time epochs the detector was recommended to declare the jump in the past.

**Definition 3** For each time  $t$  and each message history  $m_{1:t-1}$ , we define  $\mathcal{N}_d(m_{1:t-1})$  as the number of time slots where message  $d$  has been sent to the detector.

**Lemma 1** For each time  $t$ , the recommendation policy  $\rho_t^{\theta_t, m_{1:t-1}}$  needs to be designed only for message profiles  $m_{1:t-1}$  with  $\mathcal{N}_d(m_{1:t-1}) \leq 1$ . The part of



the recommendation policy related to cases where the message  $d$  has been sent more than once has no effect on either the obedience property of the mechanism or the utility it gets to the principal.

As a result of Lemma 1, an optimal recommendation policy must determine the optimal values of  $\sum_{t=1}^T t(t+1) = \frac{1}{3}T(T+1)(T+2)$  variables. This number grows polynomially rather than exponentially with the horizon length  $T$ , so the complexity of the information disclosure problem is significantly reduced.

**Lemma 2** Without loss of optimality, for each message profile  $m_{1:t-1}$ , the principal can restrict attention to recommendation policies that depend, at each time  $t$ , on the current state  $s_t$  of the Markov chain and not on the exact time the jump has occurred.

As a result of Lemma 2, at each time  $t$ , and for each message profile  $m_{1:t-1}$ , the principal needs to consider only two recommendation strategies, one when  $s_t = g$  the other when  $s_t = b$ . Therefore, the total number of design variables is  $\sum_{t=1}^T 2t = T(T+1)$ , which grows as  $T^2$  rather than  $T^3$ .

**Step 2.** We derive an optimal recommendation strategy for the principal when the Markov chain is in the good state.

**Lemma 3** If at any time  $t$  the Markov chain is in the good state, irrespective of the message profile  $m_{1:t-1}$ , it is always optimal for the principal to recommend the detector to keep silent. That is

$$\rho_t^{*g, m_{1:t-1}} = 1, \forall m_{1:t-1}, \forall t \in \mathcal{T}, \quad (14)$$

The result of Lemma 3 is intuitive, because when the state of the Markov chain is good there is no conflict of interest between the principal and the detector. In this state, the principal wants to prevent the detector from declaring a jump, and the detector herself has no incentive to create a false alarm. Therefore, there is no incentive for the principal to mislead the detector.

As a result Lemma 3, when the detector is recommended to declare a jump, she is absolutely sure that the Markov chain is in the bad state, thus, she declares a jump. Therefore, the obedience constraints (8) corresponding to situations where the detector receives recommendation  $m_t = d$  are automatically satisfied and can be neglected in the rest of the design process. Moreover, if any message  $d$  has been sent by the principal in the past, the detector declares a jump right after receiving  $d$  and the whole process terminates. Therefore, we do not need to design a recommendation policy for message profiles that contain at least one  $d$ . Consequently, the only variable we should design at any time  $t$ , is the probability of recommending the detector to declare a jump when  $s_t = b$  and  $m_{1:t-1} = (k)^{t-1}$ . Finding the optimal values of these variables is the subject of the next step.

**Step 3.** First, we show that, without loss of optimality, the principal can restrict attention to time-based prioritized mechanisms. Then, we determine such optimal mechanism.

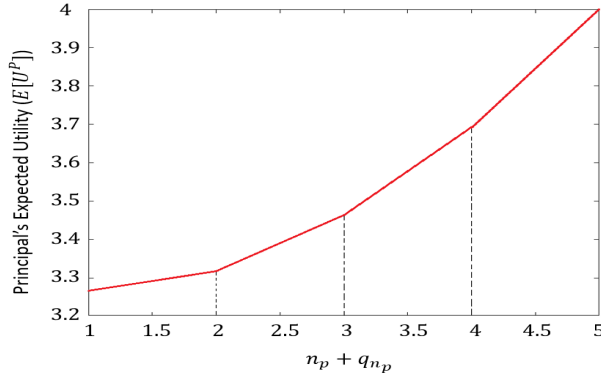


Fig. 1: An example of the Principal's utility function in a time-based prioritized mechanism

**Lemma 4** Without loss of optimality, the principal can restrict attention to the class of time-based prioritized mechanisms.

We proceed now to complete the proof of our main result (Theorem 1). The expected utility of a principal who uses time-based prioritized mechanism  $\boldsymbol{\rho} = (n_p, q_{n_p})$  is

$$\mathbb{E}^{(n_p, q_{n_p})}\{U^P\} = n_p - 1 + \mathbb{P}(\theta \leq n_p)q_{n_p} + \sum_{t=n_p}^T \mathbb{P}(\theta > t). \quad (15)$$

This can be seen as follows. Before the threshold time is reached, the detector is always recommended to keep silent. Therefore, at the first  $n_p - 1$  time slots silence is guaranteed. At the threshold period  $n_p$ , if the Markov chain is in the bad state (prob.  $\mathbb{P}(\theta \leq n_p)$ ), the detector remains silent with probability  $q_{n_p}$ . However, if the Markov chain is in the good state (prob.  $\mathbb{P}(\theta > n_p)$ ), the detector remains silent for sure. At each time  $t$  after the threshold period, the detector keeps quiet only if the jump has not occurred. The probability of this event is  $\mathbb{P}(\theta > t)$ .

Equation (15) shows that for each constant threshold  $n_p$ , the principal's expected utility is an increasing linear function of  $q_{n_p}$ . Moreover, we have  $\mathbb{E}^{(n_p, 1)}\{U^P\} = \mathbb{E}^{(n_p+1, 0)}\{U^P\}$ . This is true, simply because  $\boldsymbol{\rho} = (n_p, 1)$  and  $\boldsymbol{\rho}' = (n_p + 1, 0)$  are actually two representations of the same mechanism. We can easily conclude from these two facts that the principal's expected utility is a piecewise linear function of  $n_p + q_{n_p}$  as depicted in Fig. 1. The slope of the segments increases whenever the variable  $n_p + q_{n_p}$  takes an integer value.

The arguments above show that finding the optimal time-based prioritized mechanism is equivalent to finding the maximum value of  $n_p + q_{n_p}$  such that the mechanism  $\boldsymbol{\rho} = (n_p, q_{n_p})$  satisfies the obedience constraints. With some algebra, it can be shown that for a time-based prioritized mechanism  $\boldsymbol{\rho}$  the

obedience constraints (7) can be simplified to

$$c(\sum_{l=t}^{n_p-1} \mathbb{P}(\theta \leq l) + \mathbb{P}(\theta \leq n_p)q_{n_p}) \leq \mathbb{P}(\theta > t), \forall t \leq n_p. \quad (16)$$

These constraints are very intuitive as for each time  $t$ , 1) the right-hand side of (16) is the expected cost of declaring the jump at time  $t$ ; and 2) the left-hand side of (16) is the expected continuation cost that the detector incurs from time  $t$  onward, when she follows the recommendations made by the mechanism  $\boldsymbol{\rho} = (n_p, q_{n_p})$ . Therefore, constraints (16) simply say that a time-based prioritized mechanism is obedient if and only if the detector finds declaring the jump more costly than keeping silent at each time  $t \leq n_p$  when she is recommended to stay quiet.

The left-hand side of each obedience constraint in (16) is increasing in terms of  $q_{n_p}$ . Therefore, if the obedience conditions are satisfied for a mechanism  $\boldsymbol{\rho} = (n_p, q_{n_p})$ , they are also satisfied for mechanisms with smaller values for  $q_{n_p}$ . Given that the mechanisms  $(n_p, 0)$  and  $(n_p - 1, 1)$  are the same, we can conclude that obedience of a time-based-prioritized mechanism  $\boldsymbol{\rho} = (n_p, q_{n_p})$  implies the obedience of all time-based-prioritized mechanisms with smaller values of  $n_p + q_{n_p}$ . Therefore, there is a threshold  $k^*$  such that the set of all obedient time-based-prioritized mechanisms consists of the mechanisms for which  $n_p + q_{n_p}$  takes a value smaller than the threshold  $k^*$ . The fact that the principal's expected utility is increasing in terms of  $n_p + q_{n_p}$  implies that the mechanism with  $n_p + q_{n_p} = k^*$  is optimal. This optimal mechanism is uniquely determined by Algorithm 1.

Algorithm 1 works as follows: It iterates over  $n_p = 1, \dots, T$  and at each iteration it computes the maximum value of  $q_{n_p}$  such that  $\boldsymbol{\rho} = (n_p, q_{n_p})$  satisfies the obedience constraints (16) for all  $t \leq n_p$ . Achieving a maximum greater than 1 means that the mechanism  $(n_p, 1) = (n_p + 1, 0)$  not only satisfies all the obedience constraints, but also has no binding constraints. This means that there is still more room for improvement. Therefore, the algorithm goes to the next iteration to find a mechanism with the greater utility for the principal. If at some iteration  $n_p^*$  we obtain a maximum of less than one for  $q_{n_p}$  we stop. The mechanism  $(n_p^*, q_{n_p}^*)$  satisfies all obedience constraints and there are binding obedience constraints. Therefore,  $k^* = n_p^* + q_{n_p}^*$  is the optimal threshold which cannot be enhanced anymore, hence  $\boldsymbol{\rho}^* = (n_p^*, q_{n_p}^*)$  is an optimal information disclosure mechanism.

The proof of Theorem 1 is now complete.

## 5 Numerical Results and Discussion

In this section, we discuss and highlight some interesting features of our designed optimal mechanism obtained with Algorithm 1. We also run some numerical experiments to observe its performance.

**Feature 1.** The optimal mechanism  $\rho^* = (n_p^*, q_{n_p}^*)$  we propose is a three-phase mechanism. The principal employs three different strategies in different time regions (See Eq. (12)).

**Region 1:** In the first region which consists of times before the threshold  $n_p^*$ , irrespective of the Markov chain's state  $s_t$ , the principal recommends the detector to keep silent. During this time interval, the principal's messages are independent of the Markov chain's state. These messages give no information to the detector and hence, without loss of optimality, can be removed from the mechanism. Therefore, this region can be referred to as the *no-information region*.

**Region 2:** In the second region which takes only one time slot (i.e.  $t = n_p^*$ ), the principal runs a mixed/randomized strategy to hide his information. In this time slot, the principal always recommends the detector to keep silent if the Markov chain is in the good state. However, when the Markov chain is in the bad state, he reveals this undesirable news only with a probability of  $q_{n_p}^*$ . By employing this strategy, the detector who receives the recommendation to keep silent cannot distinguish whether this recommendation is caused by the truth-telling strategy of the principal in the good state or by his randomized strategy in the bad state. Therefore, she makes a belief about each of these two cases and takes an action that maximizes her expected continuation utility with respect to these beliefs. The probability  $q_{n_p}^*$  is chosen as the maximum probability which makes obedience the best action for the detector. We referred to this region as the *randomized region*.

**Region 3:** In the third region which consists of times after the threshold  $n_p^*$ , the state of the Markov chain can be exactly derived from the principal's messages. In this time interval, a recommendation to keep silent means that the Markov chain is in the good state and a recommendation to declare the jump means that the state of the Markov chain has switched to the bad state. This region can be referred to as the *full-information region*.

Based on the above arguments, we can depict the principal's optimal strategy as in Fig. 2. In this optimal mechanism, the principal does not give any information to the detector up to time  $n_p^* - 1$ , but he promises that if the detector remains silent until that point, then he starts to provide him with "almost accurate" information. The information provided by the principal after time  $n_p^* - 1$  would contain some noise at time  $n_p^*$ , but will be precise and fully revealing of the state after that.

**Feature 2.** In the optimal three-phase mechanism, the principal's commitment to full disclosure of information after a certain time increases the detector's patience and gives her incentives to remain silent longer. To see this we compute the length of time the detector remains silent in two instances: (1) when the principal employs a no-information disclosure strategy for the whole horizon; (2) when the principal commits to full information disclosure some time in the future. In the first instance, the detector's expected cost if

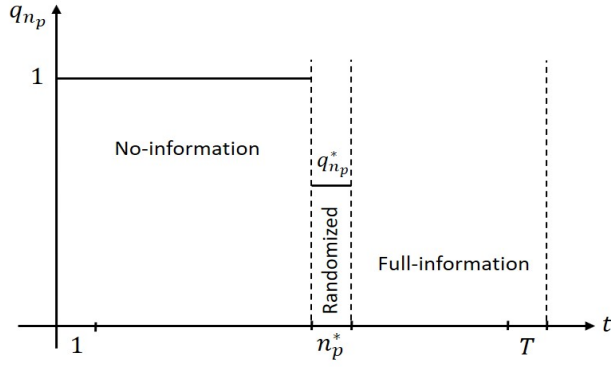


Fig. 2: The principal's strategy in the optimal mechanism

she declares the jump at  $\tau = 1, \dots, T$  is

$$\begin{aligned}
 \mathbb{E}^{No}\{J^D(\tau, \theta)\} &= \mathbb{P}(\theta > \tau) + c \sum_{t=1}^{\tau-1} \mathbb{P}(\theta \leq t) \\
 &= \mu(1-q)^{\tau-1} + c \sum_{t=1}^{\tau-1} (1 - \mu(1-q)^{t-1}) \\
 &= \mu(1-q)^{\tau-1} + c(\tau-1) - c\mu \frac{1 - (1-q)^{\tau-1}}{q}.
 \end{aligned} \tag{17}$$

If the detector keeps silent at the whole horizon, captured by  $\tau = T + 1$ , her expected cost is

$$\mathbb{E}^{No}\{J^D(T+1, \theta)\} = c \sum_{t=1}^T \mathbb{P}(\theta \leq t) = cT - c\mu \frac{1 - (1-q)^T}{q}. \tag{18}$$

Therefore, to minimize her expected cost, the detector with no additional information declares the jump at time

$$\tau^{No} = \arg \min_{\tau=1:T+1} \mathbb{E}^{No}\{J^D(\tau, \theta)\}.^* \tag{19}$$

The case  $\tau^{No} = T + 1$  means that the belief  $\mu$  the detector has in the good state of the Markov chain is high enough that the best action for her is to keep silent over the whole horizon. Using the optimal stopping time  $\tau^{No}$ , we can derive the principal's utility as follows:

$$\mathbb{E}^{No}\{U^P\} = \tau^{No} - 1. \tag{20}$$

---

\*For every fixed initial probability mass function,  $\tau^{No}$  is a deterministic quantity.

	$\eta \geq 1$	$\eta \geq 4$	$\eta \geq 7$	$\eta \geq 10$
$c = 0.1$	57.17%	13.38%	5.64%	2.94%
$q = 0.1$	69.57%	31.12%	8.13%	0%
$\mu = 0.9$	56.65%	12.83%	5.14%	2.39%

Table 1: Probabilities of having a patience enhancement above certain thresholds

The arguments above show that the detector who receives no new information remains silent for  $\tau^{No} - 1$  numbers of time slots. In the optimal time-based-prioritized mechanism (instance 2), the principal's commitment to provide almost accurate information in the future incentivizes the detector to keep silent for  $n_p^* - 1$  periods of time without receiving any new information. The difference  $\eta = n_p^* - \tau^{No}$ , which we call the *patience enhancement*, measures by how much the principal's commitment increases the detector's patience. In Fig. 3, we have plotted this measure for different values of the parameters  $\mu \in [0, 1]$ ,  $q \in [0, 1]$  and  $c \in [0, 1]$ . In each sub-figure, we fix one of the parameters and partition the space of the two remaining parameters in terms of the patience enhancement  $\eta$ . Based on Fig. 3a, the detector's patience increases by at least one time slot in 57.17% of the cases, when  $c = 1$ . For  $c = 1$ , the patience enhancement is at least 4, 7 and 10 time slots in 13.38%, 5.64%, and 2.94% of the cases, respectively. The probabilities of having a patience enhancement above thresholds 1, 4, 7, and 10, are depicted in Table 1 for each of sub-figures 3a-3c. These results show that the principal's commitment to provide almost accurate information in the future can significantly enhance the detector's patience.

**Feature 3.** The mechanism is almost independent of the horizon length  $T$ . The only effect of the horizon length on the optimal mechanism is to limit the threshold  $n_p$  from above. Therefore, as long as the optimal threshold  $n_p^*$  is an interior point, increasing the time horizon  $T$  does not change the optimal mechanism.

Up to now, we have assumed that the horizon length  $T$  is finite, fixed and deterministic. However, Feature 3 allows us to extend our results to both infinite horizon and random finite horizon cases. Below, we discuss the extension of our approach to the infinite horizon case. Then, the applicability of our approach to the random finite horizon setting is derived from a well-known fact that a random horizon can be reformulated as an infinite horizon [73].

According to Feature 3, finding the optimal information disclosure mechanism in infinite horizon setting is equivalent to solving the problem with a sufficiently large horizon length  $\bar{T}$  such that  $n_p^* < \bar{T}$ . To find a large enough horizon length  $\bar{T}$ , we can run Algorithm 1 by considering  $T = \infty$ . Assuming that the algorithm halts and the optimal values  $n_p^*$  and  $q_{n_p}^*$  are found, we can define  $\bar{T} = n_p^* + 1$ . According to Theorem 1, mechanism  $\boldsymbol{\rho}^* = (n_p^*, q_{n_p}^*)$  is an optimal information disclosure mechanism in a finite horizon setting with horizon length  $\bar{T}$ . Since  $n_p^*$  is an interior point in this setting, we can conclude from Feature 3 that increasing the horizon length has no effect on the optimal-

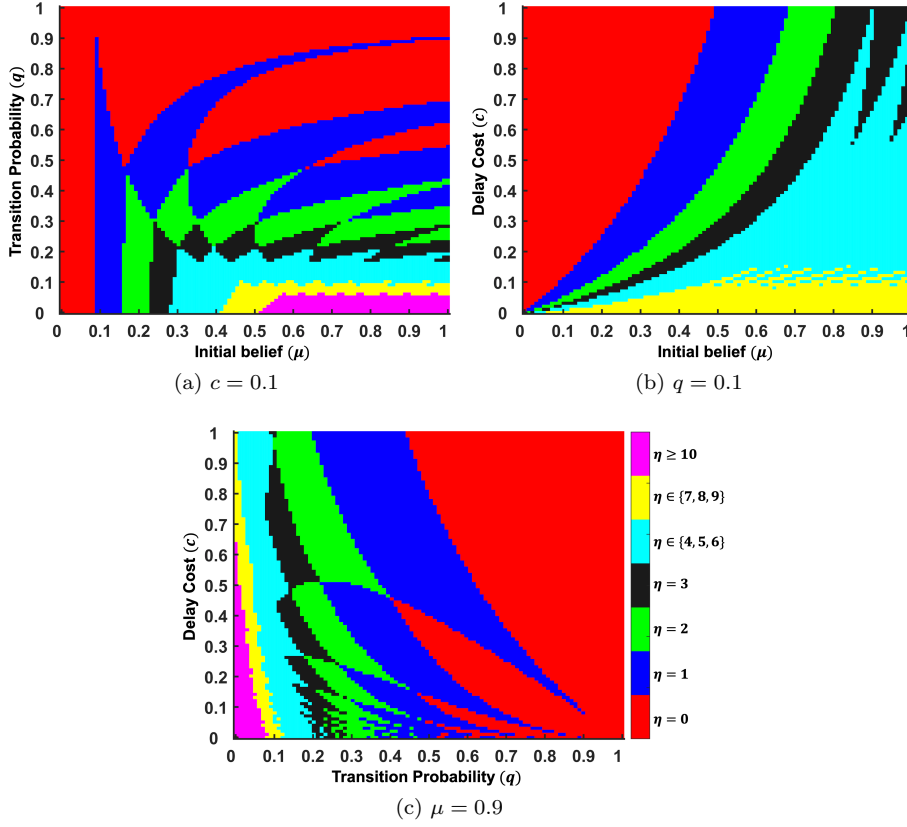


Fig. 3: Color-coded map. Red:  $\eta=0$ , Blue:  $\eta=1$ , Green:  $\eta=2$ , Black:  $\eta=3$ , Cyan:  $4 \leq \eta \leq 6$ , Yellow:  $7 \leq \eta \leq 9$ , Magentic:  $\eta \geq 10$ .

ity conditions and hence  $\rho^* = (n_p^*, q_{n_p}^*)$  is an optimal information disclosure mechanism for the infinite horizon case.

The only remaining part of the proof is to show that Algorithm 1 with  $T = \infty$  always halts. Suppose otherwise, which means that the obedience constraints are satisfied for a time-based prioritized mechanism with  $q_{n_p} = 1$  for all  $n_p \geq 1$ . In such a mechanism, the recommendations are state-independent. Therefore, based on the discussion made in Feature 1, the mechanism is equivalent to the no-information mechanism. It can be seen from (17) that in a no-information mechanism, the detector's expected cost  $\mathbb{E}^{No}\{J^D(\cdot)\}$  goes to infinity when the detection time  $\tau$  goes to infinity. Therefore, it is never optimal for the detector to obey the recommendations that want to keep her silent forever. This contradicts with the assumption that the mechanism is obedient. Therefore, Algorithm 1 always halts and outputs  $n_p^* < \infty$ .

**Feature 4.** In the optimal time-based prioritized mechanism derived by Algorithm 1, we choose the maximum value of  $n_p$  such that there exists a  $q_{n_p} \in [0, 1]$  such that  $\rho = (n_p, q_{n_p})$  satisfies all the obedience constraints for

times before  $n_p$ . To do so, at each round, the algorithm considers a fix value of  $n_p$  and computes the maximum value of  $q_{n_p}$  that together with  $n_p$  satisfies all the obedience constraints of times  $t \leq n_p$ . We can simplify Algorithm 1 by using the result of next lemma.

**Lemma 5** Suppose  $n_p \geq \tau^{No}$ . Let

$$q_{n_p}^t = \frac{1}{\mathbb{P}(\theta \leq n_p)} \left( \frac{\mathbb{P}(\theta > t)}{c} - \sum_{l=t}^{n_p-1} \mathbb{P}(\theta \leq l) \right), \quad (21)$$

denote the maximum value of  $q_{n_p}$  that together with  $n_p$  satisfies the obedience constraint of time  $t$ . Then, we have

$$q_{n_p}^{\tau^{No}} = \min_{t \leq n_p} q_{n_p}^t, \quad (22)$$

where  $\tau^{No}$  is the time at which the detector declares the jump, when the principal employs a no-information strategy.

Lemma 5 intuitively says that persuading the detector to keep silent at time  $\tau^{No}$  is the most difficult challenge faced by the principal. This lemma suggests the following simplification for Algorithm 1. To find the optimal mechanism, we can run Algorithm 1 up to  $n_p = \tau^{No} - 1$ . Then, if the terminating condition (Lines 3-5) has not been satisfied yet, we can replace Line 2 of the algorithm with

$$q_{n_p} = \frac{1}{\mathbb{P}(\theta \leq n_p)} \left( \frac{\mathbb{P}(\theta > \tau^{No})}{c} - \sum_{l=\tau^{No}}^{n_p-1} \mathbb{P}(\theta \leq l) \right), \quad (23)$$

where  $\tau^{No}$  is derived by (19). This simplification reduces the complexity of Algorithm 1, as the algorithm does not need to solve an optimization problem at each round, anymore. Instead, it can solve the optimization problem (19) once and use the result to find the maximum feasible value of  $q_{n_p}$  at each round.

**Feature 5.** In Fig. 3 and Table 1, we showed the superiority of our proposed mechanism compared to the no-information mechanism. In this part, we compare our mechanism with three other benchmark mechanisms, in terms of the expected utility they can provide for the principal.

- The first benchmark is the full information mechanism in which the principal reveals perfectly the Markov chain's state to the detector. The full information mechanism can be considered as a time-based prioritized mechanism with  $n_p = 1$ , and  $q_{n_p} = 0$ . Therefore, we can conclude from (15) and (1) that the expected utility the principal gets if he honestly shares his information with the detector is

$$\mathbb{E}^{Full}\{U^P\} = \sum_{t=1}^T \mathbb{P}(\theta > t) = \sum_{t=1}^T \mu(1-q)^{t-1} = \mu \frac{1-(1-q)^T}{q}. \quad (24)$$



- The second benchmark we consider here, is the best static mechanism that can be employed by the principal. This comparison highlights the power of dynamic mechanisms compared to static ones. In a static mechanism, the set of messages  $\mathcal{M}$  that the principal sends to the detector at each instant of time, as well as the distribution over  $\mathcal{M}$  given the current state is time-independent. By the direct revelation principle, without loss of generality, we can focus on *direct* static mechanisms in which the detector follows the principal's recommendations.

In a direct static mechanism, the principal recommends the detector to keep silent with probability  $\rho_{s_t}$ , where  $s_t \in \{g, b\}$  is the current state of the Markov chain. By an argument similar to that in Lemma 3, we can show that in the optimal static mechanism, we have  $\rho_g = 1$ . Moreover, we can show that the principal's expected utility is an increasing function of  $\rho_b$ . Therefore, the problem of finding the best static mechanism is equivalent to finding the maximum value of  $\rho_b$  such that the mechanism satisfies the obedience constraints. By some algebra, the detector's obedience constraint at each time  $t$  can be derived as follows:

$$\begin{aligned} & \frac{\mu(1-q)^{t-1}}{\mu(1-q)^{t-1} + (1-\mu)\rho_b^t + \sum_{\tau=1}^{t-1} \mu(1-q)^{\tau-1} q \rho_b^{t-\tau}} \\ & \geq c \sum_{l=t}^T \frac{(1-\mu)\rho_b^l + \sum_{\tau=1}^{l-1} \mu(1-q)^{\tau-1} q \rho_b^{l-\tau}}{\mu(1-q)^{t-1} + (1-\mu)\rho_b^t + \sum_{\tau=1}^{t-1} \mu(1-q)^{\tau-1} q \rho_b^{t-\tau}} \end{aligned} \quad (25)$$

where the left-hand side is the average cost of declaring a jump and the right-hand side is the expected cost of keeping silent at time  $t$ , when the detector is recommended to keep silent. We denote the maximum value of  $\rho_b \in [0, 1]$  that satisfies constraint (25) for each  $t \leq T$ , by  $\hat{\rho}$ . Therefore, an efficient static mechanism for the principal is a direct mechanism with  $\rho_g = 1$  and  $\rho_b = \hat{\rho}$ . This mechanism provides the principal with the following expected utility:

$$\begin{aligned} \mathbb{E}^{stat}\{U^P\} &= \sum_{t=1}^T [\mathbb{P}(\theta > t) + \sum_{\theta'=1}^t \mathbb{P}(\theta = \theta') \hat{\rho}^{t-\theta'+1}] \\ &= \sum_{t=1}^T [\mu(1-q)^{t-1} + (1-\mu)\hat{\rho}^t + \sum_{\theta'=2}^t \mu(1-q)^{\theta'-2} q \hat{\rho}^{t-\theta'+1}]. \end{aligned} \quad (26)$$

- The third benchmark is delayed mechanisms. In these mechanisms, the principal's strategy is to reveal the time of the jump to the detector with a fixed delay. Such a mechanism is shown to be optimal in the problem studied by Ely [53]. Ely assumes that the detector is myopic and uses a time-invariant threshold to detect the time of the jump; based on this assumption, he proves that the principal's optimal strategy is to reveal the time of the jump to the detector with a fixed delay. In the comparison of our mechanism with fixed delay mechanisms, we assume that the detector is long-term optimizer

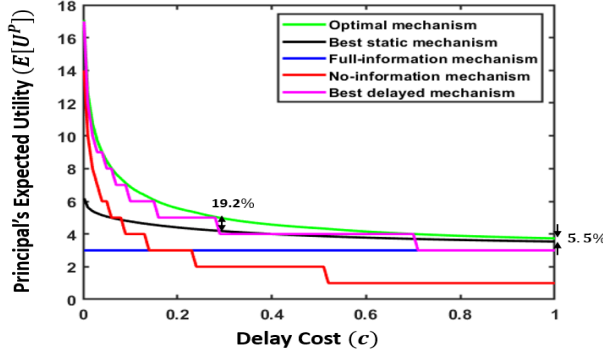


Fig. 4: Principal's expected utility vs. delay cost, for  $\mu = 0.9, q = 0.3, T = 50$ .

and uses the time-varying thresholds of the quickest detection problem corresponding to the principal's strategy. The goal of this comparison is to investigate how much the principal's performance would deteriorate if he restricts attention to the set of delayed mechanisms when the detector is long-term optimizer.

In Fig. 4, we have illustrated the principal's expected utility when he adopts the optimal dynamic, best static, full-information, no-information, and best delayed mechanisms, for different delay costs  $c$ , when  $\mu = 0.9, q = 0.3$  and  $T = 50$ . We observe that while the benchmark mechanisms outperform each other in different regions, the optimal mechanism proposed in this paper always outperforms all of them. In the comparison of different benchmarks, we can see that for low values of delay cost  $c$ , the best delayed mechanism outperforms the other benchmarks. However, when  $c$  goes up, the performance of the best static mechanism is superior to that of the three other benchmarks. We observe from Fig. 4 that the percentage of principal's utility enhancement in our proposed mechanism compared to the best available benchmark is negligible when  $c$  approaches 0, goes up to 19.2% in  $c = 0.29$ , and then approaches to 5.5% when the delay cost reaches one.

**Feature 6.** The three-phase optimal policy is not unique. First, it is a consequence of the fact that we restrict attention to direct obedient information disclosure mechanisms. There may be other indirect information disclosure mechanisms that are optimal but we don't know whether such mechanisms exist. The existence of optimal indirect information disclosure mechanisms is an open problem.

Moreover, even within the class of direct obedient information disclosure mechanisms, there are some optimal mechanisms that are different from the three-phase optimal policy. The main reason for this is as follows. In Definition 2 of the paper, we define the class of time-based prioritized mechanisms as a special class of dynamic information disclosure mechanisms. We prove through Lemmas 2-4 that restricting attention to this class of mechanisms is without

loss of optimality. However, this does not mean that no other optimal direct information disclosure mechanism exists.

To illustrate this fact more clearly, we consider a numerical example with parameters  $q = 0.4$ ,  $\mu = 0.7$ ,  $c = 0.5$ , and  $T = 4$ , and present an optimal mechanism that is different from the three-phase one. In this example, the optimal time-based prioritized mechanism which is derived by running Algorithm 1 is  $\boldsymbol{\rho}^* = (3, 0.3476)$  which provides the principal with expected utility  $\mathbb{E}^{\boldsymbol{\rho}^*}\{U^P\} = 2.6632$ . In mechanism  $\boldsymbol{\rho}^*$ , the principal recommends the detector to keep silent at times  $t = 1$  and  $t = 2$  irrespective of the Markov chain's state, runs a mixed strategy at time  $t = 3$ , and fully reveals the state at time  $t = 4$ . Now, we produce another direct mechanism that is obedient and provides the principal with the same utility. Let  $\boldsymbol{\rho} = (\rho_t^{s_t})$ , where

$$\rho_t^{s_t} = \begin{cases} 1, & \text{if } s_t = g, \\ 1, & \text{if } s_t = b, t = 1, 2, \\ 0.15, & \text{if } s_t = b, t = 3, \\ 0.694, & \text{if } s_t = b, t = 4. \end{cases} \quad (27)$$

It is easy to show that this mechanism satisfies the obedience constraints and maximizes the principal's expected utility. Comparing this mechanism with mechanism  $\boldsymbol{\rho}^*$ , we can see that in  $\boldsymbol{\rho}$ , the principal builds the detector's trust by giving her more accurate information at time  $t = 3$  ( $\rho_3^b < \rho_3^{*b} = 0.3476$ ) and then hides more information from her at time  $t = 4$  ( $\rho_4^b > \rho_4^{*b} = 0$ ). Both of these mechanisms result in the same expected utility  $\mathbb{E}\{U^P\} = 2.6632$ .

The main reason for the existence of optimal policies like (27) is Lemma 4. In the proof of Lemma 4, we showed that for any optimal policy  $\boldsymbol{\rho}$  that splits its obfuscation power into two consecutive time slots  $t$  and  $t + 1$  (i.e.  $\rho_t^b < 1$  and  $\rho_{t+1}^b > 0$ ), we can construct a time-based prioritized optimal policy that first obfuscates the information at time  $t$  as much as it can, and then puts the rest of its obfuscation power, if any, on time  $t + 1$ . It is clear from this discussion that, apart from the time-based prioritized mechanisms, there exist other optimal mechanisms similar to  $\boldsymbol{\rho}$  (27), that split their obfuscation power into several time slots. However, restricting attention to the set of time-based prioritized mechanisms involves no loss of optimality and considerably simplifies the search for an optimal mechanism.

## 6 Extensions

We discuss extension of our results in two directions: 1- When the Markov chain has a time-varying matrix of transition probabilities; 2- When the Markov chain has more than two states and one of the states is absorbing [74]. We conclude the section with a conjecture.

1. So far, for simplicity, we have assumed that the Markov chain's transition probability  $q$  is time-independent. However, the results of the paper

are readily extendable to settings with a time-varying transition probability, i.e.  $q(t)$ . For such settings, all theorems and lemmas presented in the paper still hold. This basically means that, for problems in which the Markov chain's transition matrix varies with time, the principal can still, without loss of optimality, restrict his attention to time-based prioritized mechanisms and the optimal time-based prioritized information disclosure mechanism can be obtained by running Algorithm 1. This can be readily proved by following the same approach as in Section 4.

2. Consider a Markov chain  $\{s_t, t \in \mathcal{T}\}$  with state space  $\{e_1, e_2, \dots, e_n\}$ , and a one-step transition probability matrix

$$P = \begin{pmatrix} 1 & 0 \\ \underline{P}_{n-1 \times 1} & \bar{P}_{n-1 \times n-1} \end{pmatrix}.$$

In this Markov chain, state  $e_1$  is an absorbing state and denotes the state after the jump occurs. This state is equivalent to the bad state in our main model. We denote the initial distribution of the Markov chain, which is common knowledge to both the principal and the detector, by  $\mu = (\mu_1, \mu_2, \dots, \mu_n)$ , where  $\mu_i \geq 0$  and  $\sum_{i=1}^n \mu_i = 1$ . In this setting, the detector's goal is to detect the jump to the absorbing state as accurately as possible, while the principal's goal is to delay detection of the jump.

This problem is more complicated than the two-state case for two reasons:

- (a) The probability of jumping to the absorbing state depends on the current state of the Markov chain which is unknown to the detector. Therefore, the information superiority of the principal in this case is higher than in the two-state case.
- (b) When  $|S| = 2$ , the state evolution of the Markov chain is of the form  $s_{1:t} = ((g)^{\theta_t-1}, (b)^{t-\theta_t+1})$ , where  $\theta_t$  denotes whether the jump has occurred (i.e.  $\theta_t \leq t$ ) or not (i.e.  $\theta_t = t + 1$ ), and if so, when. In this case, as described in detail in Section 3.1, the history of states at any time  $t$  can be expressed by a one-dimensional parameter  $\theta_t$ . However, when the Markov chain has more than two states, this type of simplification is not possible.

In spite of these additional difficulties, the method proposed in Section 4 can be modified as described below to address the general case. First, according to the direct revelation principle, we can restrict attention to direct dynamic information disclosure mechanisms that are obedient. In such a mechanism, at any time  $t$ , the principal directly recommends the detector to either keep silent ( $m_t = k$ ) or declare the jump ( $m_t = d$ ), and the detector must be incentivized to follow all the principal's recommendations. We describe the principal's strategy by a recommendation policy  $\rho = (\rho_t^{s_{1:t}, m_{1:t-1}}, t \in \mathcal{T})$ , where  $\rho_t^{s_{1:t}, m_{1:t-1}}$  is the probability according to which the principal recommends the detector to keep silent, when the state and message histories are  $s_{1:t}$  and  $m_{1:t}$ , respectively.

Considering direct dynamic information disclosure mechanisms and following an approach similar to that of Section 4, we can prove that it is always optimal for the principal to recommend to the detector to keep silent

when the jump has not yet occurred, i.e.  $\rho_t^{s_{1:t}, m_{1:t}} = 1$ , when  $s_t \neq e_1$ . We present the details of the proof of this result in Appendix 10. Therefore, designing an optimal information disclosure mechanism is equivalent to determining the optimal values of the parameters  $\rho_t^{s_{1:t}, m_{1:t}}$ , when  $s_t = e_1$ . These probabilities could potentially depend on the evolution path of the states (i.e.  $s_{1:t}$ ). However, since the principal uses the same strategy in all non-absorbing state, the detector's belief about the Markov chain's state is only a function of the jump time, and not exclusively related to the non-absorbing states the Markov chain has visited. Moreover, the visited non-absorbing states do not affect the future transition probabilities, as the Markov chain has already jumped to state  $e_1$  and remains there forever. Therefore, when the Markov chain is in the absorbing state, the history of the states, apart from the jump time, do not affect either the detector's belief about the past or the expectation (of the principal and the detector) about the future. Therefore, without loss of optimality, the principal can restrict attention to recommendation policies that depend only on the time of the jump (and not on the state evolution  $s_{1:t}$ ). This feature suffices for Lemmas 1-4 in Section 4 to be applicable to the general case of the problem. These lemmas allow the principal to further restrict attention to the class of time-based prioritized mechanisms.

Using this result, we can derive an algorithm, similar to Algorithm 1 of Section 4, to determine the optimal time-based prioritized mechanism that maximizes the principal's utility. The steps of both algorithms are the same; the only difference is in the probabilities that appear in line 2 of Algorithm 1. For the multi-state case, these probabilities are different from those of the two-state case (they are still common knowledge between the principal and the detector).

We conclude our discussion with the following conjecture.

*Conjecture 1* The method presented in this paper can be used to derive close-to-optimal information disclosure mechanisms for problems with general jump processes (i.e. processes that cannot be modeled as a Markov chain).

The reason we believe this conjecture to be true is the following. For Markov chains with an arbitrary number of states one of which is absorbing the distribution of the jump time into the absorbing state is a phase type (PH) distribution [74]. The family of all PH-distributions forms a dense subset of the set of all distributions [74], hence it can be used to approximate jump times into the absorbing state with an arbitrary distribution. Using this approximation and the method presented in this paper, we can design information disclosure mechanisms for general jump processes.

## 7 Conclusion

We studied a dynamic Bayesian persuasion problem whereby a strategic principal observes the evolution of a Markov chain and designs a recommendation

policy that generates a recommendation to a strategic detector at each time. The goals of the principal and the detector are different, therefore, the long-term-optimizing detector does not have to obey the principal's recommendations unless she is convinced to do so. We presented a sequential recommendation policy that maximizes the principal's utility, and ensures the detector's obedience. We proved that the optimal policy is a threshold type, with two thresholds that can be explicitly computed. As time goes by, the optimal recommendation strategy first shifts from a no-information type to a randomized type and then switches to a full-information type.

## ACKNOWLEDGMENT

This work was supported in part by Army Research Office (ARO) grant W911NF-17-1-0232. The authors wish to thank the reviewers for their comments which have significantly improved the presentation of the results appearing in this paper.

## APPENDICES

### APPENDIX 1: Proof of Lemma 1

We want to show that at each time  $t$ , part of the recommendation policy  $\rho_t^{\theta, m_{1:t-1}}$  with  $\mathcal{N}_d(m_{1:t-1}) > 1$  does not appear on either the obedience constraints or the principal's utility function. We first look at the obedience constraints (7)-(8). Substituting (10) and (11) in (7)-(8) shows that inequalities

$$\begin{aligned} c + \mathbb{E}_{t+1:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t} = (k)^t, a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\} \\ \leq (1+c) \mathbb{P}(s_t = g | m_{1:t} = (k)^t), \forall t \in \mathcal{T}, \end{aligned} \quad (28)$$

and

$$\begin{aligned} (1+c) \mathbb{P}(s_t = g | m_{1:t} = ((k)^{t-1}, d)) \\ \leq c + \mathbb{E}_{t+1:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t} = ((k)^{t-1}, d), a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\}, \forall t \in \mathcal{T}, \end{aligned} \quad (29)$$

are equivalent to the obedience constraints (7)-(8). The conditional probabilities of the Markov chain being in the good state given the histories of messages  $m_{1:t} = (k)^t$  and  $m_{1:t} = ((k)^{t-1}, d)$  can be derived as

$$\mathbb{P}(s_t = g | m_{1:t} = (k)^t) = \frac{\mathbb{P}(\theta > t) \prod_{t'=1}^t \rho_{t'}^{t'+1, (k)^{t'-1}}}{\sum_{\theta'=1}^{T+1} \mathbb{P}(\theta = \theta') \prod_{t'=1}^t \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}}}, \quad (30)$$

and

$$\begin{aligned} & \mathbb{P}(s_t = g | m_{1:t} = ((k)^{t-1}, d)) \\ &= \frac{\mathbb{P}(\theta > t) \prod_{t'=1}^{t-1} \rho_{t'}^{t'+1, (k)^{t'-1}} (1 - \rho_t^{t+1, (k)^{t-1}})}{\sum_{\theta'=1}^{T+1} \mathbb{P}(\theta = \theta') \prod_{t'=1}^{t-1} \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}} (1 - \rho_{t'}^{\min(\theta', t+1), (k)^{t-1}})}, \end{aligned} \quad (31)$$

respectively. We can also derive the expected value of the detector's future costs from time  $t+1$  onward when she obeys the recommendations as

$$\begin{aligned} & \mathbb{E}_{t+1:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t}, a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\} \\ &= \sum_{l=t+1}^T \mathbb{P}(s_l = g, m_{t+1:l-1}^{\rho} = (k)^{l-t-1}, m_l^{\rho} = d | m_{1:t}) \\ &+ \sum_{l=t+1}^T c \mathbb{P}(s_l = b, m_{t+1:l}^{\rho} = (k)^{l-t} | m_{1:t}), \end{aligned} \quad (32)$$

where the first term is the probability of occurring a false alarm and the second term is the expected cost of delay in the detection of the jump. Using Bayes' rule to calculate the probabilities appearing in (32) for  $m_{1:t} = (k)^t$  and  $m_{1:t} = ((k)^{t-1}, d)$ , we obtain

$$\begin{aligned} & \mathbb{E}_{t+1:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t} = (k)^t, a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\} \\ &= \sum_{l=t+1}^T \frac{\mathbb{P}(\theta > l) \prod_{t'=1}^{l-1} \rho_{t'}^{t'+1, (k)^{t'-1}} (1 - \rho_l^{l+1, (k)^{l-1}})}{\sum_{\theta'=1}^{T+1} \mathbb{P}(\theta = \theta') \prod_{t'=1}^t \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}}} \\ &+ \sum_{l=t+1}^T c \frac{\sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=1}^l \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}}}{\sum_{\theta'=1}^{T+1} \mathbb{P}(\theta = \theta') \prod_{t'=1}^t \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}}}, \end{aligned} \quad (33)$$

and

$$\mathbb{E}_{t+1:T}^{\rho} \{J^D(\tau, \theta) | m_{1:t} = ((k)^{t-1}, d), a_t = k, a_{t+1:T} = m_{t+1:T}^{\rho}\} = \frac{A}{B}, \quad (34)$$

where

$$\begin{aligned} A &= \sum_{l=t+1}^T \mathbb{P}(\theta > l) \prod_{t'=1}^{t-1} \rho_{t'}^{t'+1, (k)^{t'-1}} (1 - \rho_t^{t+1, (k)^{t-1}}) \prod_{t'=t+1}^{l-1} \rho_{t'}^{t'+1, (k)^{t'-1}} (1 - \rho_l^{l+1, (k)^{l-1}}) + \\ &\sum_{l=t+1}^T \sum_{\theta'=1}^l c \mathbb{P}(\theta = \theta') \prod_{t'=1}^{t-1} \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}} (1 - \rho_t^{\min(\theta', t+1), (k)^{t-1}}) \prod_{t'=t+1}^l \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}}, \\ B &= \sum_{\theta'=1}^{T+1} \mathbb{P}(\theta = \theta') \prod_{t'=1}^{t-1} \rho_{t'}^{\min(\theta', t'+1), (k)^{t'-1}} (1 - \rho_{t'}^{\min(\theta', t+1), (k)^{t-1}}), \end{aligned} \quad (35)$$

and  $(k)_{-t}^{t'}$  is a vector of length  $t'$  where all the components except the  $t$ -th one equal  $k$ , and the  $t$ -th component is  $d$ . Substituting (30)-(34) in (28)-(29), we can see that the terms of the recommendation policy that appear in the obedience constraints have at most one  $d$  in their message history. Furthermore, (9) shows that the principal's utility function depends only on the recommendation probabilities  $\rho_t^{\theta, m_{1:t-1}}$  where the message history is  $m_{1:t-1} = (k)^{t-1}$ . Hence the proof of Lemma 1 is complete.

## APPENDIX 2: Proof of Lemma 2

For each time  $t$ ,  $\theta = 1, \dots, t$  means that the jump has occurred and the state of the Markov chain becomes bad, i.e.  $s_t = b$ ; furthermore,  $\theta = t + 1$  captures the fact that the jump has not occurred yet and the state is still good, i.e.  $s_t = g$ . Therefore, to prove Lemma 2, we need to show that it is without loss of optimality if the principal restricts attention to recommendation policies with the same  $\rho_t^{\theta, m_{1:t-1}}$  for all  $\theta = 1, \dots, t$ . We claim that in this class of mechanisms, the recommendations do not depend on the exact time of the jump, but only on whether it has occurred. To prove the claim, we show that for each  $t \in \mathcal{T}$  and each  $m_{1:t-1}$  with  $\mathcal{N}_d(m_{1:t-1}) \leq 1$ , the parameters  $\rho_t^{\theta, m_{1:t-1}}$  with  $\theta = 1, \dots, t$  always appear in the obedience constraints and the principal's utility function as parts of a constant linear combination. Therefore, as long as the value of this linear combination is fixed, the values of the individual parameters can be customized. As a result, for each optimal mechanism, we can construct another optimal mechanism in which for each  $t$  and each  $m_{1:t-1}$ , the values of all parameters  $\rho_t^{\theta, m_{1:t-1}}$ , with  $\theta = 1, \dots, t$ , are the same.

We proceed by backward induction on time  $t$ .

**Basis of induction.** Let  $t = T$ . The recommendation policy of the final time  $T$  appears in the obedience constraints of time  $T$  and of times  $t' < T$ . We start our investigation by studying the obedience constraints of time  $T$ . According to (7)-(11) and due to the fact that there is no future after time  $T$ , the obedience constraints of time  $T$  are as follows:

$$c(1 - \mathbb{P}(s_T = g | m_{1:T} = (k)^T)) \leq \mathbb{P}(s_T = g | m_{1:T} = (k)^T), \quad (36)$$

and

$$\mathbb{P}(s_T = g | m_{1:T} = ((k)^{T-1}, d)) \leq c(1 - \mathbb{P}(s_T = g | m_{1:T} = ((k)^{T-1}, d))). \quad (37)$$

By substituting (30) and (31) in (36) and (37), respectively, and canceling out the denominators, we can rewrite the obedience constraints of time  $T$  as follows:

$$c \left[ \sum_{\theta'=1}^T \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^T \rho_{t'}^{\theta', (k)^{t'-1}} \right] \leq \mathbb{P}(s_T = g) \prod_{t'=1}^T \rho_{t'}^{t'+1, (k)^{t'-1}}, \quad (38)$$



and

$$\begin{aligned} & \mathbb{P}(s_T = g) \prod_{t'=1}^{T-1} \rho_{t'}^{t'+1, (k)^{t'-1}} (1 - \rho_T^{T+1, (k)^{T-1}}) \\ & \leq c \left[ \sum_{\theta'=1}^T \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^{T-1} \rho_{t'}^{\theta', (k)^{t'-1}} (1 - \rho_T^{\theta', (k)^{T-1}}) \right]. \end{aligned} \quad (39)$$

It can be seen that in both constraints, the parameters  $\rho_T^{\theta, (k)^{T-1}}$  for  $\theta = 1, \dots, T$  only appeared in the following linear combination

$$\sum_{\theta'=1}^T \rho_T^{\theta', (k)^{T-1}} \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^{T-1} \rho_{t'}^{\theta', (k)^{t'-1}}, \quad (40)$$

which we denote by  $L(\rho_T^{1, (k)^{T-1}}, \dots, \rho_T^{T, (k)^{T-1}})$ .

Now we investigate how the parameters  $\rho_T^{\theta, (k)^{T-1}}$  for  $\theta = 1, \dots, T$  appear in the obedience constraints of time  $t < T$ . According to (7)-(8), the obedience constraints of time  $t < T$  are as follows:

$$\begin{aligned} & c \mathbb{P}(s_t = b | m_{1:t} = (k)^t) \\ & + \sum_{l=t+1}^T \mathbb{P}(s_l = g, m_{t+1:l-1} = (k)^{l-t-1}, m_l = d | m_{1:t} = (k)^t) \\ & + \sum_{l=t+1}^T c \mathbb{P}(s_l = b, m_{t+1:l} = (k)^{l-t} | m_{1:t} = (k)^t) \leq \mathbb{P}(s_t = g | m_{1:t} = (k)^t), \end{aligned} \quad (41)$$

and

$$\begin{aligned} & \mathbb{P}(s_t = g | m_{1:t} = ((k)^{t-1}, d)) \leq c \mathbb{P}(s_t = b | m_{1:t} = ((k)^{t-1}, d)) \\ & + \sum_{l=t+1}^T \mathbb{P}(s_l = g, m_{t+1:l-1} = (k)^{l-t-1}, m_l = d | m_{1:t} = ((k)^{t-1}, d)) \\ & + \sum_{l=t+1}^T c \mathbb{P}(s_l = b, m_{t+1:l} = (k)^{l-t} | m_{1:t} = ((k)^{t-1}, d)). \end{aligned} \quad (42)$$

We note that the distribution of  $s_t$  given the messages received up to time  $t$  does not depend on the future policies  $\rho_{t+1:T}$ . For each  $l < T$ , the joint distribution of the state  $s_l$  and the messages received between  $t+1$  and  $l$ , does not either depend on the recommendation policy of time  $T$ . Therefore, the only terms in (41) and (42) that depend on the recommendation policy of time  $T$  are the terms appear in the summations with index  $l = T$ . The joint probability that the state  $s_T$  is good and a certain message sequence  $m_{t+1:T}$  happens between  $t+1$  and  $T$  depends on the recommendation policy the

principal adopts at time  $T$  when the jump has not occurred, that is  $\theta = T + 1$ . Therefore, the only terms in the obedience constraints (41) and (42) that depend on parameters  $\rho_T^{\theta, m_{1:T-1}}$ ,  $\theta = 1, \dots, T$ , are

$$\begin{aligned} & \mathbb{P}(s_T = b, m_{t+1:T} = (k)^{T-t} | m_{1:t} = (k)^t) \\ &= \frac{1}{\mathbb{P}(m_{1:t} = (k)^t)} \left( \sum_{\theta'=1}^T \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^T \rho_{t'}^{\theta', (k)^{t'-1}} \right) \end{aligned} \quad (43)$$

and

$$\begin{aligned} & \mathbb{P}(s_T = b, m_{t+1:T} = (k)^{T-t} | m_{1:t} = ((k)^{t-1}, d)) = \frac{1}{\mathbb{P}(m_{1:t} = ((k)^{t-1}, d))} \\ & \left( \sum_{\theta'=1}^{t-1} \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^{t-1} \rho_{t'}^{\theta', (k)^{t'-1}} (1 - \rho_{t'}^{\theta', (k)^{t'-1}}) \prod_{t'=t}^T \rho_{t'}^{\theta', (k)^{t'-1}} \right. \\ & + \mathbb{P}(\theta = t) \prod_{t'=1}^{t-1} \rho_{t'}^{t'+1, (k)^{t'-1}} (1 - \rho_{t'}^{t'+1, (k)^{t'-1}}) \prod_{t'=t}^T \rho_{t'}^{t, (k)^{t'-1}} \\ & \left. + \sum_{\theta'=t+1}^T \mathbb{P}(\theta = \theta') \prod_{t'=1}^{t-1} \rho_{t'}^{t'+1, (k)^{t'-1}} (1 - \rho_{t'}^{t'+1, (k)^{t'-1}}) \prod_{t'=t+1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^T \rho_{t'}^{\theta', (k)^{t'-1}} \right), \end{aligned} \quad (44)$$

respectively. Equation (43) is a multiple of  $L(\rho_T^{1, (k)^{T-1}}, \dots, \rho_T^{T, (k)^{T-1}})$ . Therefore, we can change the individual values of  $\rho_T^{\theta, (k)^{T-1}}$ ,  $\theta = 1, \dots, T$ , without violating the obedience constraint (41), as long as the value of the function  $L$  remains constant. Equation (44) includes parts of recommendation policy of time  $T$  which correspond to message history  $(k)_{-t}^{T-1}$ . These parameters do not appear in any other constraint or in principal's utility function. Therefore, as long as the probability  $\mathbb{P}(s_T = b, m_{t+1:T} = (k)^{T-t} | m_{1:t} = ((k)^{t-1}, d))$ , given by (44), remains constant, changing the values of parameters  $\rho_T^{\theta, (k)^{T-1}}$ ,  $\theta = 1, \dots, T$  does not violate the obedience property of the mechanism.

The last point we should check is the effect of parameters  $\rho_T^{\theta, m_{1:T-1}}$ ,  $\theta = 1, \dots, T$ , on the principal's utility. According to (9), the principal's utility is

$$\begin{aligned} & \mathbb{E}^{1:T} \{U^P(\tau) | \rho\} \\ &= \sum_{\theta'=1}^{T+1} \mathbb{P}(\theta = \theta') \left[ \sum_{t=1}^{\theta'-1} \prod_{t'=1}^t \rho_{t'}^{t'+1, (k)^{t'-1}} + \sum_{t=\theta'}^T \left( \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^t \rho_{t'}^{\theta', (k)^{t'-1}} \right) \right] \\ &= \sum_{\theta'=1}^{T+1} \mathbb{P}(\theta = \theta') \left[ \sum_{t=1}^{\theta'-1} \prod_{t'=1}^t \rho_{t'}^{t'+1, (k)^{t'-1}} + \sum_{t=\theta'}^{T-1} \left( \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^t \rho_{t'}^{\theta', (k)^{t'-1}} \right) \right] \\ &+ \sum_{\theta'=1}^T \rho_T^{\theta', (k)^{T-1}} \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^{T-1} \rho_{t'}^{\theta', (k)^{t'-1}} \end{aligned} \quad (45)$$

It can be seen that the parameters  $\rho_T^{\theta, (k)^{T-1}}$  for  $\theta = 1, \dots, T$  do not appear in the first additive term of (45). These parameters only appeared in the second component of (45) as parts of the linear function  $L(\rho_T^{1, (k)^{T-1}}, \dots, \rho_T^{T, (k)^{T-1}})$ . Therefore, for each optimal mechanism we can construct another optimal mechanism where for each message sequence  $m_{1:T-1}$ , the values of parameters  $\rho_T^{\theta, m_{1:T-1}}$ ,  $\theta = 1, \dots, T$ , are the same. For each  $m_{1:T-1}$ , we denote this common value by  $\rho_T^{b, m_{1:T-1}}$ ,  $\theta = 1, \dots, T$ .

**Induction step.** Suppose that the statement of the lemma is true for times after  $t$ . Now, we want to prove that it is also true for time  $t$ . To prove this fact, we employ an approach similar to the one used in proving the basis of induction. We show that for each  $m_{1:t-1}$ , parameters  $\rho_t^{\theta, m_{1:t-1}}$  with  $\theta = 1, \dots, t$  appear in all the obedience constraints and the principal's utility function as parts of a constant linear combination. To do so, we first study the obedience constraints of time  $t$ . The obedience constraints of time  $t$  are as in (41)-(42). By some basic algebra, one can show that the parameters  $\rho_t^{\theta, m_{1:t-1}}$  with  $\theta = 1, \dots, t$ , appear in (41) as the following linear combination

$$\sum_{\theta'=1}^t \rho_t^{\theta', (k)^{t-1}} \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^{t-1} \rho_{t'}^{\theta', (k)^{t'-1}} \left( \sum_{l=t}^T \prod_{t'=t+1}^l \rho_{t'}^{\theta', (k)^{t'-1}} \right). \quad (46)$$

By the induction hypothesis, the probabilities of recommending silence at times greater than  $t$  do not depend on the exact time of the jump. Therefore, for each  $t' \geq t+1$  we can replace  $\rho_{t'}^{\theta', (k)^{t'-1}}$  by  $\rho_{t'}^{b, (k)^{t'-1}}$ . This replacement makes the last multiplicative term of (46) independent of  $\theta'$ ; hence we can neglect it. Therefore, the simplified form of the linear combination in which parameters  $\rho_t^{\theta, (k)^{t-1}}$  with  $\theta = 1, \dots, t$  appear is

$$\sum_{\theta'=1}^t \rho_t^{\theta', (k)^{t-1}} \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^{t-1} \rho_{t'}^{\theta', (k)^{t'-1}}. \quad (47)$$

We denote this linear combination by  $L(\rho_t^{1, (k)^{t-1}}, \dots, \rho_t^{t, (k)^{t-1}})$ . Now we investigate the second obedience constraint (42) of time  $t$ . We can show that the parameters  $\rho_t^{\theta, m_{1:t-1}}$  with  $\theta = 1, \dots, t$ , appear in (42) as the following linear combination

$$\sum_{\theta'=1}^t \rho_t^{\theta', (k)^{t-1}} \mathbb{P}(\theta = \theta') \prod_{t'=1}^{\theta'-1} \rho_{t'}^{t'+1, (k)^{t'-1}} \prod_{t'=\theta'}^{t-1} \rho_{t'}^{\theta', (k)^{t'-1}} \left( \sum_{l=t}^T \prod_{t'=t+1}^l \rho_{t'}^{\theta', (k)^{t'-1}} \right). \quad (48)$$

By the induction hypothesis, we have  $\rho_{t'}^{\theta', (k)^{t'-1}} = \rho_{t'}^{b, (k)^{t'-1}}$ , for  $t' \geq t+1$ . Using this equality, the last multiplicative term of (48) becomes independent of  $\theta'$ ; hence can be neglected. This makes the linear combination of (48) exactly

the same as  $L(\rho_t^{1,(k)^{t-1}}, \dots, \rho_t^{t,(k)^{t-1}})$ . Therefore, the parameters  $\rho_t^{\theta,(k)^{t-1}}$  with  $\theta = 1, \dots, t$  appear in the obedience constraints of time  $t$  as parts of one single linear combination.

There are two other sets of obedience constraints that we should study: obedience constraints at times before  $t$  and obedience constraints at times after  $t$ . Following the same steps as above, we can show that the parameters  $\rho_t^{\theta,(k)^{t-1}}$ , with  $\theta = 1, \dots, t$ , appear in all obedience constraints as well as the principal's utility function as parts of the linear function  $L(\rho_t^{1,(k)^{t-1}}, \dots, \rho_t^{t,(k)^{t-1}})$ . Therefore, as long as the value of this linear combination remains constant, changing the values of the individual parameters  $\rho_t^{\theta,(k)^{t-1}}$ ,  $\theta = 1, \dots, T$  does not violate the obedience constraints or change the principal's benefit. Other parameters of the recommendation policy of time  $t$  that appear in the obedience constraints are  $\rho_t^{\theta,(k)^{t-1}-t'}$ , where  $\theta = 1, \dots, t$  and  $t' = 1, \dots, t-1$ . For each  $t' = 1, \dots, t-1$ , the parameters  $\rho_t^{\theta,(k)^{t-1}-t'}$ , with  $\theta = 1, \dots, t$ , appear as a linear combination only in the obedience constraint of time  $t'$  when  $m_{t'} = d$ . Therefore, as long as the value of this linear combination remains constant, changing the values of parameters  $\rho_t^{\theta,(k)^{t-1}-t'}$ ,  $\theta = 1, \dots, t$  does not make any difference. Combining all the arguments above we conclude that, for each optimal mechanism we can construct another optimal mechanism such that for each message sequence  $m_{1:t-1}$ , the values of parameters  $\rho_t^{\theta,m_{1:t-1}}$ ,  $\theta = 1, \dots, t$ , are the same. For each  $m_{1:t-1}$ , we denote this common value by  $\rho_t^{b,m_{1:t-1}}$ ,  $\theta = 1, \dots, t$ .

**Conclusion.** By the principle of induction, the statement is true for all  $t = 1, \dots, T$ . This completes the proof of lemma 2.

### APPENDIX 3: Proof of Lemma 3

We prove this Lemma by showing that we can increase the utility of principal in an obedient mechanism by setting  $\rho_{t+1}^{g,m_{1:t}} = 1$ , for any  $t \in \mathcal{T}$ , and this change does not violate any of the obedience constraints. Having  $\rho_{t+1}^{g,m_{1:t}} = 1$  guarantees that the recommendations to declare the jump always will be obeyed by the detector. Therefore, in the following we focus on the recommendations to keep silent and show that by this change, the detector will be more willing to obey this kind of recommendations.

We proved in Lemma 2 that, without loss of optimality, at any time  $t \in \mathcal{T}$  the principal can restrict attention to recommendation policies that depend only on the message profile  $m_{1:t-1}$  and the current state  $s_t$  of the Markov chain (not on the state evolution). For any fixed recommendation policy  $\rho$  of the form suggested by Lemma 2, the detector is faced with a quickest detection problem [6]. However, there is a subtle difference between the standard quickest detection problem [6] and the one arising in our setup. In the standard quickest detection problem the detector's decision at any time  $t$  depends only on her posterior belief  $\hat{\pi}_t = \mathbb{P}(s_t = g | m_{1:t})$  about the state of the Markov chain at  $t$ . In the quickest detection problem arising in our setup, the detector's

decision at any time  $t$  depends on the received messages  $m_{1:t}$ , the principal's recommendation policy  $\rho$ , and her belief

$$\pi_t = \mathbb{P}(s_t = g | m_{1:t}, \rho). \quad (49)$$

This difference is due to the fact that at any time  $t$  the principal's fixed recommendation policy  $\rho_{t+1}$  depends on  $m_{1:t}$ , therefore, the detector needs keep  $m_{1:t}$  in order to determine the statistics of  $m_{t+1}$  so as to update her information at time  $t+1$ . Consequently, in our setup, the detector's information state at any time  $t$  is  $(\pi_t, m_{1:t}, \rho)$ . The detector's optimal strategy  $\gamma^* = (\gamma_1^*, \dots, \gamma_T^*)$  is determined by the dynamic program

$$W_T(\pi_T, m_{1:T}, \rho) = \min [\pi_T, c(1 - \pi_T)], \quad (50)$$

$$\begin{aligned} W_t(\pi_t, m_{1:t}, \rho) = & \min [\pi_t, c(1 - \pi_t) + \\ & \mathbb{E}\{W_{t+1}(\pi_{t+1}, m_{1:t}, m_{t+1}, \rho) | \pi_t, m_{1:t}, \rho\}], \quad t = 1, \dots, T-1, \end{aligned} \quad (51)$$

where

$$\pi_{t+1} = T_t(\pi_t, m_{1:t}, m_{t+1}, \rho), \quad (52)$$

$T_t(\cdot)$  is determined by Bayes' rule, and  $m_{t+1}$  is a random variable that takes values in the set  $\{d, k\}$ ; the statistics of  $m_{t+1}$  are determined by  $m_{1:t}$ , the state  $s_{t+1}$  of the Markov chain at time  $t+1$ , and the recommendation policy  $\rho$ . The first term on the right hand side (RHS) of (50) and (51) represents the detector's expected cost due to her decision to stop at  $t$  and declare that jump has occurred ( $a_t = 1$ ); the second term on the RHS of (50) and (51) represents the detector's expected cost due to her decision to wait/remain silent at time  $t$  ( $a_t = 0$ ). The second term on the RHS of (51) is equal to

$$\begin{aligned} & c(1 - \pi_t) + \mathbb{P}(m_{t+1} = k | \pi_t, m_{1:t}, \rho) W_{t+1}(T_t(\pi_t, m_{1:t}, k, \rho), m_{1:t}, k, \rho) \\ & + \mathbb{P}(m_{t+1} = d | \pi_t, m_{1:t}, \rho) W_{t+1}(T_t(\pi_t, m_{1:t}, d, \rho), m_{1:t}, d, \rho). \end{aligned} \quad (53)$$

Furthermore,

$$\mathbb{P}(m_{t+1} = k | \pi_t, m_{1:t}, \rho) = \pi_t(1 - q)\rho_{t+1}^{g, m_{1:t}} + (1 - \pi_t(1 - q))\rho_{t+1}^{b, m_{1:t}}. \quad (54)$$

Using (51) and (54) and Bayes' rule to write explicitly  $T_t(\pi_t, m_{1:t}, k, \rho)$  and  $T_t(\pi_t, m_{1:t}, d, \rho)$ , we obtain

$$\begin{aligned} W_t(\pi_t, m_{1:t}, \rho) = & \min \left[ \pi_t, c(1 - \pi_t) + (\pi_t(1 - q)\rho_{t+1}^{g, m_{1:t}} + (1 - \pi_t(1 - q))\rho_{t+1}^{b, m_{1:t}}) \right. \\ & W_{t+1}\left(\frac{\pi_t(1 - q)\rho_{t+1}^{g, m_{1:t}}}{\pi_t(1 - q)\rho_{t+1}^{g, m_{1:t}} + (1 - \pi_t(1 - q))\rho_{t+1}^{b, m_{1:t}}}, m_{1:t}, k, \rho\right) \\ & + (\pi_t(1 - q)(1 - \rho_{t+1}^{g, m_{1:t}}) + (1 - \pi_t(1 - q))(1 - \rho_{t+1}^{b, m_{1:t}})) \\ & \left. W_{t+1}\left(\frac{\pi_t(1 - q)(1 - \rho_{t+1}^{g, m_{1:t}})}{\pi_t(1 - q)(1 - \rho_{t+1}^{g, m_{1:t}}) + (1 - \pi_t(1 - q))(1 - \rho_{t+1}^{b, m_{1:t}})}, m_{1:t}, d, \rho\right) \right]. \end{aligned} \quad (55)$$

We use (50) and (55) to prove Lemma 3. To do this we first establish the following auxiliary results (Lemmas 6-9). The proof of these lemmas is presented at the end of Appendix 3<sup>†</sup>.

**Lemma 6** For any fixed time  $t$  and any fixed message profile  $m_{1:t}$ ,  $W_t(\pi_t, m_{1:t}, \rho)$  is a concave function of  $\pi_t$ .

**Lemma 7** For any fixed time  $t$ , any fixed message profile  $m_{1:t}$ , and any belief  $\pi_t$ ,  $W_t(\pi_t, m_{1:t}, \rho)$  is a concave function of  $\rho_{t+1}^{g, m_{1:t}}$ .

**Lemma 8** In an obedient mechanism, we must have  $\rho_{t+1}^{g, m_{1:t}} \in [\rho_{t+1}^{b, m_{1:t}}, 1]$ , for any time  $t$  and any message profile  $m_{1:t}$ .

**Lemma 9** For any fixed time  $t$ , any fixed message profile  $m_{1:t}$ , and any belief  $\pi_t$ , we have

$$\arg \min_{\rho_{t+1}^{g, m_{1:t}}} W_t(\pi_t, m_{1:t}, \rho) = 1. \quad (56)$$

Lemma 9 shows that setting  $\rho_{t+1}^{g, m_{1:t}} = 1$  decreases the value function  $W_t(\pi_t, m_{1:t}, \rho)$ . Using this result, we complete the proof of Lemma 3, that is, we show that when the Markov chain is in the good state, irrespective of the message history, it is always optimal for the principal to recommend the detector to keep silent. That is

$$\rho_{t+1}^{*g, m_{1:t}} = 1, \forall t, m_{1:t}. \quad (57)$$

We prove this by contradiction. Consider an optimal recommendation policy  $\rho^*$  and assume that there exists at least one time  $t$  and one message profile  $m_{1:t}$  such that  $\rho_{t+1}^{*g, m_{1:t}} < 1$ . We construct another recommendation policy  $\rho$  such that  $\rho_{t+1}^{b, m_{1:t}} = \rho_{t+1}^{*b, m_{1:t}}$ ,  $\rho_{t+1}^{g, m_{1:t}} = 1$ , and for all other times the policies  $\rho$  and  $\rho^*$  are the same. We show that the policy  $\rho$  leads to a higher expected utility for the principal and satisfies the obedience constraints corresponding to situations where the detector is recommended to keep silent. Repeating the above argument for all times  $s$  such that  $\rho_{s+1}^{*g, m_{1:s}} < 1$  we construct a policy  $\hat{\rho}$  that satisfies  $\hat{\rho}_{t+1}^{*g, m_{1:t}} = 1, \forall t \in \mathcal{T}$ , it improves the principal's expected utility as compared to  $\rho^*$  and incentivizes the detector to keep silent when she is recommended to do so. In the new mechanism  $\hat{\rho}$ , when the detector is recommended to declare the jump, she is sure that the Markov chain is in the bad state; hence, she will obey the recommendations for sure. Therefore, the mechanism  $\hat{\rho}$  is a policy with higher expected utility than  $\rho^*$  that satisfies all the obedience constraints. This contradicts the optimality of  $\rho^*$ .

To prove the claim stated above, we investigate the effect of setting  $\rho_{t+1}^{g, m_{1:t}}$  to one on the detector's incentives to keep silent at times  $t' \leq t$  and  $t' > t$ .

**Times  $t'$  where  $t' \leq t$ :** In this case, we show that setting  $\rho_{t+1}^{g, m_{1:t}} = 1$  reduces the value function at time  $t'$ . Therefore, if in the original mechanism

<sup>†</sup>Lemma 6 is needed so as to prove Lemma 8. Lemmas 7 and 8 are needed so as to prove Lemma 9. Lemma 9 is essential in establishing Lemma 3.

we had  $W_{t'}(\pi_{t'}, m_{1:t'}, \boldsymbol{\rho}^*) < \pi_{t'}$  (meaning that the detector had incentives to keep silent), in the new mechanism with  $\rho_{t+1}^{g, m_{1:t}} = 1$  this situation still holds. Therefore, the recommendation to keep silent at time  $t'$  will be obeyed. We prove this claim by backward induction on time  $t'$  as follows.

**Basis of induction.  $t'=t$ .** The result follows from Lemma 9.

**Induction step.** Suppose that setting  $\rho_{t+1}^{g, m_{1:t}} = 1$  decreases the value function  $W_{k+1}(\cdot)$ ,  $k+1 < t$ . We prove that this is also true for time  $k$ . Consider the RHS of (55) for time  $k$  as follows:

$$\begin{aligned} W_k(\pi_k, m_{1:k}, \boldsymbol{\rho}) = & \min \left[ \pi_k, c(1-\pi_k) + (\pi_k(1-q)\rho_{k+1}^{g, m_{1:k}} + (1-\pi_k(1-q))\rho_{k+1}^{b, m_{1:k}}) \right. \\ & W_{k+1}\left(\frac{\pi_k(1-q)\rho_{k+1}^{g, m_{1:k}}}{\pi_k(1-q)\rho_{k+1}^{g, m_{1:k}} + (1-\pi_k(1-q))\rho_{k+1}^{b, m_{1:k}}}, m_{1:k}, k, \boldsymbol{\rho}\right) \\ & + (\pi_k(1-q)(1-\rho_{k+1}^{g, m_{1:k}}) + (1-\pi_k(1-q))(1-\rho_{k+1}^{b, m_{1:k}})) \\ & \left. W_{k+1}\left(\frac{\pi_k(1-q)(1-\rho_{k+1}^{g, m_{1:k}})}{\pi_k(1-q)(1-\rho_{k+1}^{g, m_{1:k}}) + (1-\pi_k(1-q))(1-\rho_{k+1}^{b, m_{1:k}})}, m_{1:k}, d, \boldsymbol{\rho}\right) \right]. \end{aligned} \quad (58)$$

The first term of the minimum does not depend on  $\rho_{t+1}^{g, m_{1:t}}$ . The only component of the second term of the minimum that depends on  $\rho_{t+1}^{g, m_{1:t}}$  is the value function  $W_{k+1}(\cdot)$ . By the induction hypothesis, setting  $\rho_{t+1}^{g, m_{1:t}}$  to one decreases  $W_{k+1}(\cdot)$  for any input, therefore, it decreases  $W_k(\cdot)$  (by Eq. (55)).

**Conclusion.** By the principle of induction, the statement is true for all  $t' \leq t$ .

**Times  $t'$  where  $t' > t$ :** Changing the value of  $\rho_{t+1}^{g, m_{1:t}}$  with  $m_{1:t} \neq (k)^t$  has no effect on the obedience constraints (7)-(8) at time  $t' > t$ . However, setting  $\rho_{t+1}^{g, (k)^t} = 1$  increases the detector's belief in the good state of the Markov chain at time  $t'$  when she is recommended to keep silent. This is because we have

$$\pi_{t'} = \mathbb{P}(s_{t'} = g | m_{1:t'}, \boldsymbol{\rho}) = \frac{\mathbb{P}(s_{t'} = g) \prod_{t''=1}^{t'} \rho_{t''}^{g, (k)^{t''-1}}}{\mathbb{P}(s_{t'} = g) \prod_{t''=1}^{t'} \rho_{t''}^{g, (k)^{t''-1}} + \mathbb{P}(s_{t'} = b, m_{1:t'} = (k)^{t'})}. \quad (59)$$

It is easy to show that the dependence of the RHS of (59) on the parameter  $\rho_{t+1}^{g, (k)^t}$  is of the form

$$\frac{a\rho_{t+1}^{g, (k)^t}}{b\rho_{t+1}^{g, (k)^t} + d}, \quad (60)$$

where  $a, b, d$  are positive real numbers. This function is increasing in terms of  $\rho_{t+1}^{g, (k)^t}$ . Therefore, setting  $\rho_{t+1}^{g, (k)^t} = 1$  increases the belief of the detector in the good state of the Markov chain when she is recommended to keep silent.

We proved that for any (fixed) recommendation policy  $\boldsymbol{\rho}$ , at each time  $t$ , there is a threshold  $l_t^{(k)^t, \boldsymbol{\rho}}$  such that the detector keeps silent when she is

recommended to do so if and only if  $\mathbb{P}(s_t = g|m_{1:t} = (k)^t, \boldsymbol{\rho}) > l_t^{(k)^t, \boldsymbol{\rho}}$ . In the optimal mechanism  $\boldsymbol{\rho}^*$ , the detector obeys the recommendation to keep silent at  $t'$ , i.e.  $\mathbb{P}(s_{t'} = g|(k)^{t'}, \boldsymbol{\rho}^*) > l_{t'}^{(k)^{t'}, \boldsymbol{\rho}^*}$ . In the modified mechanism  $\boldsymbol{\rho}$ , where  $\rho_{t+1}^{g, (k)^t} = 1$ , we have  $l_{t'}^{(k)^{t'}, \boldsymbol{\rho}} = l_{t'}^{(k)^{t'}, \boldsymbol{\rho}^*}$  for all  $t' > t$  because for  $t' > t$  the value function  $W_{t'}(\pi_{t'}, m_{1:t'}, \boldsymbol{\rho})$  depends only on  $\rho_{t'+1:T}$  and  $\rho_{t'+1:T} = \rho_{t'+1:T}^*$ . Since, by the arguments above,  $\mathbb{P}(s_{t'} = g|(k)^{t'}, \boldsymbol{\rho}) > \mathbb{P}(s_{t'} = g|(k)^{t'}, \boldsymbol{\rho}^*) > l_{t'}^{(k)^{t'}, \boldsymbol{\rho}^*} = l_{t'}^{(k)^{t'}, \boldsymbol{\rho}}$ , under  $\boldsymbol{\rho}$  the detector obeys the recommendation to remain silent at time  $t'$ . From the arguments above it follows that under  $\boldsymbol{\rho}$  the detector always obeys the recommendation to keep silent.

Based on Eq. (9), we conclude that the principal's utility is an increasing function of the recommendation probabilities  $\rho_{t+1}^{g, m_{1:t}}$ . Therefore, the new constructed mechanism  $\boldsymbol{\rho}$  results in a higher utility for the principal and the detector always obeys the recommendation to keep silent. Repeating the above argument for all times  $s$  such that  $\rho_{s+1}^{*, g, m_{1:s}} < 1$  we construct a policy  $\hat{\boldsymbol{\rho}}$  that satisfies  $\hat{\rho}_{t+1}^{*, g, m_{1:t}} = 1, \forall t \in \mathcal{T}$ , it improves the principal's expected utility as compared to  $\boldsymbol{\rho}^*$  and incentivizes the detector to keep silent when she is recommended to do so. In the new mechanism  $\hat{\boldsymbol{\rho}}$ , when the detector is recommended to declare the jump, she is sure that the Markov chain is in the bad state; hence, she will obey the recommendations for sure. Therefore, the mechanism  $\hat{\boldsymbol{\rho}}$  is a policy with higher expected utility than  $\boldsymbol{\rho}^*$  that satisfies all the obedience constraints. This contradicts the optimality of  $\boldsymbol{\rho}^*$  and proves Lemma 3.

#### APPENDIX 4: Proof of Lemma 6

We prove this by backward induction on time  $t$ .

**Basis of induction.** Let  $t = T$ . According to (55), for each message profile  $m_{1:T}$ ,  $W_T(\pi_T, m_{1:T}, \boldsymbol{\rho})$  is the minimum of two affine, hence concave, functions of  $\pi_T$ . Minimum conserves concavity. Therefore,  $W_T(\pi_T, m_{1:T}, \boldsymbol{\rho})$  is a concave function of  $\pi_T$ .

**Induction step.** Suppose that  $W_{t+1}(\pi_{t+1}, m_{1:t+1}, \boldsymbol{\rho})$  is a concave function of  $\pi_{t+1}$ . Now, we want to show this is also true for  $t$ . Concave functions can be written as the infimum of affine functions; so we have

$$W_{t+1}(\pi_{t+1}, m_{1:t+1}, \boldsymbol{\rho}) = \inf_i [\alpha_i^{m_{1:t+1}} \pi_{t+1} + \beta_i^{m_{1:t+1}}]. \quad (61)$$

Substituting (61) in (55), we have

$$\begin{aligned} W_t(\pi_t, m_{1:t}, \boldsymbol{\rho}) = \min & \left[ \pi_t, c(1 - \pi_t) + \inf_i \left\{ \alpha_i^{m_{1:t}, k} \pi_t (1 - q) \rho_{t+1}^{g, m_{1:t}} + \right. \right. \\ & \left. \left. \beta_i^{m_{1:t}, k} (\pi_t (1 - q) \rho_{t+1}^{g, m_{1:t}} + (1 - \pi_t (1 - q)) \rho_{t+1}^{b, m_{1:t}}) \right\} + \inf_i \left\{ \alpha_i^{m_{1:t}, d} \pi_t (1 - q) \right. \right. \\ & \left. \left. (1 - \rho_{t+1}^{g, m_{1:t}}) + \beta_i^{m_{1:t}, d} (\pi_t (1 - q) (1 - \rho_{t+1}^{g, m_{1:t}}) + (1 - \pi_t (1 - q)) (1 - \rho_{t+1}^{b, m_{1:t}})) \right\} \right]. \end{aligned} \quad (62)$$



It can be seen that the terms inside of both infima are affine in  $\pi_t$ , hence concave. Therefore, since the infimum preserve concavity, each term inside the minimum is concave in  $\pi_t$ . Using the concavity preserving of the minimum function proves the concavity of the value function  $W_t(\pi_t, m_{1:t}, \boldsymbol{\rho})$  in terms of  $\pi_t$ .

**Conclusion.** By the principle of induction, the statement is true for all  $t = 1, \dots, T$ . This completes the proof of lemma 6.

#### APPENDIX 5: Proof of Lemma 7

The assertion of this lemma easily follows from (62).

#### APPENDIX 6: Proof of Lemma 8

In Lemma 6 we proved that: at any time  $t$ , the expected cost of the detector due to her decision to stop and declare the jump ( $a_t = 1$ ) is given by the first term of the RHS of (62) and is a linear increasing function of  $\pi_t$ ; the expected cost due to her decision to remain silent is given by the second term of the RHS of (62) and is a concave function of  $\pi_t$ . Denote the above expected costs by  $V_t(\pi_t, m_{1:t}, \boldsymbol{\rho}, a_t = 1)$  and  $V_t(\pi_t, m_{1:t}, \boldsymbol{\rho}, a_t = 0)$ , respectively.

Now, we claim that if

$$V_t(\pi, m_{1:t}, \boldsymbol{\rho}, a_t = 0) < V_t(\pi, m_{1:t}, \boldsymbol{\rho}, a_t = 1) \quad (63)$$

for some  $\pi$ , then (63) is also true for all  $\pi' > \pi$ .

We prove the claim by contradiction. Suppose that

$$V_t(\pi, m_{1:t}, \boldsymbol{\rho}, a_t = 0) < V_t(\pi, m_{1:t}, \boldsymbol{\rho}, a_t = 1), \quad (*)$$

and there exists a belief  $\pi' > \pi$  where

$$V_t(\pi', m_{1:t}, \boldsymbol{\rho}, a_t = 0) > V_t(\pi', m_{1:t}, \boldsymbol{\rho}, a_t = 1). \quad (**)$$

We know that

$$V_t(0, m_{1:t}, \boldsymbol{\rho}, a_t = 0) = c > 0 = V_t(0, m_{1:t}, \boldsymbol{\rho}, a_t = 1). \quad (***)$$

Therefore, (\*)-(\*\*\*) show that the graphs of functions  $V_t(\pi_t, m_{1:t}, \boldsymbol{\rho}, a_t = 0)$  and  $V_t(\pi_t, m_{1:t}, \boldsymbol{\rho}, a_t = 1)$  have two intersections  $l^1 \in (0, \pi)$  and  $l^2 \in (\pi, \pi')$ , where the graph of  $V_t(\pi_t, m_{1:t}, \boldsymbol{\rho}, a_t = 0)$  is below the line  $V_t(\pi_t, m_{1:t}, \boldsymbol{\rho}, a_t = 1)$  in the interval  $(l^1, l^2)$ . This contradicts the concavity of the function  $V_t(\pi_t, m_{1:t}, \boldsymbol{\rho}, a_t = 0)$  and hence the claim is true. The truth of this claim shows that at any time  $t$  and for any fixed  $m_{1:t}$  and  $\boldsymbol{\rho}$ , the detector's optimal policy is of threshold type with respect to  $\pi_t$ ; that is there is a threshold  $l_t^{m_{1:t}, \boldsymbol{\rho}}$  such that the detector declares the jump if and only if her belief  $\pi_t$  about the good state of the Markov chain is below the threshold  $l_t^{m_{1:t}, \boldsymbol{\rho}}$ . Therefore, for a mechanism to be obedient we need to have the following property:

$$\mathbb{P}(s_{t+1} = g | m_{1:t}, \boldsymbol{\rho}, m_{t+1} = d) \leq \mathbb{P}(s_{t+1} = g | m_{1:t}, \boldsymbol{\rho}, m_{t+1} = k), \forall t, m_{1:t}. \quad (64)$$

Writing the probabilities appearing in (64) in terms of the belief of time  $t$ , we can simplify the necessary condition (64) as follows:

$$\begin{aligned} & \frac{\pi_t(1-q)(1-\rho_{t+1}^{g,m_{1:t}})}{\pi_t(1-q)(1-\rho_{t+1}^{g,m_{1:t}}) + (1-\pi_t(1-q))(1-\rho_{t+1}^{b,m_{1:t}})} \\ & \leq \frac{\pi_t(1-q)\rho_{t+1}^{g,m_{1:t}}}{\pi_t(1-q)\rho_{t+1}^{g,m_{1:t}} + (1-\pi_t(1-q))\rho_{t+1}^{b,m_{1:t}}}, \quad \forall t, m_{1:t}. \end{aligned} \quad (65)$$

Simplifying the above expression gives us  $\rho_{t+1}^{g,m_{1:t}} \geq \rho_{t+1}^{b,m_{1:t}}$  as a necessary condition for an obedient mechanism, and hence completes the proof of Lemma 8.

#### APPENDIX 7: Proof of Lemma 9

According to Lemma 7,  $W_t(\pi_t, m_{1:t}, \boldsymbol{\rho})$  is a concave function of  $\rho_{t+1}^{g,m_{1:t}}$ . Therefore, the minimum of the function is attained either at the beginning or at the end of the feasible interval of the parameter  $\rho_{t+1}^{g,m_{1:t}}$  which is derived in Lemma 8 as  $[\rho_{t+1}^{b,m_{1:t}}, 1]$ . Therefore, to prove Lemma 9, we only need to show that

$$W_t(\pi_t, m_{1:t}, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g,m_{1:t}}\}, \rho_{t+1}^{g,m_{1:t}}=1) \leq W_t(\pi_t, m_{1:t}, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g,m_{1:t}}\}, \rho_{t+1}^{g,m_{1:t}}=\rho_{t+1}^{b,m_{1:t}}), \quad (66)$$

for every time  $t$ , belief  $\pi_t$ , and message profile  $m_{1:t}$ , where  $\boldsymbol{\rho} \setminus \{\rho_{t+1}^{g,m_{1:t}}\}$  denotes the recommendation policy  $\boldsymbol{\rho}$  excluding the element  $\rho_{t+1}^{g,m_{1:t}}$ . Using (55), we have

$$\begin{aligned} & W_t(\pi_t, m_{1:t}, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g,m_{1:t}}\}, \rho_{t+1}^{g,m_{1:t}}=1) \\ & = \min \left[ \pi_t, c(1-\pi_t) + (\pi_t(1-q) + (1-\pi_t(1-q))\rho_{t+1}^{b,m_{1:t}}) \right. \\ & \quad W_{t+1}\left(\frac{\pi_t(1-q)}{\pi_t(1-q) + (1-\pi_t(1-q))\rho_{t+1}^{b,m_{1:t}}}, m_{1:t}, k, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g,m_{1:t}}\}, \rho_{t+1}^{g,m_{1:t}}=1\right) \\ & \quad \left. + (1-\pi_t(1-q))(1-\rho_{t+1}^{b,m_{1:t}})W_{t+1}(0, m_{1:t}, d, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g,m_{1:t}}\}, \rho_{t+1}^{g,m_{1:t}}=1) \right]. \end{aligned} \quad (67)$$

When the detector is sure that the state is bad, she will declare the jump irrespective of the message she has received, and incurs no cost. Therefore, we have

$$\begin{aligned} & W_{t+1}(0, m_{1:t}, d, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g,m_{1:t}}\}, \rho_{t+1}^{g,m_{1:t}}=1) \\ & = W_{t+1}(0, m_{1:t}, k, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g,m_{1:t}}\}, \rho_{t+1}^{g,m_{1:t}}=1) = 0. \end{aligned} \quad (68)$$

This equality allows us to replace  $W_{t+1}(0, m_{1:t}, d, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g, m_{1:t}}\}, \rho_{t+1}^{g, m_{1:t}} = 1)$  in (67) by  $W_{t+1}(0, m_{1:t}, k, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g, m_{1:t}}\}, \rho_{t+1}^{g, m_{1:t}} = 1)$  and get

$$\begin{aligned}
& W_t(\pi_t, m_{1:t}, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g, m_{1:t}}\}, \rho_{t+1}^{g, m_{1:t}} = 1) \\
&= \min \left[ \pi_t, c(1 - \pi_t) + (\pi_t(1-q) + (1-\pi_t(1-q))\rho_{t+1}^{b, m_{1:t}}) \right. \\
& \quad W_{t+1}\left(\frac{\pi_t(1-q)}{\pi_t(1-q) + (1-\pi_t(1-q))\rho_{t+1}^{b, m_{1:t}}}, m_{1:t}, k, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g, m_{1:t}}\}, \rho_{t+1}^{g, m_{1:t}} = 1\right) \\
& \quad \left. + (1 - \pi_t(1-q))(1 - \rho_{t+1}^{b, m_{1:t}})W_{t+1}(0, m_{1:t}, k, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g, m_{1:t}}\}, \rho_{t+1}^{g, m_{1:t}} = 1) \right]. \tag{69}
\end{aligned}$$

It can be concluded from (50) and (55) that for any  $t$  and any fixed  $\pi_t$  and  $m_{1:t}$ , the value function  $W_t(\pi_t, m_{1:t}, \boldsymbol{\rho})$  is independent of the recommendation policies  $\rho_{t'}$ ,  $t' \leq t$  (the effect of  $\rho_{t'}$ ,  $t' \leq t$ , on  $W_t(\pi_t, m_{1:t}, \boldsymbol{\rho})$  is captured/summarized by  $\pi_t$  and  $m_{1:t}$ ). Therefore, we can write (69) as

$$\begin{aligned}
& W_t(\pi_t, m_{1:t}, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g, m_{1:t}}\}, \rho_{t+1}^{g, m_{1:t}} = 1) \\
&= \min \left[ \pi_t, c(1 - \pi_t) + (\pi_t(1-q) + (1 - \pi_t(1-q))\rho_{t+1}^{b, m_{1:t}}) \right. \\
& \quad W_{t+1}\left(\frac{\pi_t(1-q)}{\pi_t(1-q) + (1 - \pi_t(1-q))\rho_{t+1}^{b, m_{1:t}}}, m_{1:t}, k, \boldsymbol{\rho}\right) \\
& \quad \left. + (1 - \pi_t(1-q))(1 - \rho_{t+1}^{b, m_{1:t}})W_{t+1}(0, m_{1:t}, k, \boldsymbol{\rho}) \right]. \tag{70}
\end{aligned}$$

When  $\rho_{t+1}^{g, m_{1:t}} = \rho_{t+1}^{b, m_{1:t}}$ , we have from (55) that

$$\begin{aligned}
& W_t(\pi_t, m_{1:t}, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g, m_{1:t}}\}, \rho_{t+1}^{g, m_{1:t}} = \rho_{t+1}^{b, m_{1:t}}) = \min [\pi_t, c(1 - \pi_t) + \\
& \quad \rho_{t+1}^{b, m_{1:t}}W_{t+1}(\pi_t(1-q), m_{1:t}, k, \boldsymbol{\rho}) + (1 - \rho_{t+1}^{b, m_{1:t}})W_{t+1}(\pi_t(1-q), m_{1:t}, d, \boldsymbol{\rho})]. \tag{71}
\end{aligned}$$

In this case, the message sent by the principal at time  $t$  is independent of the observed state. Therefore, the detector neglects this data in her future decisions. Thus, we have

$$W_{t+1}(\pi_t(1-q), m_{1:t}, k, \boldsymbol{\rho}) = W_{t+1}(\pi_t(1-q), m_{1:t}, d, \boldsymbol{\rho}). \tag{72}$$

Substituting (72) into (71) gives

$$\begin{aligned}
& W_t(\pi_t, m_{1:t}, \boldsymbol{\rho} \setminus \{\rho_{t+1}^{g, m_{1:t}}\}, \rho_{t+1}^{g, m_{1:t}} = \rho_{t+1}^{b, m_{1:t}}) \\
&= \min [\pi_t, c(1 - \pi_t) + W_{t+1}(\pi_t(1-q), m_{1:t}, k, \boldsymbol{\rho})]. \tag{73}
\end{aligned}$$

According to Lemma 6,  $W_{t+1}(\pi_t(1-q), m_{1:t}, k, \boldsymbol{\rho})$  is a concave function of its first element. Thus, we have

$$\begin{aligned}
& (\pi_t(1-q) + (1 - \pi_t(1-q))\rho_{t+1}^{b, m_{1:t}})W_{t+1}\left(\frac{\pi_t(1-q)}{\pi_t(1-q) + (1 - \pi_t(1-q))\rho_{t+1}^{b, m_{1:t}}}, m_{1:t}, k, \boldsymbol{\rho}\right) \\
&+ (1 - \pi_t(1-q))(1 - \rho_{t+1}^{b, m_{1:t}})W_{t+1}(0, m_{1:t}, k, \boldsymbol{\rho}) \leq W_{t+1}(\pi_t(1-q), m_{1:t}, k, \boldsymbol{\rho}). \tag{74}
\end{aligned}$$

Comparing (70) and (73) based on the result derived in (74) proves that the inequality (66) holds; hence the statement of Lemma 9 is true.

#### APPENDIX 8: Proof of Lemma 4

We show that for each optimal sequential information disclosure mechanism  $\rho$  there is a time-based prioritized mechanism which obtains the same expected utility for the principal and satisfies all the obedience constraints.

Based on Lemma 3, we set  $\rho_t^{g,(k)^{t-1}} = 1$ , for all  $t, t = 1, 2, \dots, T$ . Furthermore, we let  $\rho_t^{s_t, m_{1:t-1}}$  be arbitrary  $0 \leq \rho_t^{s_t, m_{1:t-1}} \leq 1$ , whenever  $m_{1:t-1} \neq (k)^{t-1}$ . Thus, we concentrate on  $\rho_t^{b,(k)^{t-1}}, t = 1, 2, \dots, T$ . For ease of notation, throughout the remainder of the proof we denote  $\rho_t^{b,(k)^{t-1}}$  by  $\rho_t$ .

Consider an optimal mechanism  $\rho$  and let time  $t$  be the last time period where  $\rho_t < 1$  and  $\rho_{t+1} > 0$ . Using the result of Lemma 3 and the new notation, we find that the principal's expected utility according to  $\rho$  is

$$\mathbb{E}^\rho\{U^P\} = \sum_{l=1}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \rho_{t'} + \sum_{l=1}^T \sum_{\theta'=l+1}^{T+1} \mathbb{P}(\theta = \theta'). \quad (75)$$

It is clear from (75) that the principal's utility is a continuous and increasing function of each  $\rho_{t'}, t' = 1, 2, \dots, T$ . Since  $\rho_{t+1} > 0$  there exists a  $\gamma > 0$  such that  $\hat{\rho}_{t+1} = \rho_{t+1} - \gamma \geq 0$ . Replacing  $\rho_{t+1}$  by  $\hat{\rho}_{t+1}$  reduces the expected utility of the principal. However, since  $\mathbb{E}\{U^P(\tau)\}$  is a continuously increasing function of  $\rho_t$ , there exists some  $\epsilon(\gamma) > 0$  such that increasing  $\rho_t$  by the amount of  $\epsilon(\gamma)$  can compensate this decrease. This compensation is feasible if  $\hat{\rho}_t = \rho_t + \epsilon(\gamma)$  does not exceed its upper limit 1. To ensure this, we take  $\gamma = \min(\epsilon^{-1}(1 - \rho_t), \rho_{t+1})$ , where  $\epsilon^{-1}(\cdot)$  is the inverse of function  $\epsilon(\cdot)$ . The function  $\epsilon(\cdot)$  is one to one and increasing in  $\gamma$ , hence it has an inverse  $^\dagger$ . Choosing this  $\gamma$  results in either  $\hat{\rho}_t = 1$  or  $\hat{\rho}_{t+1} = 0$ .

The arguments above show that the principal's expected utility when he discloses his information based on the mechanism  $\hat{\rho}$ , where  $\hat{\rho}_{t'} = \rho_{t'}$ , for all  $t' \neq t, t+1$ , and

$$\hat{\rho}_t = \rho_t + \epsilon(\gamma), \quad \hat{\rho}_{t+1} = \rho_{t+1} - \gamma, \quad (76)$$

is the same as the average utility he gets when he uses the optimal mechanism  $\rho$ ; i.e.

$$\sum_{l=1}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \hat{\rho}_{t'} = \sum_{l=1}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \rho_{t'}. \quad (77)$$

Therefore, showing the new mechanism  $\hat{\rho}$  satisfies the obedience constraints will prove its optimality.

Since the recommendation to declare that the jump has occurred is always obeyed, we only need to investigate the obedience constraints (7) when the

---

<sup>†</sup>The explicit declaration of function  $\epsilon(\cdot)$  can be derived from equation (80) that will be derived later.

detector is recommended to keep silent. Using the results derived in (28) and (33), we can show that the mechanism  $\hat{\boldsymbol{\rho}}$  satisfies the obedience constraints if and only if

$$c \sum_{l=t''}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \hat{\rho}_{t'} \leq 1 - \sum_{\theta'=1}^{t''} \mathbb{P}(\theta = \theta'), \forall t'' \in \mathcal{T}. \quad (78)$$

Using (77) we can show that for  $t'' \leq t$ , we have

$$c \sum_{l=t''}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \hat{\rho}_{t'} = c \sum_{l=t''}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \rho_{t'} \leq 1 - \sum_{\theta'=1}^{t''} \mathbb{P}(\theta = \theta'), \quad (79)$$

where the last inequality follows from the obedience property of the original mechanism  $\boldsymbol{\rho}$ . Therefore, the obedience constraint for all times before  $t$  is satisfied. To show this is also true for times after  $t$  we need to take a closer look at (77). Canceling out the equal terms from both sides of the equality derived in (77), we get

$$\begin{aligned} & \epsilon(\gamma) \sum_{\theta'=1}^t \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^{t-1} \rho_{t'} + \left( -\gamma \mathbb{P}(\theta = t+1) \right. \\ & \left. + (\epsilon(\gamma)\rho_{t+1} - \gamma\rho_t - \gamma\epsilon(\gamma)) \sum_{\theta'=1}^t \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^{t-1} \rho_{t'} \right) \sum_{l=t+1}^T \prod_{t'=t+2}^l \rho_{t'} = 0. \end{aligned} \quad (80)$$

The first term in (80) and  $\sum_{l=t+1}^T \prod_{t'=t+2}^l \rho_{t'}$  are both non-negative. Therefore, we have

$$-\gamma \mathbb{P}(\theta = t+1) + (\epsilon(\gamma)\rho_{t+1} - \gamma\rho_t - \gamma\epsilon(\gamma)) \sum_{\theta'=1}^t \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^{t-1} \rho_{t'} \leq 0. \quad (81)$$

By substituting (76) in the left hand side of the obedience constraint (78) for time  $t'' \geq t+1$ , we obtain

$$\begin{aligned} & c \sum_{l=t''}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \hat{\rho}_{t'} = c \sum_{l=t''}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \rho_{t'} \\ & + \left( -\gamma \mathbb{P}(\theta = t+1) + (\epsilon(\gamma)\rho_{t+1} - \gamma\rho_t - \gamma\epsilon(\gamma)) \sum_{\theta'=1}^t \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^{t-1} \rho_{t'} \right) \sum_{l=t''}^T \prod_{t'=t+2}^l \rho_{t'} \\ & \leq c \sum_{l=t''}^T \sum_{\theta'=1}^l \mathbb{P}(\theta = \theta') \prod_{t'=\theta'}^l \rho_{t'} \leq 1 - \sum_{\theta'=1}^{t''} \mathbb{P}(\theta = \theta'), \end{aligned} \quad (82)$$

where the first inequality is based on (81), and the second one follows from the fact that the original mechanism  $\boldsymbol{\rho}$  is obedient, hence it satisfies the obedience

constraints (78). The results derived in (79) and (82) show that the new mechanism  $\hat{\rho}$  satisfies the obedience constraints, hence it is optimal. In this new mechanism, we have either  $\hat{\rho}_t = 1$  or  $\hat{\rho}_{t+1} = 0$ . Therefore, irrespective of the original mechanism  $\rho$ , the new (modified) mechanism satisfies the condition of having a time-based priority for recommending the detector to keep silent at time  $t$  and at the next time period  $t + 1$ .

It is clear from the above arguments that by repeating this procedure for any time  $t$  that violates the priority condition in a backward direction we can construct a time-based prioritized mechanism which is optimal. This completes the proof of Lemma 4.

#### APPENDIX 9: Proof of Theorem 5

After removing the terms in (21) that do not depend on  $t$  and doing some simple algebra, we have

$$\begin{aligned}
\arg \min_{t \leq n_p} q_{n_p}^t &= \arg \min_{t \leq n_p} \left[ \mathbb{P}(\theta > t) - c \sum_{l=t}^{n_p-1} \mathbb{P}(\theta \leq l) \right] \\
&= \arg \min_{t \leq n_p} \left[ \mathbb{P}(\theta > t) + c \sum_{l=1}^{n_p-1} \mathbb{P}(\theta \leq l) - c \sum_{l=t}^{n_p-1} \mathbb{P}(\theta \leq l) \right] \\
&= \arg \min_{t \leq n_p} \left[ \mathbb{P}(\theta > t) + c \sum_{l=1}^{t-1} \mathbb{P}(\theta \leq l) \right] \\
&= \arg \min_{t \leq n_p} \mathbb{E}^{N_o} \{ J^D(t, \theta) \}.
\end{aligned} \tag{83}$$

In Feature 2, we showed that  $J^D(t, \theta)$  attains its minimum at  $t = \tau^{N_o}$ . Therefore, since  $n_p \geq \tau^{N_o}$ , we have

$$\arg \min_{t \leq n_p} q_{n_p}^t = \tau^{N_o}. \tag{84}$$

This completes the proof of Theorem 5.

#### APPENDIX 10: Proof of Extension 2

**Lemma 10** Suppose that the state evolves according to a Markov chain with  $n$  states, one of which is absorbing. Then, it is always optimal for the principal to recommend the detector to keep silent when the Markov chain is in a non-absorbing state.

**Proof:** We consider the finite horizon case with  $T$  number of time slots and formulate the problem according to the Ely-Doval approach presented in [64]. The state in the Ely-Doval model is the evolution of the states in the original problem. We denote these states by  $\theta = (s_1, s_2, \dots, s_T)$ , where  $s_t \in \{e_1, e_2, \dots, e_n\}$  represents the state of the Markov chain at time  $t \in \mathcal{T} = \{1, 2, \dots, T\}$  and  $e_1$  represents the absorbing state. We denote the set of all states in the Ely-Doval model by  $\Theta$ .

We define a plan  $\mathbf{P} = (m_1, m_2, \dots, m_T)$ , where  $m_t \in \{k, d\}$  is the principal's recommendation at time  $t$  when the detector has not announced the jump yet. If the detector has already declared the jump, the principal does not send her any recommendation.

Based on these notations, the problem of designing an optimal information disclosure mechanism is equivalent to finding a probability distribution  $\pi(\boldsymbol{\theta}, \mathbf{P})$  such that 1) it satisfies the constraints related to obedience, prior belief, and causality, and 2) it maximizes the principal's expected utility. The obedience constraint at each history of messages  $m_{1:t}$  of time  $t$  can be written as

$$\sum_{\boldsymbol{\theta}, \mathbf{P}: m_{1:t} \subseteq \mathbf{P}} \pi(\boldsymbol{\theta}, \mathbf{P}) [U^D(\text{Dev}(\mathbf{P}, t), \mathbf{P}, \boldsymbol{\theta}) - U^D(\mathbf{P}, \mathbf{P}, \boldsymbol{\theta})] \leq 0, \quad (85)$$

where  $\text{Dev}(\mathbf{P}, t)$  represents the plan that is the same as  $\mathbf{P}$  everywhere but at time  $t$ . In fact,  $\text{Dev}(\mathbf{P}, t)$  is the result of the one-step deviation from plan  $\mathbf{P}$  at time  $t$ . The prior-belief constraints can be written for each  $\boldsymbol{\theta} \in \Theta$  as

$$\sum_{\mathbf{P}} \pi(\boldsymbol{\theta}, \mathbf{P}) = \mu(\boldsymbol{\theta}). \quad (86)$$

The causality constraints can also be written for each history of messages  $m_{1:t}$  and each history of states  $s_{1:t}$  at time  $t$  as

$$\frac{1}{\mu(\boldsymbol{\theta})} \sum_{\mathbf{P}: m_{1:t} \subseteq \mathbf{P}} \pi(\boldsymbol{\theta}, \mathbf{P}) = \frac{1}{\mu(\boldsymbol{\theta}_r(s_{1:t}))} \sum_{\mathbf{P}: m_{1:t} \subseteq \mathbf{P}} \pi(\boldsymbol{\theta}_r(s_{1:t}), \mathbf{P}), \quad \forall \boldsymbol{\theta} : s_{1:t} \subseteq \boldsymbol{\theta}, \quad (87)$$

where  $\boldsymbol{\theta}_r(s_{1:t}) = (h_t^s, e_1, e_1, \dots, e_1)$  is the reference state of history  $s_{1:t}$ .

The obedience constraints are linear inequalities in terms of the distribution. The prior-belief and causality constraints are also linear equalities. Moreover, the principal's expected utility is a linear function of the distribution  $\pi$ . Therefore, the problem can be written as the following linear program

$$\begin{aligned} & \max_{\pi(\boldsymbol{\theta}, \mathbf{P})} \sum_{\boldsymbol{\theta}, \mathbf{P}} \pi(\boldsymbol{\theta}, \mathbf{P}) U^P(\boldsymbol{\theta}, \mathbf{P}), \\ \text{s.t. } & \sum_{\boldsymbol{\theta}, \mathbf{P}: m_{1:t} \subseteq \mathbf{P}} \pi(\boldsymbol{\theta}, \mathbf{P}) [U^D(\text{Dev}(\mathbf{P}, t), \mathbf{P}, \boldsymbol{\theta}) - U^D(\mathbf{P}, \mathbf{P}, \boldsymbol{\theta})] \leq 0, \forall t, m_{1:t}, \\ & \sum_{\mathbf{P}} \pi(\boldsymbol{\theta}, \mathbf{P}) = \mu(\boldsymbol{\theta}), \quad \forall \boldsymbol{\theta}, \\ & \frac{1}{\mu(\boldsymbol{\theta})} \sum_{\mathbf{P}: m_{1:t} \subseteq \mathbf{P}} \pi(\boldsymbol{\theta}, \mathbf{P}) = \frac{1}{\mu(\boldsymbol{\theta}_r(s_{1:t}))} \sum_{\mathbf{P}: m_{1:t} \subseteq \mathbf{P}} \pi(\boldsymbol{\theta}_r(s_{1:t}), \mathbf{P}), \forall \boldsymbol{\theta} : s_{1:t} \subseteq \boldsymbol{\theta}, \\ & \pi(\boldsymbol{\theta}, \mathbf{P}) \geq 0, \forall \boldsymbol{\theta}, \mathbf{P}. \end{aligned} \quad (88)$$

Now, we use this formulation to show that the statement of Lemma 1 is true. Our proof is based on contradiction. Suppose that at the optimal solution of problem (88), there exists a pair of  $\boldsymbol{\theta}$  and  $\mathbf{P} = (m_1, m_2, \dots, m_T)$  with  $\pi(\boldsymbol{\theta}, \mathbf{P}) > 0$ , for which there exists at least one  $t \in \mathcal{T}$  such that  $m_t = d$  and  $s_t \neq e_1$ . Now, we construct another distribution  $\pi'$  from  $\pi$  with a higher utility for the principal.

To this end, for each  $\theta \in \Theta$ , we define a plan  $Truth(\theta) = (\tilde{m}_1(s_1), \tilde{m}_2(s_2), \dots, \tilde{m}_T(s_T))$ , where

$$\tilde{m}_t(s_t) = \begin{cases} d, & \text{if } s_t = e_1, \\ 0, & \text{otherwise.} \end{cases} \quad (89)$$

We also define by  $N(\theta)$  the set of non-desirable plans for state  $\theta$ . A plan  $\mathbf{P} = (m_1, m_2, \dots, m_T)$  is called non-desirable for state  $\theta$  if there exists at least one  $t \in \mathcal{T}$  such that  $m_t = d$  and  $s_t \neq e_1$ .

Now, we construct  $\pi'$  from  $\pi$  as follows:

1. We make  $\pi'(\theta, \mathbf{P}) = 0$ , for all  $\mathbf{P} \in N(\theta)$ ;
2. To satisfy the prior-belief constraints, we add all the probabilities that we zeroed in Step 1 to  $\pi'(\theta, Truth(\theta))$ , i.e.  $\pi'(\theta, Truth(\theta)) = \pi(\theta, Truth(\theta)) + \sum_{\mathbf{P} \in N(\theta)} \pi(\theta, \mathbf{P})$ .

It can be seen that the principal's utility is higher in  $\pi'$  compared to  $\pi$ . This is because in  $\pi$ , for some  $\theta$ , the detector is recommended to declare the jump when the Markov chain is still in a non-absorbing state. However, in  $\pi'$  the recommendations to declare the jump for those  $\theta$  have been postponed until the jump is actually occurred. Therefore, the detector remains silent for a longer period of time and hence the principal yields more utility.

Now, we want to show that the distribution  $\pi'$  satisfies all the constraints of problem (88).

- The prior-belief constraints are satisfied since  $\sum_{\mathbf{P}} \pi'(\theta, \mathbf{P}) = \sum_{\mathbf{P}} \pi(\theta, \mathbf{P})$ .
- The causality constraints are also satisfied. This is because all the changes we have made were based on the current and past states. Therefore, if the policy  $\pi$  is causal, then the policy  $\pi'$  is causal as well.
- Now, we should investigate whether  $\pi'$  satisfies the obedience constraints. First of all, increasing the probability of the pair  $(\theta, Truth(\theta))$  does not violate the obedience constraints. The reason is that  $U^D(Truth(\theta), Truth(\theta), \theta) = 0$  and  $U^D(Dev(Truth(\theta), t), Truth(\theta), \theta) < 0$ . Therefore,

$$U^D(Dev(Truth(\theta), t), Truth(\theta), \theta) - U^D(Truth(\theta), Truth(\theta), \theta) < 0, \quad (90)$$

which basically means that when the principal gives full information to the detector, it is never optimal for her to deviate from the recommendations. Due to this feature, the coefficient corresponding to  $\pi(\theta, Truth(\theta))$  is negative for all  $\theta$ ,  $t$ , and  $m_{1:t}$ . Therefore, increasing this probability reduces the left hand side of the obedience constraints and hence does not violate them.

Second, we need to show that making  $\pi'(\theta, \mathbf{P}) = 0$  for  $\mathbf{P} \in N(\theta)$  does not violate the obedience constraints. These terms appear in constraints related to both desirable and non-desirable histories  $m_{1:t}$ . For a non-desirable history of message (i.e. a history in which the detector is recommended to declare the jump when the state is non-absorbing), we have  $\pi'(\theta, \mathbf{P}) = 0$ ,



for all  $\mathbf{P} : m_{1:t} \subseteq \mathbf{P}$ . Therefore, the left hand sides of the obedience constraints related to non-desirable histories are always zero and hence the inequalities hold.

For a desirable history  $m_{1:t}$  and a non-desirable plan  $\mathbf{P}$  such that  $m_{1:t} \subseteq \mathbf{P}$  (consider  $\boldsymbol{\theta} = (e_2, e_4, e_5, e_2, e_3, e_1, e_1)$ ,  $t = 2$ ,  $m_{1:t} = (k, k)$  and  $\mathbf{P} = (k, k, k, d, d, d, d)$  for example), if the detector follows the recommendation, her utility will be  $-1$  (i.e. the cost of false alarm). If she deviate at time  $t$ , then again a false alarm occurs. Therefore,  $U^D(\text{Dev}(\mathbf{P}, t), \mathbf{P}, \boldsymbol{\theta}) - U^D(\mathbf{P}, \mathbf{P}, \boldsymbol{\theta}) = -1 - (-1) = 0$ . The coefficients of non-desirable plans is zero. Therefore, reducing their values does not change the obedience constraints.

Based on the discussion above, distribution  $\pi'$  satisfies all the constraints and increases the principal's utility. This contradicts the assumption that distribution  $\pi$  was optimal and hence completes the proof of Lemma 10.

## References

1. G. de Veciana and R. Baldick, "Resource allocation in multi-service networks via pricing: statistical multiplexing," *Computer Networks and ISDN Systems*, vol. 30, 1998.
2. P. Anussornnitisarn, S. Y. Nof, and O. Etzion, "Decentralized control of cooperative and autonomous agents for solving the distributed resource allocation problem," *International Journal of Production Economics*, vol. 98, 2005.
3. M. Rasouli and D. Teneketzis, "Electricity pooling markets with elastic demand: A mechanism design approach," in *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2014.
4. S. Das, E. Kamenica, and R. Mirka, "Reducing congestion through information design," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2017, pp. 1279–1284.
5. H. Tavafoghi and D. Teneketzis, "Informational incentives in congestion games," in *Proc. of the Allerton*, Oct. 2017.
6. A. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.
7. H. S. Witsenhausen, "On the structure of real-time source coders," *The Bell System Technical Journal*, vol. 58, no. 6, July 1979.
8. J. Walrand and P. Varaiya, "Optimal causal coding - decoding problems," *IEEE Transactions on Information Theory*, Nov. 1983.
9. A. Mahajan and D. Teneketzis, "Optimal design of sequential real-time communication systems," *IEEE Transactions on Information Theory*, vol. 55, no. 11, Nov 2009.
10. A. Nayyar and D. Teneketzis, "On the structure of real-time encoding and decoding functions in a multiterminal communication system," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6196–6214, Sep. 2011.
11. D. Teneketzis, "On the structure of optimal real-time encoders and decoders in noisy communication," *IEEE Transactions on Information Theory*, vol. 52, no. 9, Sept 2006.
12. V. P. Crawford and J. Sobel, "Strategic information transmission," *Econometrica*, vol. 50, no. 6, 1982.
13. M. Battaglini, "Multiple referrals and multidimensional cheap talk," *Econometrica*, vol. 70, no. 4, 2002.
14. S. Saritas, S. Yuksel, and S. Gezici, "On multi-dimensional and noisy quadratic signaling games and affine equilibria," in *American Control Conference (ACC)*, 2015.
15. A. Ambrus and S. Takahashi, "Multi-sender cheap talk with restricted state spaces," *Theoretical Economics*, vol. 3, no. 1, 2008.
16. M. Ivanov, "Dynamic information revelation in cheap talk," *The B.E. Journal of Theoretical Economics*, vol. 15, no. 2, July 2015.

17. —, “Dynamic learning and strategic communication,” *International Journal of Game Theory*, vol. 45, no. 3, Aug 2016.
18. A. Spence, “Job market signaling,” *The Quarterly Journal of Economics*, vol. 87, no. 3, pp. 355–374, 1973.
19. J. Hedlund, “Bayesian persuasion by a privately informed sender,” *Journal of Economic Theory*, vol. 167, pp. 229 – 268, 2017.
20. H. V. Stackelberg, *Marktform und Gleichgewicht*. Springer Verlag, 1934.
21. T. C. Schelling, “An essay on bargaining,” *The American Economic Review*, 1956.
22. T. Schelling, *The Strategy of Conflict*. Harvard University Press, 1980.
23. A. T. Kalai, E. Kalai, E. Lehrer, and D. Samet, “A commitment folk theorem,” *Games and Economic Behavior*, vol. 69, 2010.
24. I. Goldstein and Y. Leitner, “Stress tests and information disclosure,” *Journal of Economic Theory*, vol. 177, 2018.
25. N. Inostroza and A. Pavan, “Persuasion in global games with application to stress testing,” *Working paper*, 2018.
26. D. Bergemann, B. Brooks, and S. Morris, “The limits of price discrimination,” *American Economic Review*, vol. 105, no. 3, pp. 921–57, March 2015.
27. E. Akyol, C. Langbort, and T. Basar, “Information-theoretic approach to strategic communication as a hierarchical game,” *Proceedings of the IEEE*, vol. 105, no. 2, Feb 2017.
28. E. Kamenica and M. Gentzkow, “Bayesian persuasion,” *American Economic Review*, vol. 101, no. 6, October 2011.
29. D. Bergemann and S. Morris, “Information design: A unified perspective,” *Cowles Foundation Discussion Paper No. 2075R.*, 2017.
30. R. J. Aumann and M. Maschler, *Repeated Games with incomplete information*. MIT Press, 1995.
31. S. Zamir, “Chapter 5 repeated games of incomplete information: Zero-sum,” ser. Handbook of Game Theory with Economic Applications. Elsevier, 1992.
32. F. Forges, “Chapter 6 repeated games of incomplete information: Non-zero-sum,” ser. Handbook of Game Theory with Economic Applications. Elsevier, 1992.
33. J. Hedlund, “Persuasion with communication costs,” *Games and Economic Behavior*, vol. 92, pp. 28 – 40, 2015.
34. W. Tamura, “A Theory of Multidimensional Information Disclosure,” Institute of Social and Economic Research, Osaka University, ISER Discussion Paper 0828, Jan. 2012.
35. L. Rayo and I. Segal, “Optimal information disclosure,” *Journal of Political Economy*, vol. 118, 2010.
36. M. Gentzkow and E. Kamenica, “Bayesian persuasion with multiple senders and rich signal spaces,” *Games and Economic Behavior*, vol. 104, pp. 411 – 429, 2017.
37. F. Li and P. Norman, “On bayesian persuasion with multiple senders,” *Economics Letters*, vol. 170, pp. 66 – 70, 2018.
38. D. Bergemann and S. Morris, “Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium,” *American Economic Review*, vol. 106, no. 5, May 2016.
39. —, “Bayes correlated equilibrium and the comparison of information structures in games,” *Theoretical Economics*, vol. 11, no. 2, pp. 487–522, 2016.
40. R. Alonso and O. Camara, “Bayesian persuasion with heterogeneous priors,” *Journal of Economic Theory*, vol. 165, pp. 672 – 706, 2016.
41. C. Li, “A model of bayesian persuasion with transfers,” *Economics Letters*, vol. 161, pp. 93 – 95, 2017.
42. R. Boleslavsky and C. Cotton, “Grading standards and education quality,” *American Economic Journal: Microeconomics*, vol. 7, no. 2, pp. 248–79, 2015.
43. N. Yoder, “Designing incentives for heterogeneous researchers,” *Working paper*, 2019. [Online]. Available: <https://ssrn.com/abstract=3154143>
44. N. Schweizer and N. Szech, “Optimal revelation of life-changing information,” *Management Science*, vol. 64, no. 11, pp. 5250–5262, 2018.
45. D. Garcia, R. Teper, and M. Tsur, “Information design in insurance markets: Selling peaches in a market for lemons,” *Working paper*, 2018. [Online]. Available: <https://ssrn.com/abstract=3140361>
46. I. Kremer, Y. Mansour, and M. Perry, “Implementing the wisdom of the crowd,” *Journal of Political Economy*, vol. 122, no. 5, pp. 988–1012, 2014.

47. Y. Zhu and K. Savla, "On the stability of optimal bayesian persuasion strategy under a mistrust dynamics in routing games," in *56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2018, pp. 92–99.
48. M. Wu and S. Amin, "Information design for regulating traffic flows under uncertain network state," in *57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2019.
49. E. Kamenica, "Bayesian persuasion and information design," *Annual Review of Economics*, vol. 11, no. 1, pp. 249–272, 2019.
50. F. Farokhi, A. M. H. Teixeira, and C. Langbort, "Estimation with strategic sensors," *IEEE Trans. on Automatic Control*, vol. 62, 2017.
51. M. O. Sayin, E. Akyol, and T. Basar, "Strategic control of a tracking system," in *Proc. of the CDC*, Dec 2016.
52. D. Lingenbrink and K. Iyer, "Optimal signaling mechanisms in unobservable queues with strategic customers," in *Proceedings of the 2017 ACM Conference on Economics and Computation*. ACM, 2017, pp. 347–347.
53. J. Ely, "Beeps," *American Economic Review*, Jan. 2017.
54. J. Renault, E. Solan, and N. Vieille, "Optimal dynamic information provision," *Games and Economic Behavior*, vol. 104, pp. 329 – 349, 2017.
55. J. W. Best and D. P. Quigley, "Honestly dishonest : A solution to the commitment problem in bayesian persuasion," 2016.
56. J. Best and D. Quigley, "Persuasion for the long-run," Economics Group, Nuffield College, University of Oxford, Economics Papers 2016-W12, 2016.
57. M. O. Sayin, E. Akyol, and T. Basar, "Hierarchical multistage gaussian signaling games in noncooperative communication and control systems," *Automatica*, vol. 107, 2019.
58. M. O. Sayin and T. Basar, "Deception-as-defense framework for cyber-physical systems," *Arxiv*, 2019.
59. M. O. Sayin and T. Başar, "On the optimality of linear signaling to deceive kalman filters over finite/infinite horizons," in *GameSec*. Springer LNCS 11836, 2019.
60. T. Honryo, "Dynamic persuasion," *Journal of Economic Theory*, vol. 178, pp. 36 – 58, 2018.
61. D. Orlov, A. Skrzypacz, and P. Zryumov, "Persuading the principal to wait," *Stanford University Graduate School of Business Research Paper*, no. 16-20, Sep. 2019.
62. J. Horner and A. Skrzypacz, "Selling Information," *Journal of Political Economy*, vol. 124, no. 6, pp. 1515–1562, 2016.
63. J. Ely and M. Szydlowski, "Moving the goalposts," *Journal of Political Economy*, 2019.
64. L. Doval and J. Ely, "Sequential information design," *Econometrica*, 2019.
65. P. H. Au, "Dynamic information disclosure," *The RAND Journal of Economics*, vol. 46, no. 4, pp. 791–823, 2015.
66. F. Farhadi, D. Teneketzis, and S. J. Golestani, "Static and dynamic informational incentive mechanisms for security enhancement," in *2018 European Control Conference (ECC)*, June 2018.
67. E. Meigs, F. Parise, A. Ozdaglar, and D. Acemoglu, "Optimal dynamic information provision in traffic routing," *Arxiv*, 2020.
68. P. Basu, "Dynamic bayesian persuasion with a privately informed receiver," 2017.
69. A. Kolotilin, T. Mylovanov, A. Zapechelnuk, and M. Li, "Persuasion of a privately informed receiver," *Econometrica*, vol. 85, pp. 1949–1964, 2017.
70. B. Ginzburg, "Optimal information censorship," *Journal of Economic Behavior & Organization*, vol. 163, 2019.
71. R. B. MYERSON, *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.
72. D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.
73. D. P. Bertsekas, *Dynamic Programming and Optimal Control, Two Volume Set*. Athena Scientific, 2001.
74. V. Krishnamurthy, "Bayesian sequential detection with phase-distributed change time and nonlinear penalty - a pomdp lattice programming approach," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 7096–7124, 2011.