CONTROLLED SOCIAL SENSING: A POMDP APPROACH

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sujay Bhatt Hodrali Ramesh

August 2019

CONTROLLED SOCIAL SENSING:

A POMDP APPROACH

Sujay Bhatt Hodrali Ramesh, Ph.D.

Cornell University 2019

When human atoms are knit into an organization in which they are used, not in their full right as responsible human beings, but as cogs and levers and rods, it matters little that the raw material is flesh and blood. What is used as an element in a machine is in fact an element in a machine ... The hour is very late, and the choice of good and evil knocks at our door.

Norbert Wiener

The creation and widespread adoption of social media has positively contributed to the sensing ability of people in the society. People are influenced by the views and opinions shared by other members of the society (the social network), and in turn share their experiences, contributing to the societal wisdom. An approach to decision making with the information continuously generated online is learning the mechanism of the data generation process by people in the social network, and using these models to perform decision analysis. This approach not only provides insights about the collective behavior of the people in the social network, but also provides a means to make testable predictions when exogenous interventions or influences are present. We thus take a model based approach to people-centric decision analysis in this thesis and marry ideas from microeconomics & social psychology with traditional model estimation and decision analysis techniques from statistical signal processing & stochastic control.

We consider two well studied and empirically validated models from social psychology and microeconomics, namely, communication flow model based on social influence and social learning model, for modeling the behavior of people when they are situated in a social network. With these models to capture the behavior of people, we consider three frameworks for controlled social sensing:

i.) Framework I: People sequentially take one-shot decisions by learning from the social network. A controller learns the underlying parameters that are a driving force behind their decisions.

Application studied: Quickest change detection of market shocks using the actions generated by risk-averse traders.

ii.) Framework II: People sequentially take one-shot decisions by learning from the social network and a controller exogenously intervenes in their decision making process. The controller learns the underlying parameters that are a driving force behind their decisions by controlling the intervention.

Applications studied: (1) Monopoly pricing with risk-averse customers to maximize sales and revenue, by learning about the product quality. (2) Honest information elicitation via incentivization to learn about the product quality.

iii.) Framework III: People form opinions about an event or object of interest after repeated interactions with their social network. A controller learns the underlying parameters that are a driving force behind their opinions.

Application studied: Adaptive polling in hierarchical social influence networks to learn the parameters driving public opinion.

The key unifying theme of this thesis is to provide structural results for people-centric stochastic control problems, i.e, to derive mathematical insights about the interaction between the controller and the social network.

BIOGRAPHICAL SKETCH

Sujay Bhatt Hodrali Ramesh received the Master of Technology degree in Electrical Engineering from Indian Institute of Technology Bombay, India in 2014. He worked as a visiting researcher in U.S. Army Research Laboratory in the Computational and Information Sciences Directorate, Adelphi Maryland from September 2018 to January 2019. His doctoral research at Cornell University has been advised by Prof. Vikram Krishnamurthy.

To my dear mother,

who is the reason for every success in my life.

ACKNOWLEDGEMENTS

Don't judge each day by the harvest you reap but by the seeds that you plant.

Robert Louis Stevenson

It is hard to believe that close to 5 years have gone by since I started on this journey; it seems like only a year ago I was submitting applications to be accepted into graduate schools. It is thus a strange feeling acknowledging everyone who has made this journey possible and possibly successful.

It has been a rewarding journey; the diversity of emotions and experiences only matched by sheer uncertainty of it all. Looking back, I reckon that a PhD life is mostly about solving a time-inconsistent stochastic control problem – it's never optimal to plan for the whole journey.

My advisor: Vikram Krishnamurthy. Thank you for all the guidance and support; the journey would not have been possible without your encouragement and ideas. It has been a pleasure working with you – your sheer interest, energy, and passion for research has kept the fire burning through the night. Thank you for being a vociferous critic at every step of the way and providing ample opportunities to be creative. For me, the most important takeaway from this fruitful collaboration is learning to learn.

My committee: Itai Gurvich & Qing Zhao. Thank you for the comments and suggestions on my work. It has been a privilege having you both on my dissertation committee.

Thanks to all my other collaborators: Tavis Pedersen, Anup Aprem and Buddhika Nettasinghe. It has certainly been refreshing and beneficial working with you lot.

Thank you amma – you have been my biggest cheerleader on this journey. I am forever indebted to you, appa and bhatta, for always being there for me.

A big shout out to putti – you have been the birth of a rainbow after the rainstorm.

Finally, a warm expression of thanks is due to "me" for not losing sanity, and "myself" for making sure that "I" didn't lose it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Given the rapid success and proliferation of social network and social media, there is an urgent need to consider the following challenges: (i) understand the mechanism of data generation by people – what are their needs and motivations, and why they behave the way they do, (ii) how to effectively use the data generated by people to draw inferences and make judgments, and lastly (iii) how to influence the data generation process itself.

Consider a simple application in social media marketing, which is the process of gaining website traffic for businesses through the use of social media; the main challenges being – getting and making sense of the data, extracting information from it for decision making, and possibly influencing the data generation process itself to inform and improve the business.

Addressing the above challenges requires one to take a *people-centric* approach to the design and analysis of automated systems for decision making. Therefore, the main aim of this thesis is to integrate ideas from microeconomics, social psychology, statistical signal processing, and stochastic control; for the purpose of designing automated systems that can interact with, predict, and influence the data generation process by people in a social network.

1.1 Social Psychology & Microeconomics

A major influence on people's opinions and actions is the social network they belong to. Social psychology is a scientific field that studies the nature and causes of individual

human behavior in social situations [16]. It considers human behavior as influenced by other people and the social context in which this occurs.

Social Influence: Katz and Lazarsfeld [68] formulated a breakthrough theory of public opinion formation– individuals are influenced more by exposure to their social network than to the source of underlying ground truth or the environment. “Opinion leaders” or “Influencers" act as intermediaries between the environment and the social network. Because information, and thereby influence “flows” from the environment through opinion leaders to their respective followers, Katz and Lazarsfeld [68] called their model the “two-step flow” of communication.

We consider an extension of Katz and Lazarsfeld [68] idea considering the present day social network: a "N-step flow" model of communication, in other words, a hierarchical social influence network [84, 5, 133, 118, 22, 86, 25, 50] – people are stratified into different levels and each level influences the level below it. Many social networks have a hierarchical influence structure: knowledge dissemination on the web via Wikipedia, where a minority (2%) of internet users produce the content the great majority consumes [118]; in social media like Twitter [22], where small group of ‘opinion leaders’ gather most of the attention; in judicial hierarchy [25], where lower court judges are influenced by their direct superiors and courts above them; etc.

While social psychology provides reasonable models to analyze opinion formation, we need a model for capturing the decision making behavior in people or individual decision makers. Microeconomics is the study of decision making in individual human decision makers [88]. It has its roots in understanding supply and demand, and it posits that people try to get the most from what they have to sell and aim satisfy their desires as much as possible. This is usually modeled as maximizing an “utility" function or minimizing a “cost" function; a notion popularized by von Neumann-Morgenstern expected

utility theory.

Social learning. Social learning is a well studied model in (micro)economics that models learning to act from the observation of other's actions. Social learning theory posits that actions are the only trustworthy means of communication between the decision makers (social sensors). A rationale is that actions are a quantization of the information/ opinions of the social sensors, and language as a means of communication is not effective as its not universal and words can be deceiving. People who have taken actions in the past were self-serving, and acted according to private information. By looking at their behavior/ actions, one can infer something about what they knew.

In this thesis, we consider a sequential decision making model of Bayesian social learning introduced in [19, 134, 12], where the social sensors learn from their predecessors and make one shot decisions. Each social sensor has a private signal on the underlying state and considers this in addition to the (bounded) information gathered by its predecessors. This interplay results in the well known inefficiencies such as the formation of herds and information cascades. [123] show that some of the inefficiencies in the sequential social learning model arise due to the bounded nature of the information or beliefs used in decision making, and show that the true state is aggregated when the beliefs are unbounded.

There are many works in Bayesian social learning, where the social sensors repeatedly make decisions. [98] considers a model of repeated decision making where each sensor considers the group opinion as a summary of all the observed decisions. It is shown that cascades need not appear and weight attached to the decision of others determines the influence of private information on decision making. [11] consider a situation where only the payoffs as opposed to signals are observed and show that there is asymptotic learning – beliefs converge and all social sensors herd on the same action. [52] consider

rational learning over a social network, where the social sensors not always observe the decisions of other sensors, but instead make inferences on the unseen actions. The relation between network structure and herding is established. [1] establish asymptotic learning on general social network topologies and establish conditions for asymptotic learning. See also [117] and the references therein for social learning with repeated decision making. In case of repeated decision making, inefficiencies like cascades and herds can be avoided, however, that comes at the cost of increased computational complexity for the individual social sensors.

Though microeconomic theory has its shortcomings[1], it is essential to characterize optimal choices and it serves as a benchmark on which to build behavioral economic theories. Hence, as a first step towards people-centric stochastic control, we first resort to modeling decision making behavior of people using ideas from microeconomic theory and social psychology, and analyze how opinions and information structures[2] play a role in individual decision making.

1.2 Statistical Signal Processing & Stochastic Control

Statistical Signal Processing [69, 37, 57, 70] deals with the analysis of random signals for the purpose of information recovery using appropriate statistical techniques, namely, formulation of appropriate models to describe the behavior of the system generating the signals, the development of appropriate techniques for estimation of model parameters, and the assessment of model performances. This provides the tools for dealing with the data generated by the social network.

[1] Kahneman and Tversky [65] demonstrated that Prospect Theory better explains the decision making behavior in individual decision makers.

[2] Information structures, for example, indicate the historical choices or neighbors' opinions available to decision makers.

The environment or the ground truth (referred to as state) influences the opinions and decisions (observations) of the people (social sensors) in the social network. There is uncertainty in the observations and the noise that drives the evolution of the state. Stochastic control [8, 18, 17, 122, 106, 75] deals with decision making under uncertainty, where the probability distribution of the noise that affects the evolution and observation of the state variables is assumed to be known. In other words, there is complete knowledge of the model parameters (precomputed using statistical signal processing techniques or Monte Carlo methods [121, 110, 47, 29]). The aim in stochastic control is to design the time path of the control inputs that realizes the desired control objective with minimum cost, despite the presence of uncertainty. This provides the tools for information fusion for decision making, and influencing the data generation process of the social network.

1.3 Main Thesis Contributions

Below we summarize the main contributions of this thesis:

1. Chapter 2 presents a systematic analysis of quickest change detection with risk-averse social sensors. Here the risk is modeled using the Conditional Value-at-Risk (CVaR), a coherent risk measure that is one of the very important measures used to model risk in finance. The main results are: (i) We establish that the risk-averse (CVaR) social sensors exhibit monotone ordinal behavior; a trait commonly observed in human decision making. (ii) We establish that the stopping set is non-convex, and the optimal policy has a multi-threshold structure. This is unlike traditional quickest detection problems considered in the statistical signal processing literature, and has consequences on the implementation of the change detector. (iii) We establish the intuition that risk-averse (CVaR) sensors

herd sooner (or stick to the safer option) and don't prefer to learn from the crowd. The results appear in:

- Krishnamurthy, V. and **Bhatt, S.**. *Sequential detection of market shocks with risk-averse CVaR social sensors.* IEEE Journal of Selected Topics in Signal Processing, 10(6), pp.1061-1072, 2016.

2. Chapter 3 presents a systematic analysis of monopoly pricing with risk-averse (CVaR) social sensors. The main results are: (i) We establish that the monopoly can use price differentiation to improve the revenue. As the sensors are performing social learning, the monopoly can take advantage of the changing valuation to increase profits. (ii) We establish that the time path of the prices offered to the risk-averse social sensors, viewed as a stochastic process, is a super-martingale. This implies that the monopoly should start at a higher price and gradually reduce the price over time to capture the market consisting of risk-averse social sensors. The results appear in:

   - **Bhatt, S.** and Krishnamurthy, V. *Controlled information fusion with risk-averse CVaR social sensors.* IEEE 56^{th} Annual Conference on Decision and Control (CDC), pp. 2605-2610, December 2017.

3. Chapter 3 presents a systematic analysis of controlled information fusion with social sensors. Traditionally, information fusion is open-loop; we use feedback control to choose the control inputs to influence information acquisition from social sensors. The control inputs are chosen to motivate the social sensors to reveal honest information. The main results are: (i) We establish that truthful information revelationis a Markov perfect equilibrium for the social sensors. This implies that the social sensors have no incentive to delay information revelation and do not display contrarian behavior. (ii) We establish that the time path of the control

inputs, viewed as a stochastic process, is a sub-martingale. The value of the information held by the social sensors changes over time due to learning from other social sensors. Under this changing valuation, to motivate the social sensors to reveal truthful information, the control inputs (incentives or compensation) should increase on average over time. (iii) We establish the asymptotic consistency of the fusion estimates, and provide uniform bounds on the additional cost incurred for consistency. Social learning is well known to be in-efficient, i.e, a finite number of social sensors' decisions can derail the aggregation of true parameters. However, we establish that a suitable choice of control inputs can reverse this phenomenon. (iv) We provide uniform bounds on the budget saved as result of estimating the true parameters only upto a degree of confidence. This guarantees finite time estimation of the true parameters to within the specified level of confidence.

The results appear in:

- **Bhatt, S.** and Krishnamurthy, V. *Incentivized Information Fusion with Social Sensors.* ACM SIGMETRICS Performance Evaluation Review, 45(2), pp.90-95, 2018.
- **Bhatt, S.** and Krishnamurthy, V. *Controlled Sequential Information Fusion with Social Sensors.* ACM Transactions on Economics and Computation. (submitted) 2018.

4. Chapter 4 considers adaptive polling on social influence networks. Here, we deviate from the traditional sequential learning model of social learning to allow repeated interactions over a network to influence and form opinions. We use feedback control to select the social sensors in the social influence network to poll, to obtain their opinion about the parameters that drive public opinion. Our main results exploit the structure of the polling problem to determine novel conditions for Blackwell dominance that arise in hierarchical social influence networks. The

main results are: (i) We develop an adaptive version of the important intent polling algorithm that is inexpensive to implement. We show that in case of adaptive intent polling, the sensing channels between the pollster and the social influence network are polynomial channels, and establish an interesting link between Hurwitz stability and the Shannon capacity in terms of Blackwell dominance. (ii) We develop an adaptive version of the important expectation polling algorithm that is inexpensive to implement. We show that in case of adaptive expectation polling, the sensing channels between the pollster and the social influence network are ultrametric (hidden) channels, and establish an interesting link between hidden channels and the Shannon capacity in terms of Blackwell dominance. (iii) We provide an approximation procedure using Le Cam deficiency to deal with polling on general social networks. We use this procedure to develop an adaptive version of the Neighborhood Expectation Polling algorithm for hierarchical social influence networks. (iv) We provide results on performance bounds to account for mis-specification of the parameters and mis-classification of the influence level of social sensors for the pollster.

The results appear in:

- **Bhatt, S.**, Krishnamurthy, V. and Rangaswamy, M. *Controlled Sentiment Sampling for Information Fusion in Social Networks.* 21^{st} International Conference on Information Fusion (FUSION), pp. 1157-1162, July, 2018.
- **Bhatt, S.** and Krishnamurthy, V. *Adaptive Polling in Hierarchical Social Networks using Blackwell Dominance.* IEEE Transactions on Signal and Information Processing over Networks. (accepted) 2019.

1.4 Other Publications

Below we list other publications that are related but not a part of the main thesis. We briefly discuss the main results in the appendix of this thesis.

1. Krishnamurthy, V., **Bhatt, S.** and Pedersen, T. *Tracking infection diffusion in social networks: Filtering algorithms and threshold bounds.* IEEE Transactions on Signal and Information Processing over Networks, 3(2), pp.298-315, 2017.
2. Krishnamurthy, V., Aprem, A. and **Bhatt, S.** *Multiple stopping time POMDPs: Structural results.* 54^{th} Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 115-120, September 2016.
3. Krishnamurthy, V., Aprem, A. and **Bhatt, S.** *Multiple stopping time POMDPs: Structural results & application in interactive advertising on social media.* Automatica, 95, pp.385-398, 2018.
4. **Bhatt, S.**, Koppel, A., and Krishnamurthy, V. *Policy Gradient using Weak Derivatives for Reinforcement Learning.* (submitted) Conference on Decision and Control (CDC) 2019.
5. **Bhatt, S.**, Nettasinghe, B., and Krishnamurthy, V. *Efficient Polling Algorithms using Friendship Paradox and Blackwell Dominance.* (submitted) FUSION 2019.

1.5 Thesis Outline

Chapter 2 presents controlled social sensing under Framework-I. Here the individual decision makers (people or social sensors) are assumed to interact with and learn from each other in the decision making process. A controller makes use of the social sensor's

choices to realize a global objective. We first illustrate the problems considered in the literature that can be studied systematically using Framework-I. Then we study a novel application of the framework in analyzing quickest detection of market shocks with risk-averse social sensors, after providing an in-depth overview of relevant literature.

Chapter 3 presents controlled social sensing under Framework-II. Here the individual decision makers (people or social sensors) are assumed to interact with and learn from each other, while the controller intervenes in their decision making process. The controller makes use of the social sensor's choices to realize a global objective. We first illustrate the problems considered in the literature that can be studied systematically using Framework-II. Then we study two novel applications of the framework in analyzing monopoly pricing with risk-averse social sensors, and honest information elicitation using social sensors, after providing an in-depth overview of relevant literature.

Chapter 4 presents controlled social sensing under Framework-II. Here the social sensors are assumed to influence each other a network, where the sensors form opinions after repeated interaction with their neighbors. A controller makes use of social sensor's opinions to realize a global objective. We first illustrate the problems considered in the literature that can be studied systematically using Framework-III. Then we study a novel application of the framework in adaptive polling on hierarchical social influence networks, after providing an in-depth overview of relevant literature.

Chapter 5 presents the main concluding remarks and mathematical insights derived from the work carried out in this thesis.

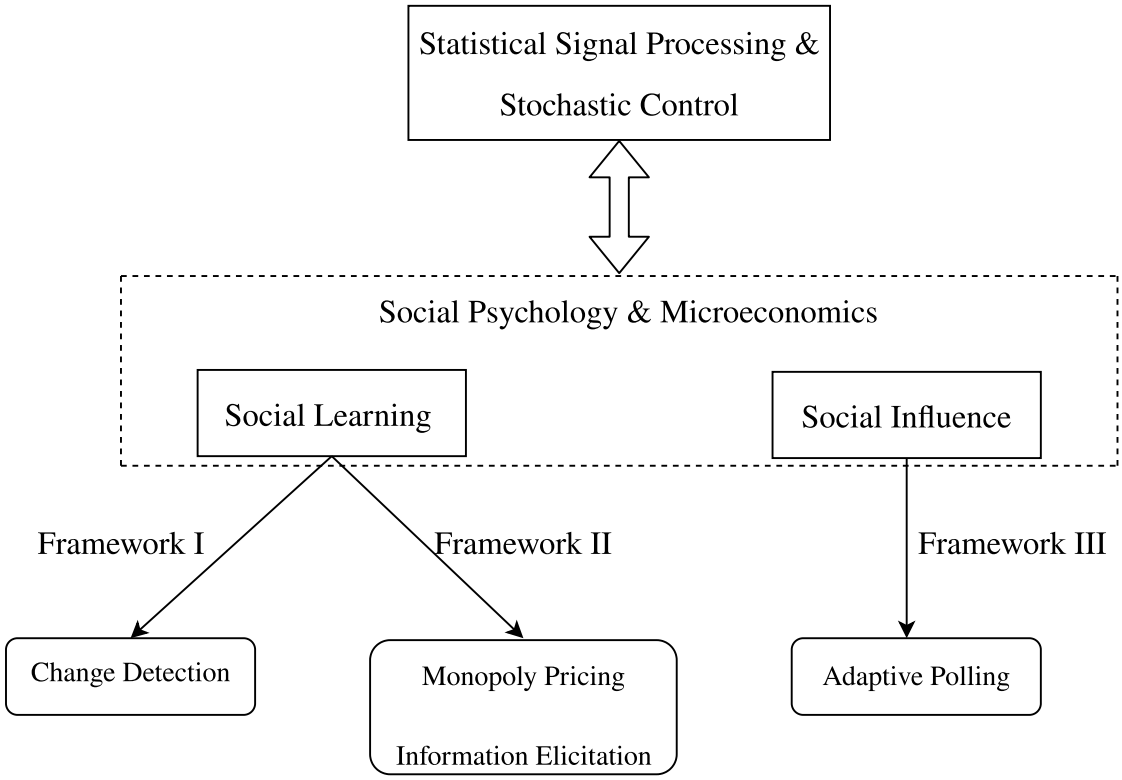

Figure 1.1: Outline of the thesis. The problems considered are at the intersection of statistical signal processing, stochastic control, social psychology and microeconomics. Three different frameworks are discussed to provide sufficiently general paradigms to deal with contemporary applications. Four novel applications and key results are derived in the considered frameworks.

CHAPTER 2

STOCHASTIC CONTROL WITH SOCIAL LEARNING: FRAMEWORK-I

Framework-I: Consider the following specifications: (i) the individual decision makers act once sequentially by observing all the predecessors' actions, (ii) the individual decision maker's reward is not directly influenced by the operator's control, (iii) the optimal policy of the operator is determined as the solution that achieves the stochastic control objective, which is to estimate the state with minimum incurred cost.

We first describe the problems considered in the literature that fit into Framework-I. We then illustrate an application of Framework-I in quickest detection of market shocks using social sensors that display an aversion to risk[1].

2.1 Relevant Problems in Literature

Framework-I is applicable in a wide range of situations. Below we discuss a few.

Designing mechanisms to perform information aggregation for the purpose of making sales forecasts. In [103], an Information Aggregation Mechanism (IAM) was deployed in inside Hewlett-Packard Corporation for sales prediction. [103] reports that performed better than traditional methods employed inside Hewlett-Packard.

Parimutuel betting markets as information aggregation devices, where empirical studies demonstrate the existence of a clear monotone relationship between odds and observed relative frequencies of winning. [104] considers designing parimutuel type betting systems to aggregate information held by decentralized system of individuals.

Learning and Information Aggregation in Stopping Games. [93] considers information aggregation in a stopping game with uncertain pay-offs that are correlated across players

[1] Krishnamurthy, V., & Bhatt, S. (2016). Sequential detection of market shocks with risk-averse CVaR social sensors. IEEE Journal of Selected Topics in Signal Processing, 10(6), 1061-1072.

and show that information aggregates in randomly occurring exit waves.

2.2 Application: Quickest Change Detection of Market Shocks

Consider an asset that experiences a shock in it's value at a phase distributed time (which generalizes geometric distributed change times). The change point detection problem is formulated as a Market Observer[2] seeking to detect a shock in the stock value (modeled as a Markov chain) by observing individual decision makers (traders) perform social learning. The market observer seeks to determine if the underlying asset value has changed based on the agent behaviour by balancing the natural trade-off between detection delay and false alarm. The problem of market shock detection considered in this chapter is different from the classical Bayesian quickest detection [119], [105], [51] where, non-human observations are used to detect the change.

Quickest detection in the presence of social learning was considered in [74] where it was shown that making global decisions (stop or continue) based on local decisions (buy or sell) leads to discontinuous value function and the optimal policy has multiple thresholds. However, unlike [74] which deals with expected cost, we consider a more general measure to account for the local agents' attitude towards risk. It is well documented in various fields like economics [40], behavioural economics, psychology [46] that people prefer a certain but possibly less desirable outcome over an uncertain but potentially larger outcome. To model this risk averse behaviour, commonly used risk measures[3] are

[2]The market observer could be the securities dealer (investment bank or syndicate) that underwrites the stock which is later traded in a secondary market.

[3]A risk measure $\varrho : \mathcal{L} \to \mathbb{R}$ is a mapping from the space of measurable functions to the real line which satisfies the following properties: (i) $\varrho(0) = 0$. (ii) If $S_1, S_2 \in \mathcal{L}$ and $S_1 \leq S_2$ a.s then $\varrho(S_1) \leq \varrho(S_2)$. (iii) if $a \in \mathbb{R}$ and $S \in \mathcal{L}$, then $\varrho(S + a) = \varrho(S) + a$. The risk measure is coherent if in addition ϱ satisfies: (iv) If $S_1, S_2 \in \mathcal{L}$, then $\varrho(S_1 + S_2) \leq \varrho(S_1) + \varrho(S_1)$. (v) If $a \geq 0$ and $S \in \mathcal{L}$, then $\varrho(aS) = a\varrho(S)$. The expectation operator is a special case where subadditivity is replaced by additivity.

Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), Entropic risk measure and Tail value at risk; see [91]. We consider social learning under CVaR risk measure. CVaR [111] is an extension of VaR that gives the total loss given a loss event and is a coherent risk measure [7]. In this work, we choose CVaR risk measure as it exhibits the following properties [7], [111]: (i) It associates higher risk with higher cost. (ii) It ensures that risk is not a function of the quantity purchased, but arises from the stock. (iii) It is convex. CVaR as a risk measure has been used in solving portfolio optimization problems [78], [82] credit risk optimization [6] and also order execution [49]. For an overview of risk measures and their application in finance, see [91].

From a signal processing point of view, the formulation and solutions presented here are non-standard due to the following three properties:

1. Traders (or social sensors) influence the behaviour of other traders, whereas in standard SP sensors typically do not affect other sensors.
2. Traders reveal quantized information (decisions) and have dynamics, whereas in standard SP sensors are static with the dynamics modelled in the state equation.
3. Standard SP is expectation centric. We use *coherent risk measures* which generalizes the concept of expected value and is much more relevant in financial applications. Such coherent risk measures [7] are now widely used in finance to model risk averse behaviour.

2.2.1 Formulation of Quickest Change Detection

The quickest detection problem is formulated as a partially observed Markov decision process (POMDP). We first present the Bayesian social learning model and define the

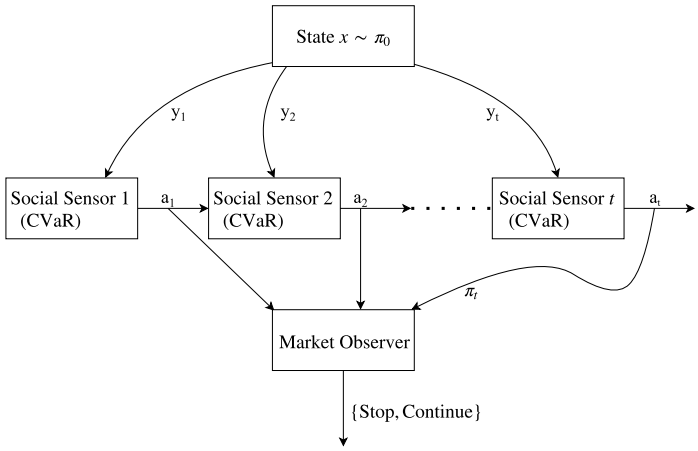

Figure 2.1: Sequential detection with risk-averse traders. Each trader (social sensor) receives a noisy observation on the state and chooses an action to minimize its CVaR measure of trading. The traders communicate their actions. The market observer seeks to determine if there is a change in the value of the underlying asset from the actions of the traders.

objective of the market observer. The problem solution is obtained using stochastic dynamic programming.

A: CVaR Social Learning Model

The market micro-structure is modelled as a discrete time dealer market motivated by algorithmic and high-frequency tick-by-tick trading [31]. There is a single traded stock or asset, a market observer and a countable number of traders. The asset has an initial true underlying value $x_0 \in \mathcal{X} = \{1, 2, \ldots, X\}$. The market observer does not receive direct information about $x \in \mathcal{X}$ but only observes the public buy/sell actions of the traders, $a_k \in \mathcal{A} = \{1(\text{buy}), 2(\text{sell})\}$. The traders themselves receive noisy private observations of the underlying value x and consider this in addition to the trading decisions of the other traders visible in the order book [3], [9], [71]. At a random time, τ^0 determined by the transition matrix P, the asset experiences a jump change in its value to a new value. The aim of the market observer is to detect the change time (global decision) with minimal

cost, having access to only the actions of these social traders. Let $y_k \in \mathcal{Y} = \{1, 2, \ldots, Y\}$ denote trader k's private observation. The initial distribution is $\pi_0 = (\pi_0(i), i \in \mathcal{X})$ where $\pi_0(i) = \mathbb{P}(x_0 = i)$. The agent based model has the following dynamics:

1. *Shock in the asset value*: At time $\tau^0 > 0$, the asset experiences a jump change (shock) in its value due to exogenous factors. The change point τ^0 is modeled by a *phase type (PH) distribution*. The family of all PH-distributions forms a dense subset for the set of all distributions [96] i.e., for any given distribution function F such that $F(0) = 0$, one can find a sequence of PH-distributions $\{F_n, n \geq 1\}$ to approximate F uniformly over $[0, \infty)$. The PH-distributed time τ^0 can be constructed via a multi-state Markov chain x_k with state space $\mathcal{X} = \{1, \ldots, X\}$ as follows: Assume state '1' is an absorbing state and denotes the state after the jump change. The states $2, \ldots, X$ (corresponding to beliefs $e_2, \ldots, e_X$) can be viewed as a single composite state that x resides in before the jump. So $\tau^0 = \inf\{k : x_k = 1\}$ and the transition probability matrix P is of the form

$$P = \begin{bmatrix} 1 & 0 \\ \underline{P}_{(X-1)\times 1} & \bar{P}_{(X-1)\times(X-1)} \end{bmatrix} \tag{2.1}$$

   The distribution of the absorption time to state 1 is

$$\nu_0 = \pi_0(1), \quad \nu_k = \bar{\pi}_0' \bar{P}^{k-1} \underline{P}, \quad k \geq 1, \tag{2.2}$$

   where $\bar{\pi}_0 = [\pi_0(2), \ldots, \pi_0(X)]'$. The key idea is that by appropriately choosing the pair (π_0, P) and the associated state space dimension X, one can approximate any given discrete distribution on $[0, \infty)$ by the distribution $\{\nu_k, k \geq 0\}$; see [96, pp.240-243]. The event $\{x_k = 1\}$ means the change point has occurred before time k according to PH-distribution (2.2). In the special case when x is a 2-state Markov chain, the change time τ^0 is geometrically distributed.

2. *Trader's Private Observation*: Trader k's private (local) observation denoted by y_k is a noisy measurement of the true value of the asset. It is obtained from the observation likelihood distribution as,

$$B_{xy} = \mathbb{P}(y_k = y | x_k = x). \tag{2.3}$$

The discreteness of the observation distribution captures the *boundedness* or the limited processing capabilities of the trader.

3. *Private Belief update*: Trader k updates its private belief using the observation y_k and the prior public belief $\pi_{k-1}(i) = \mathbb{P}(X = i | a_1, \ldots, a_{k-1})$ as the following Hidden Markov Model update

$$\eta_k = \frac{B_{y_k} P' \pi_{k-1}}{\mathbf{1}' B_{y_k} P' \pi_{k-1}} \tag{2.4}$$

where $\mathbf{1}$ denotes the X-dimensional vector of ones.

4. *Trading decision*: Agent k executes an action $a_k \in \mathcal{A} = \{1(\text{buy}), 2(\text{sell})\}$ to myopically minimize its cost. Let $c(i, a)$ denote the cost incurred if the trader takes action a when the underlying state is i. Let the local cost vector be

$$c_a = [c(1, a)\ c(2, a) \ldots\ c(X, a)] \tag{2.5}$$

The costs for different actions are taken as

$$c(i, j) = p_j - \beta_{ij} \text{ for } i \in \mathcal{X}, j \in \mathcal{A} \tag{2.6}$$

where β_{ij} corresponds to the trader's demand. Here demand is the trader's desire and willingness to trade at a price p_j for the stock. Here p_1 is the quoted price for purchase and p_2 is the price demanded in exchange for the stock. We assume that the price is the same during the period in which the value changes. As a result, the willingness of each agent only depends on the degree of uncertainty on the value of the stock.

Remark. The analysis provided in this chapter straightforwardly extends to the case when different traders are facing different prices like in an order book [3], [9], [71]. For notational simplicity we assume the costs are time invariant.

The trader considers measures of risk in the presence of uncertainty in order to overcome the losses incurred in trading. To illustrate this, let $c(x, a)$ denote the loss incurred with action a while at unknown and random state $x \in \mathcal{X}$. When a trader solves an optimization problem involving $c(x, a)$ for selecting the best trading decision, it will take into account not just the expected loss, but also the "riskiness" associated with the trading decision a. The agent therefore chooses an action a_k to minimize the CVaR measure[4] of trading as

$$\begin{aligned} a_k &= \underset{a\in\mathcal{A}}{\operatorname{argmin}}\{\mathrm{CVaR}_\alpha(c(x_k, a))\} \\ &= \underset{a\in\mathcal{A}}{\operatorname{argmin}}\{\min_{z\in\mathbb{R}} \{z + \frac{1}{\alpha}\mathbb{E}_{y_k}[\max\{(c(x_k, a) - z), 0\}]\}\} \end{aligned} \tag{2.7}$$

Here $\alpha \in (0, 1]$ reflects the degree of risk-aversion for the agent (the smaller α is, the more risk-averse the trader is). Define

$$\mathcal{H}_k := \sigma\text{- algebra generated by } (a_1, a_2, \ldots, a_{k-1}, y_k) \tag{2.8}$$

$\mathbb{E}_{y_k}$ denotes the expectation with respect to private belief, i.e, $\mathbb{E}_{y_k} = \mathbb{E}[.|\mathcal{H}_k]$ when the private belief is updated after observation y_k.

5. *Social Learning and Public belief update*: Trader k's action is recorded in the order book and hence broadcast publicly. Subsequent traders and the market observer update the public belief on the value of the stock according to the social

[4]For the reader unfamiliar with risk measures, it should be noted that CVaR is one of the 'big' developments in risk modelling in finance in the last 15 years. In comparison, the value at risk (VaR) is the percentile loss namely, $\mathrm{VaR}_\alpha(x) = \min\{z : F_x(z) \geq \alpha\}$ for cdf F_x. While CVaR is a coherent risk measure, VaR is not convex and so not coherent. CVaR has other remarkable properties [111]: it is continuous in α and jointly convex in (x, α). For continuous cdf F_x, $\mathrm{CVaR}_\alpha(x) = \mathbb{E}\{X|X > \mathrm{VaR}_\alpha(x)\}$. Note that the variance is not a coherent risk measure.

learning Bayesian filter as follows

$$\pi_k = T^{\pi_{k-1}}(\pi_{k-1}, a_k) = \frac{R_{a_k}^{\pi_{k-1}} P' \pi_{k-1}}{\mathbf{1}' R_{a_k}^{\pi_{k-1}} P' \pi_{k-1}} \tag{2.9}$$

Here, $R_{a_k}^{\pi_{k-1}} = \text{diag}(\mathbb{P}(a_k|x = i, \pi_{k-1}), i \in \mathcal{X})$, where $\mathbb{P}(a_k|x = i, \pi_{k-1}) = \sum_{y \in \mathcal{Y}} \mathbb{P}(a_k|y, \pi_{k-1})\mathbb{P}(y|x_k = i)$ and

$$\mathbb{P}(a_k|y, \pi_{k-1}) = \begin{cases} 1 & \text{if } a_k = \underset{a \in \mathcal{A}}{\text{argmin}}\ \text{CVaR}_\zeta(c(x_k, a)); \\ 0 & \text{otherwise.} \end{cases}$$

Note that π_k belongs to the unit simplex $\Pi(X) \stackrel{\Delta}{=} \{\pi \in \mathbb{R}^X : \mathbf{1}_X' \pi = 1, 0 \leq \pi \leq 1 \text{ for all } i \in \mathcal{X}\}$.

6. *Market Observer's Action*: The market observer (securities dealer) seeks to achieve quickest detection by balancing delay with false alarm. At each time k, the market observer chooses action[5] u_k as

$$u_k \in \mathcal{U} = \{1(\text{stop}), 2(\text{continue})\} \tag{2.10}$$

Here 'Stop' indicates that the value has changed and the dealer incorporates this information before selling new issues to investors. The formulation presented considers a general parametrization of the costs associated with detection delay and false alarm costs. Define

$$\mathcal{G}_k := \sigma\text{- algebra generated by } (a_1, a_2, \ldots, a_{k-1}, a_k). \tag{2.11}$$

i) *Cost of Stopping*: The asset experiences a jump change(shock) in its value at time τ^0. If the action $u_k = 1$ is chosen before the change point, a false alarm penalty is incurred. This corresponds to the event $\bigcup_{i \geq 2} \{x_k = i\} \cap \{u_k = 1\}$. Let $\mathcal{I}$ denote the indicator function. The cost of false alarm in state $i, i \in \mathcal{X}$ with

[5]It is important to distinguish between the "local" decisions a_k of the traders and "global" decisions u_k of the market observer. Clearly the decisions a_k affect the choice of u_k as will be made precise below.

$f_i \geq 0$ is thus given by $f_i \mathcal{I}(x_k = i, u_k = 1)$. The expected false alarm penalty is

$$C(\pi_k, u_k = 1) = \sum_{i \in \mathcal{X}} f_i \mathbb{E}\{\mathcal{I}(x_k = i, u_k = 1) | \mathcal{G}_k\}$$
$$= \mathbf{f}' \pi_k \tag{2.12}$$

where $\mathbf{f} = (f_1, \ldots, f_X)$ and it is chosen with increasing elements, so that states further from '1' incur higher false alarm penalties. Clearly, $f_1 = 0$.

ii) *Cost of delay*: A delay cost is incurred when the event $\{x_k = 1, u_k = 2\}$ occurs, i.e, even though the state changed at k, the market observer fails to identify the change. The expected delay cost is

$$C(\pi_k, u_k = 2) = d\, \mathbb{E}\{\mathcal{I}(x_k = i, u_k = 1) | \mathcal{G}_k\}$$
$$= d e_1' \pi_k \tag{2.13}$$

where $d > 0$ is the delay cost and e_1 denotes the unit vector with 1 in the first position.

B: Market Observer's Quickest Detection Objective

The market observer chooses its action at each time k as

$$u_k = \mu(\pi_k) \in \{1(\text{stop}), 2(\text{continue})\} \tag{2.14}$$

where μ denotes a stationary policy. For each initial distribution $\pi_0 \in \Pi(X)$ and policy μ, the following cost is associated

$$J_\mu(\pi_0) = \mathbb{E}^\mu_{\pi_0} \left\{ \sum_{k=1}^{\tau-1} \rho^{k-1} C(\pi_k, u_k = 2) + \rho^{\tau-1} C(\pi_k, u_k = 1) \right\} \tag{2.15}$$

Here $\rho \in [0, 1]$ is the discount factor which is a measure of the degree of impatience of the market observer. (As long as $\mathbf{f}$ is non-zero, stopping is guaranteed in finite time and so $\rho = 1$ is allowed.)

Given the cost, the market observer's objective is to determine τ^0 with minimum cost by computing an optimal policy μ^* such that

$$J_{\mu^*}(\pi_0) = \inf_{\mu \in \boldsymbol{\mu}} J_\mu(\pi_0) \tag{2.16}$$

The sequential detection problem (2.16) can be viewed as a partially observed Markov decision process (POMDP) where the belief update is given by the social learning filter.

C: Stochastic Dynamic Programming Formulation

The optimal policy of the market observer $\mu^* : \Pi(X) \to \{1, 2\}$ is the solution of (2.15) and is given by Bellman's dynamic programming equation as follows:

$$V(\pi) = \min \left\{ C(\pi, 1), C(\pi, 2) + \rho \sum_{a \in \mathcal{A}} V(T^\pi(\pi, a)) \sigma(\pi, a) \right\} \tag{2.17}$$

$$\mu^*(\pi) = \operatorname{argmin} \left\{ C(\pi, 1), C(\pi, 2) + \rho \sum_{a \in \mathcal{A}} V(T^\pi(\pi, a)) \sigma(\pi, a) \right\}$$

where $T^\pi(\pi, a) = \frac{R_a^\pi P' \pi}{\mathbf{1}' R_a^\pi P' \pi}$ is the CVaR-social learning filter and $\sigma(\pi, a) = \mathbf{1}' R_a^\pi P' \pi$ is the normalization factor of the Bayesian update. $C(\pi, 1)$ and $C(\pi, 2)$ from (2.12) and (2.13) are the market observer's costs. As $C(\pi, 1)$ and $C(\pi, 2)$ are non-negative and bounded for $\pi \in \Pi(X)$, the stopping time τ is finite for all $\rho \in [0, 1]$.

The aim of the market observer is then to determine the stopping set $\mathcal{S} = \{\pi \in \Pi(X) : \mu^*(\pi) = 1\}$ given by:

$$\mathcal{S} = \left\{ \pi : C(\pi, 1) < C(\pi, 2) + \rho \sum_{a \in \mathcal{A}} V(T^\pi(\pi, a)) \sigma(\pi, a) \right\}$$

As will be shown below, because of the social learning dynamics, quite remarkably, $\mathcal{S}$ is not necessarily a convex set. This is in stark contrast to classical quickest detection where the stopping region is always convex irrespective of the change time distribution [73].

2.2.2 Main Results

A: Assumptions

(A1) Observation matrix B and transition matrix P are TP2 (all second order minors are non-negative)

(A2) Agents' local cost vector c_a is sub-modular. That is $c(x,2) - c(x,1) \leq c(x+1,2) - c(x+1,1)$.

The matrices being TP2 [67] ensures that the public belief Bayesian updates can be compared [87] and sub-modular [125] costs ensure that if it is less risky to choose $a = 2$ when in x, it is also less risky to choose it when in $x + 1$.

B: Monotone behavior of risk-averse traders

The following result says that traders choose a trading decision that is monotone and ordinal in their private observation. Humans typically convert numerical attributes to ordinal scales before making decisions. For example, it does not matter if the cost of a meal at a restaurant is \$200 or \$205; an individual would classify this cost as "high". Also credit rating agencies use ordinal symbols such as AAA, AA, A. According to Theorem 2.1, risk-averse traders take decisions that are monotone and ordinal in the observations and monotone in the prior; and its monotone ordinal behaviour implies that a Bayesian model chosen in this work is a useful idealization.

The $\mathcal{Y} \times \mathcal{A}$ local decision likelihood probability matrix R^{π} (analogous to observation

likelihood) can be computed as

$$R^{\pi} = BM^{\pi}, \text{where } M^{\pi}_{y,a} \stackrel{\Delta}{=} \mathbb{P}(a|y,\pi) \tag{2.18}$$

$$\mathbb{P}(a|y,\pi) = \mathcal{I}(\text{CVaR}_{\alpha}(c(x_k,a)) < \text{CVaR}_{\alpha}(c(x_k,a')))$$

where $a' = A - \{a\}$. Here $\mathcal{I}$ denotes the indicator function. Let $H^{\alpha}(y,a) = \text{CVaR}_{\alpha}(c(x_k,a))$ denote the cost with CVaR measure, associated with action a and observation y for convenience i.e,

$$H^{\alpha}(y,a) = \min_{z\in\mathbb{R}} \{z + \frac{1}{\alpha}\mathbb{E}_y[\max\{(c(x,a)-z),0\}]\} \tag{2.19}$$

Here $\mathbb{E}_y = \mathbb{E}[.|\mathcal{H}_k]$. y indicates the dependence of $\mathbb{E}$ and hence H^{α} on the observation. Let $a^*(\pi,y) = \text{argmin } H^{\alpha}(y,a)$ denote the optimal action of the trader with explicit dependence on the distribution and observation.

Theorem 2.1. Under (A1) and (A2), the action $a^*(\pi,y)$ made by each agent is increasing and hence ordinal in y for any prior belief π. Under (A2), $a^*(\pi,y)$ is increasing in π with respect to the monotone likelihood ratio order.

Theorem 2.1 says that traders exhibit monotone ordinal behaviour. The condition that $a^*(\pi,y)$ is monotone in the observation y is required to characterize the local decision matrices on different regions in the belief space which is stated next.

C: Non-convex stopping regions for quickest change detection

Here, we explore the link between local and global behavior for detection of market shocks. We show that the stopping region for the sequential detection problem is non-convex; this is in contrast to standard signal processing quickest detection problems where the stopping set is convex. The optimal policy for the market observer hence exhibits a multi-threshold structure. This has the implication that, the market observer

is not very confident in implementing the optimal policies: announce a change when the belief is high, but stay put when the belief is larger.

Theorem 2.2. Under (A1) and (A2), there are at most $Y + 1$ distinct local decision likelihood matrices R^π and the belief space $\Pi(X)$ can be partitioned into the following $Y + 1$ polytopes:

$$\mathcal{P}_1^\alpha = \{\pi \in \Pi(X) : H(1,1) - H(1,2) \geq 0\}$$

$$\mathcal{P}_l^\alpha = \{\pi \in \Pi(X) : H(l-1,1) - H(l-1,2) < 0 \ \cap \ H(l,1) - H(l,2) \geq 0\}, \ l = 2, \ldots, Y$$

$$\mathcal{P}_{Y+1}^\alpha = \{\pi \in \Pi(X) : H(Y,1) - H(Y,2) < 0\}$$

Also, the matrices R^π are constant on each of these polytopes.

From Theorem 2.2, the polytopes $\mathcal{P}_1^\alpha, \mathcal{P}_2^\alpha$ and $\mathcal{P}_3^\alpha$ are subsets of $[0, 1]$. Under assumptions (A1) and (A2), $\mathcal{P}_3^\alpha = [0, \pi^{**}(2)), \mathcal{P}_2^\alpha = [\pi^{**}(2), \pi^*(2)), \mathcal{P}_1^\alpha = [\pi^*(2), 1]$, where π^{**} and π^* are the belief states at which $H^\alpha(2,1) = H^\alpha(2,2)$ and $H^\alpha(1,1) = H^\alpha(1,2)$ respectively.

We now illustrate the solution to the Bellman's stochastic dynamic programming equation (2.17), which determines the optimal policy for quickest market shock detection with two states. Clearly the traders and market observer interact – the local decisions a_k taken by the agents determines the public belief π_k and hence determines decision u_k of the market observer via (2.14). From Theorem 2.2 and (2.17), the value function can be written as,

$$V(\pi) = \min\{C(\pi,1), C(\pi,2) + \rho V(\pi)\mathcal{I}(\pi \in \mathcal{P}_1^\alpha) + \rho \sum_{a \in \mathcal{A}} V(T^\pi(\pi,a))\sigma(\pi,a)\mathcal{I}(\pi \in \mathcal{P}_2^\alpha) + \rho V(\pi)\mathcal{I}(\pi \in \mathcal{P}_3^\alpha)\}$$

The explicit dependence of the filter on the belief π results in discontinuous value function. The optimal policy in general has multiple thresholds and the stopping region in general is non-convex.

D: Illustrative Example

The structural results for the risk averse social learning filter, namely Theorem 2.1 and Theorem 2.2, apply to multi-state Markov chains. However, in numerical examples, to illustrate the optimal policy, we have used 2-state Markov chains. Multi-state Markov chain examples can also be considered, but the numerical solution is substantially more expensive and one has to resort to suboptimal methods such as open loop feedback control(OLFC) [18] to compute a policy.

Fig. 2.2 displays the value function and optimal policy for a toy example having the following parameters:

$$B = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}, P = \begin{bmatrix} 1 & 0 \\ 0.06 & 0.94 \end{bmatrix}, c = \begin{bmatrix} 1 & 2 \\ 2.5 & 0.5 \end{bmatrix}$$

From Fig. 2.2 it is clear that the market observer has a double threshold policy and the value function is discontinuous. The double threshold policy is unusual from a signal processing point of view. Recall that $\pi(2)$ depicts the posterior probability of no change. The market observer “changes its mind" - it switches from no change to change as the posterior probability of change decreases! Thus the global decision (stop or continue) is a non-monotone function of the posterior probability obtained from local decisions in the agent based model. The example illustrates the unusual behaviour of the social learning filter.

2.2.3 Conclusion

We provided a Bayesian formulation of the problem of quickest detection of change in the value of a stock using the decisions of risk averse traders. The main conclusions

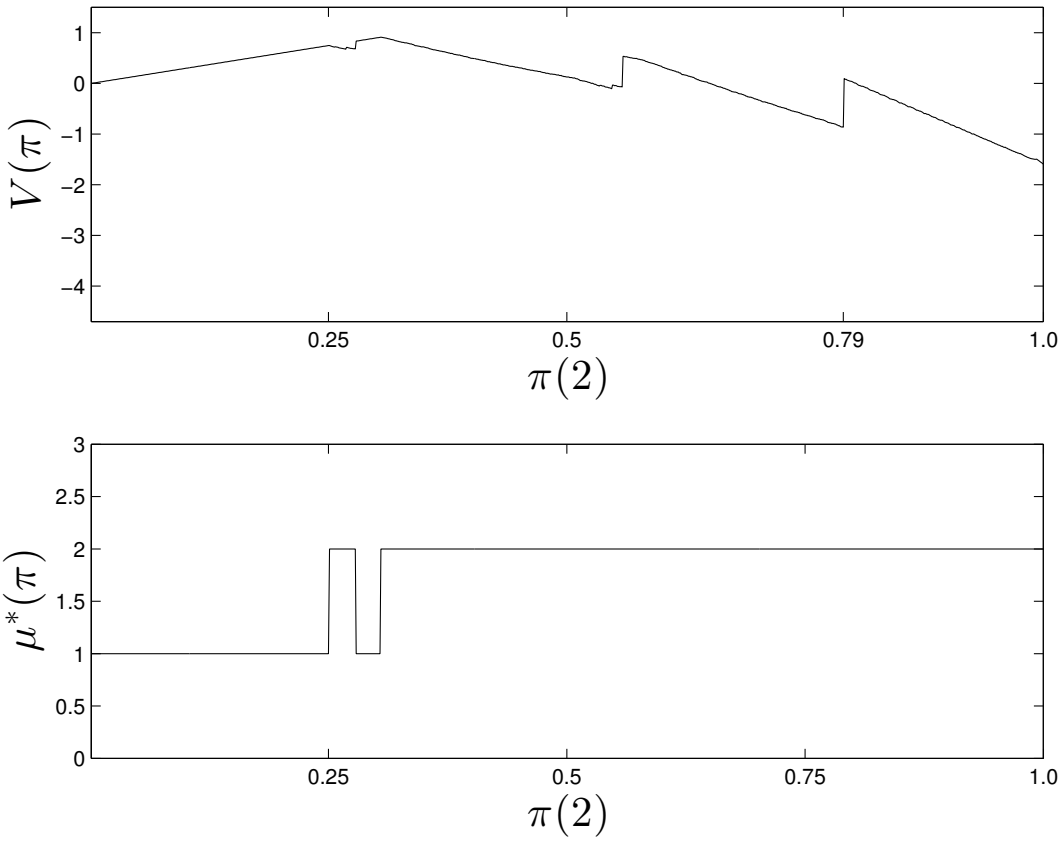

Figure 2.2: The value function $V(\pi)$ and the double threshold optimal policy $\mu^*(\pi)$ are plotted over $\pi(2)$. The parameters for the market observer are chosen as: $d = 1.25$, $\mathbf{f} = [0\ 3]$, $\alpha = 0.8$ and $\rho = 0.9$. The significance of the double threshold policy is that the stopping regions are non-convex. The implication of non-convex stopping set for the market observer is that - if it believes that it is optimal to stop, it need not be optimal to stop when his belief is larger.

drawn were:

1. Under reasonable assumptions on the costs, the trading decisions taken by risk-averse traders are ordinal functions of their private observations and monotone in the prior information. This implies that the Bayesian social learning follows simple intuitive rules.

2. Risk averse traders herd sooner and don't prefer to "learn" from the crowd. The stopping region for the sequential detection problem with risk averse traders is

non-convex; this is in contrast to standard signal processing quickest detection problems where the stopping set is convex. This affects the confidence of the market observer in implementing the optimal quickest detection policy.

There is an opportunity to apply the framework to study the behaviour of interacting sensors in online prediction markets such as Iowa Electronic Markets, Trade Sports and Foresight Exchange.

2.2.4 Appendix: Proofs

The following lemmas are required to prove Theorem 2.1 and Theorem 2.2. The results will be proved for general state and observation spaces having two actions.

Lemma 2.3. For a finite state and observation alphabet, $\underset{z\in\mathbb{R}}{\operatorname{argmin}}\{z + \frac{1}{\alpha}\mathbb{E}_y[\max\{(c(x,a) - z), 0\}]\}$ is equal to $c(i,a)$ for some $i \in \{1, 2, \ldots, X\}$.

Proof. Let η_y be the belief update (p.m.f) with observation y, i.e, $\eta_y(i) = \mathbb{P}_y(x = i)$. Let $F_y(x)$ denote the cumulative distribution function. For simplicity of notation, let $h_y(z) = z + \frac{1}{\alpha}\mathbb{E}_y[\max\{(c(x,a) - z), 0\}]$. The extremum of $h_y(z)$ is attained where the derivative is zero. It is obtained as follows.

$$
\begin{aligned}
h_y(z) &= z + \frac{1}{\alpha}\mathbb{E}_y[\max\{(c(x,a) - z), 0\}] \\
h'_y(z) &= 1 + \frac{1}{\alpha}\lim_{\Delta z\to 0}\frac{\mathbb{E}_y[\max\{c(x,a) - z - \Delta z, 0\}] - \mathbb{E}_y[\max\{(c(x,a) - z), 0\}]}{\Delta z} \\
&= 1 + \frac{1}{\alpha}\mathbb{E}_y\left(\lim_{\Delta z\to 0}\frac{\max\{c(x,a) - z - \Delta z, 0\} - \max\{(c(x,a) - z), 0\}}{\Delta z}\right) \\
&= 1 + \frac{1}{\alpha}\mathbb{E}_y\left(0 \times \mathcal{I}_{0>(c(x,a)-z)} - 1 \times \mathcal{I}_{(c(x,a)-z)>0}\right) \\
&= 1 - \frac{1}{\alpha}\mathbb{P}_y(c(x,a) > z).
\end{aligned}
$$

Also, $h_y''(z) = \frac{1}{\alpha}\frac{d}{dz}(F_y(z))$ and therefore $h_y''(z) \geq 0$. We have, $\underset{z\in\mathbb{R}}{\text{argmin}}\ \{h_y(z)\} = \{z : \mathbb{P}_y(c(x,a) > z) = \alpha\}$. Since X is a random variable, $c(x,a)$ is a random variable with realizations $c(i,a)$ for $i \in \{1,\dots X\}$. Hence $z = c(i,a)$ for some $i \in \{1,2,\dots,X\}$. □

The result of Lemma 2.3 is similar to Proposition 8 in [112]. It was shown in [87] that $\eta_{y+1} \geq_r \eta_y$. Also, MLR dominance implies first order dominance, i.e, $\eta_{y+1} \geq_s \eta_y$.

Lemma 2.4. Let l and k be the indices such that

$$\underset{z\in\mathbb{R}}{\text{argmin}}\{h_y(z)\} = c(l,a)$$

$$\underset{z\in\mathbb{R}}{\text{argmin}}\{h_{y+1}(z)\} = c(k,a)$$

For all $y \in \{1,2\dots,Y\}$, $k \geq l$.

Proof. Proof is by contradiction. From Lemma (2.3), we have $F_y(c(l,a)) = 1-\alpha$ and $F_{y+1}(c(k,a)) = 1-\alpha$. Suppose $l > k$. We know that $F_{y+1}(z)$ is a monotone function in z. Since $l > k$, $F_{y+1}(c(l,a)) > 1-\alpha$. But, by definition of first order stochastic dominance, $F_y(z) \geq F_{y+1}(z)$ for all z. Therefore, $F_y(c(l,a)) \geq F_{y+1}(c(l,a)) > 1-\alpha$, a contradiction. □

From Lemma 2.3 and equation (2.19), we have

$$H^\alpha(y,2) = c(l,2) + \frac{1}{\alpha}\sum_{i=1}^{l-1}\eta_y(i)(c(i,2) - c(l,2)),$$

$$H^\alpha(y+1,2) = c(k,2) + \frac{1}{\alpha}\sum_{i=1}^{k-1}\eta_{y+1}(i)(c(i,2) - c(k,2))$$

Lemma 2.5. $H^\alpha(y,2) \geq H^\alpha(y+1,2)$ if $\alpha \geq 1 - \mathbb{P}_y(x = X)$.

Proof. From the definitions of $H^\alpha(y,2)$ and $H^\alpha(y+1,2)$ we have,

$$\begin{aligned} H^\alpha(y,2) - H^\alpha(y+1,2) &= c(l,2) - c(k,2) + \\ &\frac{1}{\alpha}\sum_{i=1}^{l-1}\eta_y(i)(c(i,2) - c(l,2)) + \frac{1}{\alpha}\sum_{i=1}^{k-1}\eta_{y+1}(i)(c(k,2) - c(i,2)) \\ &\geq c(l,2) - c(k,2) + \\ &\frac{1}{\alpha}\sum_{i=1}^{l-1}\eta_y(i)(c(i,2) - c(l,2)) + \frac{1}{\alpha}\sum_{i=1}^{k-1}\eta_y(i)(c(k,2) - c(i,2)) \end{aligned} \tag{2.20}$$

Equation (2.20) follows from Lemma A.1 and can be simplified as

$$\begin{aligned} H^\alpha(y,2) - H^\alpha(y+1,2) &\geq c(l,2) - c(k,2) + \\ \frac{1}{\alpha}\sum_{i=1}^{l-1}\eta_y(i)(c(k,2) - c(l,2)) &+ \frac{1}{\alpha}\sum_{i=l}^{k-1}\eta_y(i)(c(k,2) - c(i,2)) \\ &\geq c(l,2) - c(k,2) - \frac{1}{\alpha}\Gamma'\eta_y \end{aligned}$$

where Γ is such that $\Gamma_i = c(l,2) - c(k,2)$ for $i = 1,\ldots,l-1$ and $\Gamma_i = c(i,2) - c(k,2)$ for $i = l,\ldots k-1$. Clearly, $\Gamma_i \geq 0$ and decreasing. Right hand side of inequality attains its maximum when $k = X$ and $l = 1$ and $\Gamma_i = c(l,2) - c(k,2)$ for all i. Therefore, we have

$$\begin{aligned} H^\alpha(y,2) - H^\alpha(y+1,2) &\geq c(l,2) - c(k,2) - \frac{1}{\alpha}\Gamma'\eta_y \\ &\geq (c(l,2) - c(k,2)) - \frac{1}{\alpha}(c(l,2) - c(k,2))(1 - \mathbb{P}_y(x = X)) \end{aligned}$$

After rearrangement we have,

$$H^\alpha(y,2) - H^\alpha(y+1,2) \geq \frac{\alpha - (1 - \mathbb{P}_y(x = X))}{\alpha}(c(l,2) - c(k,2))$$

Since $\alpha \geq 1 - \mathbb{P}_y(x = X)$ and $(c(l,2) - c(k,2)) \geq 0$ (follows from Lemma 2.4 and assumption (A2)), we have $H^\alpha(y,2) \geq H^\alpha(y+1,2)$. □

From Lemma 2.3 and (2.19), we have

$$H^\alpha(y,1) = c(l,1) + \frac{1}{\alpha}\sum_{i=l+1}^{X}\eta_y(i)(c(i,1)-c(l,1)),$$
$$H^\alpha(y+1,1) = c(k,1) + \frac{1}{\alpha}\sum_{i=k+1}^{X}\eta_{y+1}(i)(c(i,1)-c(k,1))$$

Lemma 2.6. $H^\alpha(y+1,1) \geq H^\alpha(y,1)$ if $\alpha \geq 1 - \mathbb{P}_{y+1}(x = X)$.

Proof. From the definitions of $H^\alpha(y+1,1)$ and $H^\alpha(y,1)$ we have,

$$\begin{aligned} H^\alpha(y+1,1) - H^\alpha(y,1) &= c(k,1) - c(l,1) + \\ \frac{1}{\alpha}\sum_{i=k+1}^{X}\eta_{y+1}(i)(c(i,1)-c(k,1)) &- \frac{1}{\alpha}\sum_{i=l+1}^{X}\eta_y(i)(c(i,1)-c(l,1)) \\ &\geq c(k,1) - c(l,1) + \\ \frac{1}{\alpha}\sum_{i=k+1}^{X}\eta_{y+1}(i)(c(i,1)-c(k,1)) &- \frac{1}{\alpha}\sum_{i=l+1}^{X}\eta_{y+1}(i)(c(i,1)-c(l,1)) \end{aligned} \tag{2.21}$$

Equation (2.21) follows from Lemma A.2 and can be simplified as

$$\begin{aligned} H^\alpha(y+1,1) - H^\alpha(y,1) &\geq c(k,1) - c(l,1) + \\ \frac{1}{\alpha}\sum_{i=k+1}^{X}\eta_{y+1}(i)(c(i,1)-c(k,1)) &- \frac{1}{\alpha}\sum_{i=l+1}^{X}\eta_{y+1}(i)(c(i,1)-c(l,1)) \\ &\geq c(k,1) - c(l,1) - \frac{1}{\alpha}\Delta'\eta_{y+1} \end{aligned}$$

where Δ is such that $\Delta_i = c(i,1) - c(l,1)$ for $i = l, \ldots, k$ and $\Delta_i = c(k,1) - c(l,1)$ for $i = k+1, \ldots X$. Clearly, $\Delta_i \geq 0$ and decreasing. Right hand side of inequality attains its maximum when $k = X$ and $l = 1$ and $\Delta_i = c(k,1) - c(l,1)$ for all i. Therefore, we have

$$H^\alpha(y+1,1) - H^\alpha(y,1) \geq (c(k,1) - c(l,1)) - \frac{1}{\alpha}(c(k,1) - c(l,1))(1 - \mathbb{P}_{y+1}(x = X))$$

After rearrangement we have,

$$H^{\alpha}(y+1,1) - H^{\alpha}(y,1) \geq \frac{\alpha - (1 - \mathbb{P}_{y+1}(x = X))}{\alpha}(c(k,1) - c(l,1))$$

Since $\alpha \geq 1 - \mathbb{P}_{y+1}(x = X)$ and $c(k,1) - c(l,1) \geq 0$ (follows from Lemma 2.4 and assumption (A2)), we have $H^{\alpha}(y+1,1) \geq H^{\alpha}(y,1)$. □

Lemma 2.7. Let $\alpha \geq (1 - \mathbb{P}_{y}(x = X))$. The function $H^{\alpha}(y,a)$ satisfies the single crossing condition i.e,

$$(H^{\alpha}(y,1) - H^{\alpha}(y,2)) \geq 0 \Rightarrow (H^{\alpha}(y+1,1) - H^{\alpha}(y+1,2)) \geq 0$$

Proof. Assume $(H^{\alpha}(y,1) - H^{\alpha}(y,2)) \geq 0$. We have,

$$\begin{aligned} H^{\alpha}(y,1) - H^{\alpha}(y,2) &\geq 0 \\ \Rightarrow H^{\alpha}(y,1) - H^{\alpha}(y+1,2) &\geq 0 \end{aligned} \tag{2.22}$$

Equation (2.22) follows from Lemma 2.5. And,

$$\begin{aligned} H^{\alpha}(y,1) - H^{\alpha}(y+1,2) &\geq 0 \\ \Rightarrow H^{\alpha}(y+1,1) - H^{\alpha}(y+1,2) &\geq 0 \end{aligned} \tag{2.23}$$

Equation (2.23) follows from Lemma 2.6. □

Lemma 2.7 is a crucial result which helps us to prove Theorem 2.1 and Theorem 2.2.

Proof of Theorem 2.1: From Lemma 2.7, $H^{\alpha}(y,a)$ satisfies the single crossing condition and hence is sub-modular in (y,a). Using Theorem A.3, we get $a^{*}(\pi,y) = \text{argmin } H^{\alpha}(y,a)$ is increasing in y.

Proof of Theorem 2.2: From Lemma 2.7, $H^{\alpha}(y,a)$ satisfies the single crossing condition. It is easily verified that the belief states satisfy the following property

$$\{\pi : H^{\alpha}(y,1) - H^{\alpha}(y,2) \geq 0\} \subseteq \{\pi : H^{\alpha}(y+1,1) - H^{\alpha}(y+1,2) \geq 0\} \tag{2.24}$$

Equation (2.24) says that the curves $\{\pi : H^{\alpha}(y,1) - H^{\alpha}(y,2) = 0\}$ for all $y \in \mathcal{Y}$ do not intersect. Also from (2.18) and (2.24), it is easily verified that there are at most $Y+1$ local decision likelihood matrices R^{π} (can be less than $Y+1$ when $H^{\alpha}(\bar{y},1) - H^{\alpha}(\bar{y},2) > 0$ for some $\bar{y} \in \mathcal{Y}$, for all π). The matrices R^{π} from (2.18) and Theorem 2.1 are constant on each of the $Y+1$ polytopes.

CHAPTER 3

STOCHASTIC CONTROL WITH SOCIAL LEARNING: FRAMEWORK-II

Framework-II: Consider the following specifications: (i) the individual decision makers act once sequentially by observing all the predecessors' actions, (ii) the individual decision maker's reward is directly influenced by the operator's control, (iii) the optimal policy of the operator is determined as the solution that achieves the stochastic control objective, which is to estimate the state with minimum incurred cost.

We first describe the problems considered in the literature that fit into Framework-II. We then illustrate an application of Framework-II in monopoly pricing[1] and information elicitation using social sensors[2].

3.1 Relevant Problems in Literature

Framework-II is applicable in a wide range of situations. Below, we discuss a few.

Information aggregation and market manipulation. Markets are susceptible to manipulation by agents who wish to distort decision making [4, 27, 59, 32, 60, 35].

Differential pricing. Here the prices of the goods or services are adjusted to the changing valuation and the stochastic demand to improve the revenue. [129, 10, 44, 131, 101] discuss the related literature and applications.

Information elicitation for decision making. Here beliefs about events such as the likely outcome of an election or sporting event are elicited and aggregated by an automated system. [36, 136, 107] consider mechanism design problems for truthful and robust

[1] Bhatt, S., & Krishnamurthy, V. Controlled information fusion with risk-averse CVaR social sensors. In 2017 IEEE 56th Annual Conference on Decision and Control (CDC) (pp. 2605-2610). IEEE.

[2] Bhatt, S., & Krishnamurthy, V. Controlled sequential information fusion with social sensors. Submitted to ACM Transactions on Economics and Computation, 2018.

information elicitation.

3.2 Application 1: Monopoly Pricing

Suppose a monopoly offers a product at a particular price to its customers. The customers who purchase the product review the product on online social media platforms. Future customers are influenced by the past reviews, and hence have a different valuation of the product. How could the monopoly alter the prices dynamically to utilize the changing valuations and possibly make a profit on the total revenue?

[23, 24] consider monopoly pricing in the presence of social learning and establish various properties of the value function and the optimal policy for the monopoly. It is shown that using discriminatory pricing the monopoly is able to delay the process of herding in risk-neutral social sensors, to suit its needs. The optimal price (control) sequence is shown to be a super-martingale. We consider monopoly pricing and social learning under CVaR risk-measure; see [91] for an overview of risk measures.

3.2.1 Formulation of Monopoly Pricing

The monopoly pricing with risk-averse customers is formulated as a partially observed Markov decision process (POMDP). We first present the Bayesian social learning model for risk-averse customers and define the objective of monopoly. The problem solution is obtained using stochastic dynamic programming.

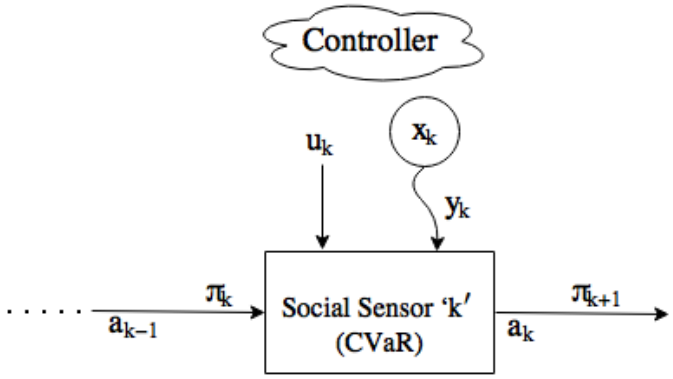

Figure 3.1: CVaR Social Learning model. The customer k receives the public belief π_{k-1} from all its predecessors. y_k denotes the private valuation of the quality and u_k denotes the price charged by the monopoly for the product/ services. The decision a_k is shared by the controller (monopoly) and the updated public belief π_{k+1} is received by the successive customers.

A: CVaR Social Learning Model and Monopoly Pricing Protocol

We consider the classical sequential social learning framework [33, 74, 76]. Fig. 3.1 shows the framework for social learning and pricing. Let $x \in \mathcal{X} = \{1(\text{Low}), 2(\text{High})\}$ denote the state. Let the initial distribution (on the quality) be denoted as $\pi_0 = (\pi_0(i), i \in \mathcal{X})$, where $\pi_0(i) = \mathbb{P}(x_0 = i)$.

Each customer acts once in a predetermined sequential order indexed by $k = 1, 2, \cdots$. The index k can also be viewed as the discrete time instant when customer k acts. Each customer k obtains noisy private valuations, $y_k \in \mathcal{Y} = \{1(\text{Low}), 2(\text{High})\}$, of the quality x_k and considers this in addition to the actions of its predecessors. The monopoly does not have any information about $x_k \in \mathcal{X}$ but infers it from the information revealed by the actions of the individual customers, $a_k \in \mathcal{A} = \{1(\text{Don't Utilize}), 2(\text{Utilize})\}$, and chooses the price inputs[3] $u_k \in [0, 1]$ at each time k (or at each sensor k). The social learning model and the pricing protocol of the monopoly is as follows:

[3]The range of prices chosen by the controller is normalized to $[0, 1]$ for convenience.

1. *Customer's Private Observation*: Customer k's private observation denoted by $y_k \in \mathcal{Y} = \{1, 2\}$ is a noisy measurement of the true quality. It is obtained from the observation likelihood distribution as,

$$B_{ij} = \mathbb{P}(y_k = j | x_k = i) \tag{3.1}$$

The discreteness of the observation distribution captures the *boundedness* or the limited processing capabilities of the customer.

2. *Social Learning and Private Belief update*: Customer k updates its private belief by fusion of the observation y_k and the prior public belief $\pi_{k-1}(i) = \mathbb{P}(x_k = i | a_1, \ldots, a_{k-1})$ as the following Hidden Markov Model (HMM) update

$$\eta_k^{y_k} = \frac{B_{y_k} \pi_{k-1}}{\mathbf{1}' B_{y_k} \pi_{k-1}} \tag{3.2}$$

where B_{y_k} denotes the diagonal matrix having $[\mathbb{P}(y_k | x_k = 1)\ \mathbb{P}(y_k | x_k = 2))]$ along the diagonal and $\mathbf{1}$ denotes the 2-dimensional vector of ones. HMM update is a consequence of Bayes' rule, information on the state conditioned on the new observation.

3. *Customer's Action*: Customer k executes an action $a_k \in \mathcal{A} = \{1, 2\}$ to myopically minimize its cost. Let $c(x_k, a_k)$ denote the cost incurred if the customer takes action a_k when the underlying state is x_k.

The form of the state-action dependent cost is taken as $c(x_k, a_k) = u_k - v(x_k)$ (see [33] for a justification), where v is the valuation of the services by each customer and u_k is the price chosen by the controller at time k. It is assumed without loss of generality that

$$v(x_k) = \begin{cases} 0 & \text{if } x_k = 1; \\ 1 & \text{if } x_k = 2. \end{cases}$$

The state-action dependent costs for $x \in \mathcal{X}$ are thus given as:

$$c(x_k, a_k) = \begin{cases} \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{if } a_k = 1; \\ \begin{bmatrix} u_k \\ u_k - 1 \end{bmatrix} & \text{if } a_k = 2. \end{cases}$$

The customer chooses an action a_k to minimize the CVaR measure as

$$\begin{aligned} a_k &= \underset{a \in \mathcal{A}}{\operatorname{argmin}} \{\text{CVaR}_\alpha(c(x_k, a))\} \\ &= \underset{a \in \mathcal{A}}{\operatorname{argmin}} \{\min_{z \in \mathbb{R}} \{z + \frac{1}{\alpha} \mathbb{E}_{y_k}[\max\{(c(x_k, a) - z), 0\}]\}\} \end{aligned} \tag{3.3}$$

Here $\alpha \in (0, 1]$ reflects the degree of risk-aversion for the customer (the smaller α is, the more risk-averse the customer is). Note that when $\alpha = 1$, the cost function is the risk-neutral cost function as in [33, 23, 24]. Define

$$\mathcal{G}_k := \sigma\text{- algebra generated by } (u_1, a_1, u_2, a_2, \ldots, u_k, y_k) \tag{3.4}$$

$\mathbb{E}_{y_k}$ denotes the expectation with respect to private belief, i.e, $\mathbb{E}_{y_k} = \mathbb{E}[.|\mathcal{G}_k]$ when the private belief is updated after observation y_k.

4. *Monopoly Reward*: We consider two possible reward functions for the monopoly. The monopoly chooses one of the following reward functions at $k = 0$ and accrues the corresponding reward at each time k as

   Case 1. Self-Interested: The monopoly accrues a reward when the customers utilize the services,

   $$r_{u_k} = (u_k - \beta)\mathcal{I}(a_k = 2|\pi_k). \tag{3.5}$$

   Case 2. Altruistic: The monopoly accrues a reward when the customers act according to their valuations,

   $$r_{u_k} = (u_k - \beta)\mathcal{I}(a_k = y_k|\pi_k). \tag{3.6}$$

Here $\mathcal{I}$ denotes the indicator function and $\beta \in (0,1)$ is a fixed[4] cost incurred by the monopoly. It could denote the cost of service. The monopoly being self-interested can be seen as profit maximizing, and being altruistic can be seen as social welfare maximizing[5].

5. *Public Belief update*: Customer k's action is shared by the monopoly on public platforms and the public belief on the quality is updated according to the social learning Bayesian filter (see [74, 76]) as follows

$$\pi_k = T^{\pi}(\pi_{k-1}, a_k) = \frac{R_{a_k}^{\pi_{k-1}} \pi_{k-1}}{\mathbf{1}' R_{a_k}^{\pi_{k-1}} \pi_{k-1}}. \tag{3.7}$$

Here, $\pi_k(i) = \mathbb{P}(x_k = i | a_1, \ldots, a_k)$, $R_{a_k}^{\pi_{k-1}} = \text{diag}(\mathbb{P}(a_k | x = i, \pi_{k-1}), i \in \mathcal{X})$, where $\mathbb{P}(a_k | x = i, \pi_{k-1}) = \sum_{y \in \mathcal{Y}} \mathbb{P}(a_k | y, \pi_{k-1}) \mathbb{P}(y | x_k = i)$ and

$$\mathbb{P}(a_k | y, \pi_{k-1}) = \begin{cases} 1 & \text{if } a_k = \underset{a \in \mathcal{A}}{\text{argmin}} \{\text{CVaR}_\alpha(c(x_k, a))\}; \\ 0 & \text{otherwise.} \end{cases}$$

Note that π_k belongs to the unit simplex

$$\Pi(2) \stackrel{\Delta}{=} \{\pi \in \mathbb{R}^2 : \pi(1) + \pi(2) = 1, 0 \leq \pi(i) \leq 1 \text{ for } i \in \{1, 2\}\}$$

Here the expectation in CVaR measure is with respect to the sigma-algebra $\mathcal{G}_k$. Social learning filter update is a consequence of Bayes' rule, information on the state conditioned on the new action.

6. *Monopoly Price*: Let the history recorded by the monopoly be denoted as $\mathcal{H}_k = \{\pi_0, u_1, a_1, \cdots, u_k, a_k\}$. The monopoly chooses $u_{k+1} = \mu_{k+1}(\mathcal{H}_k) \in [0, 1]$ for the customer $k+1$ and the protocol is repeated for all the customers in the system. Here μ_{k+1} denotes the pricing policy at time $k+1$.

[4]Note that β could be made state dependent without affecting the nature of the results in this work. Here it is assumed to be independent of the state for simplicity.

[5]We shall see later that $\mathcal{I}(a = y)$ improves the value of information fused by the successive customers, thereby promoting welfare.

B: Monopoly Pricing Objective

The monopoly chooses the price offered to customers sequentially as

$$u_k = \mu_k(\mathcal{H}_{k-1}) \in [0, 1] \tag{3.8}$$

where $\mathcal{H}_k = \{\pi_0, u_1, a_1, \cdots, u_{k-1}, a_{k-1}\}$. Since $\mathcal{H}_k$ is increasing with time k, it is useful to obtain a sufficient statistic that does not grow in dimension. The public belief π_{k-1} computed via the social learning filter (3.25) forms a sufficient statistic for $\mathcal{H}_k$ and[6] (3.8) can be written as

$$u_k = \mu_k(\pi_{k-1}). \tag{3.9}$$

The monopoly maximizes the cumulative discounted reward

$$J_\mu(\pi) = \mathbb{E}_\mu\{\sum_{k=1}^{\infty} \rho^k r_{u_k} | \pi_0 = \pi\}. \tag{3.10}$$

Here $u_k = \mu(\pi_{k-1})$ and $\rho \in [0, 1)$ denotes the economic discount factor indicating the degree of impatience of the monopoly. In (3.10), the monopoly seeks to find the optimal stationary policy μ^* such that

$$J_{\mu^*}(\pi_0) = \sup_{\mu \in \boldsymbol{\mu}} J_\mu(\pi_0). \tag{3.11}$$

in a class of stationary policies $\boldsymbol{\mu}$.

[6]The rewards are a function of the price and the state (see Lemma 3.6 and Lemma 3.7), and hence restriction to Markov policies is without loss of generality.

C: Stochastic Dynamic Programming Formulation

The optimal policy μ^* and the value function $V(\pi)$ for the POMDP satisfy the Bellman's dynamic programming equation

$$\begin{aligned} Q(\pi, u) &= r_u + \rho \sum_{a \in \mathcal{A}} V(T^\pi(\pi, a))\sigma(\pi, a), \\ \mu^*(\pi) &= \arg\max_{u \in [0,1]} Q(\pi, u), \\ V(\pi) &= \max_{u \in [0,1]} Q(\pi, u), \quad J_{\mu^*}(\pi_0) = V(\pi_0). \end{aligned} \tag{3.12}$$

Here r_u is the instantaneous reward accrued by the monopoly, $T^\pi(\pi, a) = \frac{R_a^\pi P'\pi}{\mathbf{1}' R_a^\pi P'\pi}$ is the CVaR-social learning filter and $\sigma(\pi, a) = \mathbf{1}' R_a^\pi P'\pi$ is the normalization factor of the Bayesian update.

3.2.2 Main Results

In this section, we characterize the nature of the optimal pricing policy for (3.10). It is shown that due to the structure of the social learning filter, the choice of price inputs reduces from a continuum to a finite number at every belief.

Assumptions

(A1) The observation distribution $B_{xy} = \mathbb{P}(y|x)$ is TP2(total positive of order 2), i.e, the determinant of the matrix B is non-negative; see [75].

Theorem 3.1. Given a risk-aversion factor $\alpha \in (0, 1]$, let $u^H(\pi) = 1 - \frac{\eta^{y=2}(1)}{\alpha}$ and $u^L(\pi) = 1 - \frac{\eta^{y=1}(1)}{\alpha}$ denote two possible prices at the belief π. Under (A1), the Q function (3.12) can be simplified for the rewards (3.5) and (3.6) as

Case 1.) Self-Interested:

$$Q(\pi, u) = \begin{cases} (u - \beta) + \rho V(\pi) & \text{if } u \in [0, u^L(\pi)]; \\ (u - \beta) \times \mathbf{1}' B_{y=2}\pi + \rho \mathbb{E}V(\pi) & \text{if } u \in (u^L(\pi), u^H(\pi)]; \\ 0 & \text{otherwise.} \end{cases}$$

and $V(\pi) = \max Q(\pi, u)$, where $V(\pi) \geq 0$.

Case 2.) Altruistic:

$$Q(\pi, u) = \begin{cases} (u - \beta) + \rho \mathbb{E}V(\pi) & \text{if } u \in (u^L(\pi), u^H(\pi)]; \\ 0 & \text{otherwise.} \end{cases}$$

and $V(\pi) = \max Q(\pi, u)$, where $V(\pi) \geq 0$.

Here,

$$\mathbb{E}V(\pi) = \mathbf{1}' B^{\pi}_{y=1}\pi \times V(\eta^{y=1}) + \mathbf{1}' B^{\pi}_{y=2}\pi \times V(\eta^{y=2}).$$

The prices $u^H(\pi)$ and $u^L(\pi)$ are such that customers utilize the services when $y = 2$ and $y = \{1, 2\}$ respectively. Theorem 3.14 represents the Q function (3.32) over a price input range $[0, 1]$ for rewards (3.5) and (3.6) respectively, in *three* and *two* regions. The following corollaries highlight why such partitions are useful.

Corollary 3.2.1. Let the monopoly reward be given by (3.5). At every public belief $\pi \in \Pi(2)$, it is sufficient to choose one of the three prices $\{u^L(\pi), u^H(\pi), u^H(\pi) + \epsilon\}$ for any $\epsilon > 0$. □

Corollary 3.2.2. Let the monopoly reward be given by (3.6). At every public belief $\pi \in \Pi(2)$, it is sufficient to choose one of the two prices $\{u^H(\pi), u^H(\pi) + \epsilon\}$ for any $\epsilon > 0$. □

The following theorem completely characterizes the optimal pricing policy when the monopoly aims to maximize the reward.

Theorem 3.2. For every public belief $\pi \in \Pi(2)$ and an $\epsilon > 0$, the optimal policy $\mu^*(\pi) = \arg\max_u Q(\pi, u)$ is given as

Case 1.) Self-Interested:

$$\mu^*(\pi) = \begin{cases} u^H(\pi) + \epsilon & \text{if } \pi(2) \in [0, \pi^*(2)); \\ u^H(\pi) & \text{if } \pi(2) \in [\pi^*(2), \pi^{**}(2)); \\ u^L(\pi) & \pi(2) \in [\pi^{**}(2), 1]. \end{cases} \tag{3.13}$$

for $\pi^*(2),\ \pi^{**}(2) \in [0, 1]$.

Case 2.) Altruistic:

$$\mu^*(\pi) = \begin{cases} u^H(\pi) + \epsilon & \text{if } \pi(2) \in [0, \hat{\pi}^*(2)); \\ u^H(\pi) & \text{if } \pi(2) \in [\hat{\pi}^*(2), 1]. \end{cases} \tag{3.14}$$

for $\hat{\pi}^*(2) \in (0, 1)$.

From Theorem 3.14, Corollary 3.3.2 and Corollary 3.2.2, the value function in (3.32) can be represented as

(Self-Interested)

$$V(\pi) = \max\{0, (u^L(\pi) - \beta) + \rho V(\pi), (u^H(\pi) - \beta) \times \mathbf{1}' B_{y=2}\pi + \rho \mathbb{E}V(\pi)\}$$

(Altruistic)

$$V(\pi) = \max\{0, (u^H(\pi) - \beta) + \rho \mathbb{E}V(\pi)\}$$

The key takeaway is that due to the structure of the social learning filter, the choice of price inputs at every belief is reduced to a finite number of values instead of the range $[0, 1]$. Characterizing the optimal policy amounts to selecting among the these price inputs as a function of the public belief. Theorem 3.8 completely determines the regions in the belief space $\Pi(2)$ where it is optimal to choose a particular price input. Fig. 3.2 and Fig. 3.3 show the value function and the optimal policy for two different risk-aversion factors (α) in a simple numerical example.

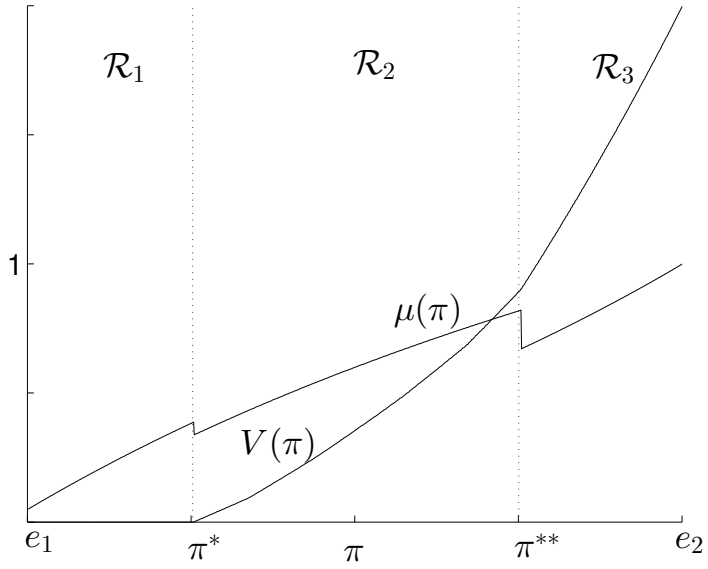

Figure 3.2: Risk-aversion factor $\alpha = 0.9$

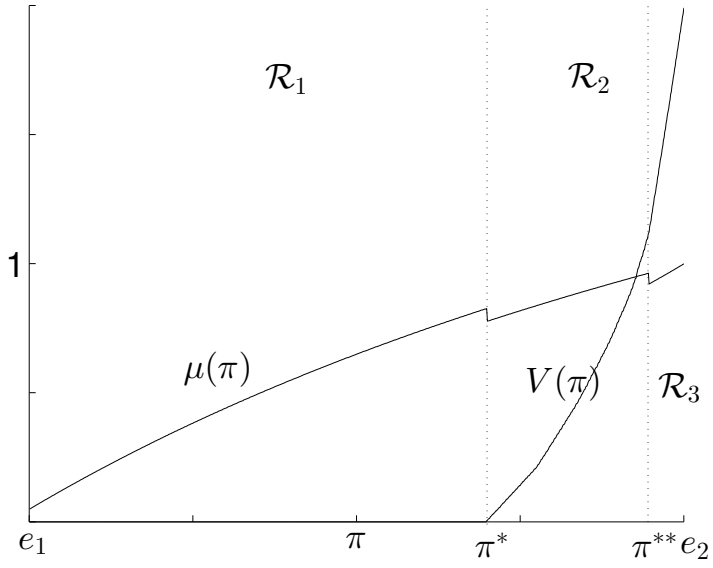

Figure 3.3: $\alpha = 0.3$

Value function and optimal pricing policy of the monopoly in the *self-interested* case. $B = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$, the discount factor $\rho = 0.7$, and $\mathcal{R}_1 - [0, \pi^*)$, $\mathcal{R}_2 - [\pi^*, \pi^{**})$, and $\mathcal{R}_3 - [\pi^{**}, 1]$ are the cut-off, social learning and herding regions respectively. It can be seen that the width of $\mathcal{R}_1$ increases with increased aversion to risk. This is equivalent to saying that risk-averse customers that show an increased aversion to risk, choose to utilize the services only when they are reasonably certain about the quality. So it is profitable to the monopoly if it offers services only when it believes that the quality is high.

Let $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$ denote the three regions determined by Theorem 3.8 where $u^H(\pi) + \epsilon$, $u^H(\pi)$, $u^L(\pi)$ respectively are optimal. $\mathcal{R}_1$ is the *cut-off* region - the monopoly terminates the services to the customers. In the *Self-Interested* case, the price inputs are such that no customer has an incentive to utilize the services. $\mathcal{R}_2$ is the *social learning* region - the customers act according to their private valuations. The price inputs are such that the customer having a high valuation $y = 2$ will utilize the services, while the customer having low valuation $y = 1$ finds it prohibitive. Since the customers act according to their valuation, customer deciding at a future instant can successfully infer the private valuation of its predecessors; in other words, the information fusion reduces uncertainty about the quality of the service. $\mathcal{R}_3$ is the *herding* region - every customer utilizes the services. The monopoly chooses a low input $u^L(\pi)(< u^H(\pi))$, which prompts the customer with even a low valuation $y = 1$ to utilize the service. Notice that when the monopoly chooses $u = u^L(\pi)$ (when $\pi(2) \in [\pi^{**}(2), 1]$), the value function is $V(\pi) = \frac{(u^L(\pi)-\beta)}{(1-\alpha)}$ - a fixed payoff. This means the monopoly induces a herd (customers choose the same action irrespective of their private valuation) that leads to an information cascade (information fusion results in no improvement in uncertainty) - public belief is frozen.

In the *Altruistic* case, the price inputs (two at every belief) are chosen so as to encourage the customers to act according to their valuations. This implies that the monopoly

chooses inputs to maximize the width of the *social learning* region $\mathcal{R}_2$. The *herding* region $\mathcal{R}_3$ is absent as $u = u^L(\pi)$ is not chosen by the monopoly. The *cut-off* region indicates the flexibility to terminate the services when the expected valuation is less than the cost of service.

Theorem 3.3. Let $\mathcal{F}_k$ be the σ-algebra generated by $(u_1, a_1, u_2, a_2, \ldots, u_{k-1}, a_{k-1}, u_k, a_k)$, where π_0 is the initial belief. The optimal price sequence $u_k = \mu^*(\pi_{k-1})$ is a supermartingale[7] when the quality is a random variable for any $\alpha \in (0, 1]$.

When the monopoly is profit maximizing or self-interested, it initially chooses higher price inputs to encourage customers with higher valuation to utilize the services. Decisions at higher prices are more informative[8], which in turn results in higher public beliefs when $a = 2$. Due to the concavity of the pricing policy, higher belief causes the future price inputs to increase. Once sufficient information about the quality is accumulated, the monopoly either chooses low price inputs to allow every customer to utilize the services or terminates its services. When the monopoly is altruistic, it always chooses high price inputs to encourage the customers to act according to their private valuations.

3.2.3 Conclusions

We considered the problem of monopoly pricing risk-averse customers that are learning and influencing each other. The following conclusions were drawn:

[7] Decreases on average over time.

[8] Informativeness is in the sense of Blackwell; see [75]. For any two observation matrices B_1 and B_2, B_1 is more informative than B_2 in the Blackwell sense ($B_1 \succ_B B_2$) if $B_2 = B_1 Q$, for any stochastic matrix Q. Note here that when $u = u^H(\pi)$, the action likelihood matrix in (3.25) $R^H = B$; and when $u = u^L(\pi)$, the action likelihood matrix $R^L = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. We have for $Q = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, $R^L = R^H Q \Rightarrow R^H \succ_B R^L$.

1. The monopoly can use price differentiation to improve the revenue, when the risk-averse customers are learning from and influencing each other.
2. The time path of the prices offered to the risk-averse customers is a super-martingale– it decreases on average. This implies that the monopoly should start with a high price to establish an elite clientèle, and then reduces the price to capture the market.

3.2.4 Appendix: Proofs

Lemma 3.4 ([75])**.** Let η^y denote the private belief update (3.22) with a public prior belief π. Under (A1), η^y is increasing[9] in y, i.e, $\eta^{y=1}(1) > \eta^{y=2}(1)$. □

Theorem 3.5 ([75])**.** Let the instantaneous rewards be non-decreasing in π. Under (A1), the value function $V(\pi)$ with finite number of actions at every belief, is monotone and convex. □

Lemma 3.6. The instantaneous reward $(u-\beta)\mathcal{I}(a=2|\pi)$ is given as

$$\sum_{j\in\mathcal{Y}}\sum_{i\in\mathcal{X}}(u-\beta)\mathcal{I}(u\leq 1-\frac{\eta^{y=j}(1)}{\alpha})B_{ij}\pi(i). \tag{3.15}$$

Lemma 3.7. The instantaneous reward $(u-\beta)\mathcal{I}(a=y|\pi)$ is given as

$$(u-\beta)\ \mathcal{I}(u^L(\pi)<u\leq u^H(\pi)). \tag{3.16}$$

The proofs follow from the structure of the social learning filter (see Theorem 2, [76]), property of the CVaR measure (see Lemma 6, [76]), and Bayes' rule. It is omitted.

[9] $\pi_2 \geq \pi_1$ if the determinant

$$\begin{vmatrix}\pi_1(1) & \pi_1(2)\\ \pi_2(1) & \pi_2(2)\end{vmatrix} \geq 0$$

We will prove Theorem 3.14 and Theorem 3.8 for the *Self-Interested* case. The proof for the *Altruistic* case follows similarly.

Proof of Theorem 3.14:

Consider $Q(\pi, u)$ as in (3.30) for $u \in [0, 1]$.

i.) Let $u \in [0, u^L(\pi)]$. Recall that $u^L(\pi) = 1 - \frac{\eta^{y=1}(1)}{\alpha}$. The instantaneous reward in (3.15) is $(u - \beta)$. The continuation payoff $\sum_{a \in \mathcal{A}} V(T^\pi(\pi, a))\sigma(\pi, a)$ is given as follows. From (3.25), $R_a^\pi = \begin{bmatrix} 0 & 1 \\ & \\ 0 & 1 \end{bmatrix}$.

$$(\Rightarrow) \sum_{a \in \mathcal{A}} V(T^\pi(\pi, a))\sigma(\pi, a) = V(\pi).$$

$$\therefore Q(\pi, u) = (u - \beta) + \rho V(\pi). \tag{3.17}$$

ii.) Let $u \in (u^L(\pi), u^H(\pi)]$. The instantaneous reward in (3.15) is $(u - \beta) \times \mathbf{1}' B_{y=2}\pi$. From (3.25), $R_a^\pi = B$.

$$(\Rightarrow) \sum_{a \in \mathcal{A}} V(T^\pi(\pi, a))\sigma(\pi, a) = \mathbb{E}V(\pi).$$

$$\therefore Q(\pi, u) = (u - \beta) \times \mathbf{1}' B_{y=2}\pi + \rho \mathbb{E}V(\pi). \tag{3.18}$$

iii.) Let $u \in (u^H(\pi), 1]$. This implies that $u > 1 - \frac{\eta^{y=2}(1)}{\alpha}$. The instantaneous reward in (3.15) is 0. From (3.25), $R_a^\pi = \begin{bmatrix} 1 & 0 \\ & \\ 1 & 0 \end{bmatrix}$. Since $\mathbb{P}(a = 1) = 1$, the controller doesn't accrue any profit by offering services. Therefore the instantaneous and continuation payoff is 0.

$$(\Rightarrow) Q(\pi, u) = 0. \tag{3.19}$$

The result follows from (3.17), (3.18) and (3.19). □

Proof of Theorem 3.8:

Define the following:

$$\delta^* = \min\{\pi(2)|\ \eta^{y=1}(2) \geq 1-\alpha\},$$

$$\gamma^* = \{\pi|(u^H(\pi)-\beta)\times \mathbf{1}'B_{y=2}\pi + \rho\mathbb{E}V(\pi) = 0\},$$

$$\pi^*(2) = \max\{\delta^*, \gamma^*\},$$

$$\pi^{**}(2) = \{\pi(2)|(u^L(\pi)-\beta) + \rho V(\pi) = (u^H(\pi)-\beta)\times \mathbf{1}'B_{y=2}\pi + \rho\mathbb{E}V(\pi)\}.$$

i.) Consider $\pi(2) \in [0, \pi^*(2))$. We will show that $\{\max\ Q(\pi, u) = 0\}$.

Let $V(0)$ denote the value at $\pi = \begin{bmatrix}1\\0\end{bmatrix}$. As $\pi(2) \to 0$, we have $\mathbb{E}V(0) \to V(0)$ and $u^H(0) = u^L(0) \to 1 - \frac{1}{\alpha}$.

Assume on the contrary $V(\pi) = (u^L(\pi)-\beta) + \rho V(\pi)$. As $\pi(2) \to 0$, $V(0) = \frac{(1-\frac{1}{\alpha}-\beta)}{(1-\rho)}$. Since $\alpha \in (0,1]$, $\frac{1}{\alpha} \geq 1$ and $V(0) < 0$. From Theorem 3.14, $V(\pi) \geq 0$. Contradiction.

Similarly if $V(\pi) = (u^H(\pi)-\beta)\times \mathbf{1}'B_{y=2}\pi + \rho\mathbb{E}V(\pi)$, we have $V(0) < 0$. Therefore, $V(0) = 0$ and we have $Q(0, u^H(0)) < 0$ and $Q(0, u^L(0)) < 0$.

From the convexity of the value function, $\mathbb{E}V(\pi) \geq V(\pi)$. Since $Q(\pi, u^H(\pi)) < 0$ for $\pi(2) = [0, \pi^*(2))$, by definition of $\pi^*(2)$, we have

$$(u^H(\pi)-\beta)\times \mathbf{1}'B_{y=2}\pi + \rho\mathbb{E}V(\pi) < 0$$

$$V(\pi) \geq 0 \to \mathbb{E}V(\pi) \geq 0 \text{ by Jensen's Inequality.}$$

$$\therefore (u^H(\pi)-\beta) < 0.$$

$$(u^H(\pi)-\beta) < 0 \to (u^L(\pi)-\beta) < 0 \text{ from Lemma 3.4.}$$

If on the contrary $V(\pi) = (u^L(\pi)-\beta) + \rho V(\pi)$, then $V(\pi) < 0$; a contradiction.

$$\therefore Q(\pi, u^L(\pi)) < 0 \text{ for all } \pi(2) = [0, \pi^*(2)].$$

$$\Rightarrow V(\pi) = 0 \text{ for all } \pi(2) = [0, \pi^*(2)].$$

ii.) $\pi^{**}(2) = \{\pi(2)|Q(\pi, u^H(\pi) = Q(\pi, u^L(\pi)))\}$. We will show that for $\pi(2) \in (\pi^{**}(2), 1]$, $Q(\pi, u^L(\pi))) > Q(\pi, u^H(\pi))) > 0$.

Assume $Q(\pi, u^H(\pi)) > Q(\pi, u^L(\pi))$ on the contrary. Consider $\pi(2) \to 1$. Let $V(1)$ and $b(= \mathbf{1}' B_{y=2}\pi) \in [0,1]$ denote the values at $\pi = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. We have

$$u^H(1) = u^L(1) \to 1 \text{ and } \mathbb{E}V(1) \to V(1)$$

$$\Rightarrow (1-\beta) \times b + \rho \mathbb{E}V(1) > (1-\beta) + \rho V(1)$$

$$\Rightarrow b > 1, \text{ a contradiction as } \beta > 0.$$

$$\Rightarrow Q(\pi, u^L(\pi)) > Q(\pi, u^H(\pi)).$$

From Theorem 3.14, $V(\pi) \geq 0$ and therefore, $\mathbb{E}V(\pi) \geq 0$.

$$\text{For } \pi(2) \in [\pi^{**}(2), 1], \ (u^H(\pi) - \beta) \times \mathbf{1}' B_{y=2}\pi > 0$$

$$(\Rightarrow) Q(\pi, u^H(\pi)) > 0.$$

iii.) Since $\pi^{**}(2) = \{\pi(2) | Q(\pi, u^H(\pi) = Q(\pi, u^L(\pi)))\}$ and $Q(\pi, u^L(\pi)) < 0$ for all $\pi(2) = [0, \pi^*(2)]$, from part (ii) we have

$$Q(\pi, u^H(\pi))) > Q(\pi, u^L(\pi))) \text{ for all } \pi(2) \in [\pi^*(2), \pi^{**}(2)).$$

Note that $Q(\pi, u^H(\pi))) > 0$ for all $\pi(2) \in [\pi^*(2), \pi^{**}(2))$ by definition of $\pi^*(2)$ and the fact that $Q(\pi, u) \uparrow \pi$ (Theorem 3.5). □

Proof of Theorem 3.3:

The public belief π_k is a martingale when the state is a random variable, i.e, $\mathbb{E}[\pi_{k+1} | \mathcal{F}_k] = \pi_k$; see [33, 24].

It can easily be verified[10] that $u^H(\pi)$ is a concave function and $u^L(\pi)$ is a convex function of π for $\alpha \in (0,1]$.

i.) *Self-Interested*: For $\epsilon \to 0$, we have for $\pi_k(2),\ \pi_{k+1}(2) \in [0, \pi^{**}(2))$, $u_k = u^H(\pi_k)$ and it satisfies $\mathbb{E}[u^H(\pi_{k+1}) | \mathcal{F}_k] \leq u_k$ by Jensen's inequality.

[10]Note that the matrix B is TP2. It can be seen that derivative of $u^H(\pi)$ is strictly decreasing and the derivative of $u^L(\pi)$ is strictly increasing with respect to $\pi(2)$ for any $\alpha \in (0,1]$.

We know that $u^L(\pi) \leq u^H(\pi)$ from Lemma 3.4. For the case of $\pi_k(2) \in [\pi^*(2), \pi^{**}(2))$ and $\pi_{k+1}(2) \in [\pi^{**}(2), 1]$, we have

$$\mathbb{E}[u_{k+1}|\mathcal{F}_k] = \mathbb{E}[u^L(\pi_{k+1})|\mathcal{F}_k] \leq \mathbb{E}[u^H(\pi_{k+1})|\mathcal{F}_k] \leq u_k.$$

Note that the belief is frozen in $[\pi^{**}(2), 1]$, so $\pi_{k+1}(2) \in [\pi^*(2), \pi^{**}(2))$ and $\pi_k(2) \in [\pi^{**}(2), 1]$ is irrelevant.

ii.) *Altruistic*: Here $\pi^{**}(2) = 1$. For $\epsilon \to 0$, we have for $\pi_k(2)$, $\pi_{k+1}(2) \in [0, 1]$, $u_k = u^H(\pi_k)$ and it satisfies $\mathbb{E}[u^H(\pi_{k+1})|\mathcal{F}_k] \leq u_k$ by Jensen's inequality.

3.3 Application 2: Honest Information Elicitation

Suppose online social media platforms like Yelp or TripAdvisor offer incentives to the consumers of a product or a service to reveal a truthful account of their experiences in using the product or services. This informative (as it is honest) feedback from the social sensors can be used by the retailers to improve the quality of their product or services, and it will also benefit the future customers in that they are well informed before making a decision. In this sense, the objective of truthful information elicitation improves the overall welfare. How should the fusion center (Yelp or TripAdvisor) dynamically alter the incentives to improve the welfare[11]?

The problem of information fusion with social sensors considered in this paper deals with combining the decisions of the social sensors, correlated due to social learning, to make an informed decision regarding the underlying state (that is being estimated). Information fusion with social sensors is challenging due to the fact that social learning leads to inefficiencies [19, 12, 134] like herds (sensors choose the same action irrespec-

[11]Social welfare is maximized when the fusion center and the customers take decisions considering network externalities; see (Chapter 4, [33]).

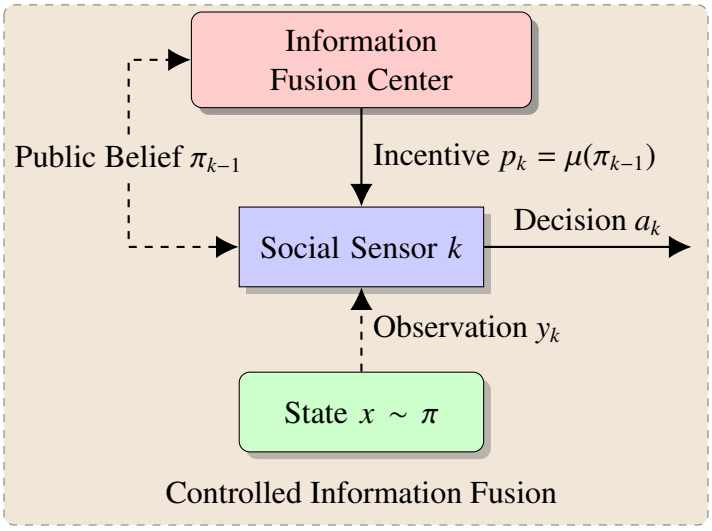

Figure 3.4: A sequence of social sensors perform Bayesian social learning to estimate the underlying state x, and take a decision a_k after myopically optimizing a reward function. The fusion center provides incentives $p_k \in [0, 1]$ at each time k (or at each sensor k) and fuses the information gathered in a Bayesian way. Each incentive p_k is computed as a function μ of the posterior probability mass function (public belief) of the state π_{k-1} at time $k-1$. The public belief π_{k-1} is computed from the decisions of the first $k-1$ sensors. The decision a_k of social sensor k depends on the incentive p_k, the public belief π_{k-1}, and the private observation y_k of the state x.

tive of their private information) and informational cascades (information fusion results in no improvement in uncertainty). So having more social sensors need not always be advantageous (in terms of reduced mean square error between the state estimate and the true state).

Multi-sensor data fusion [130] on the other hand, refers to the problem of data acquisition, processing, and fusion of information, to provide a better estimate of the underlying state. A data fusion center gathers the information from the peripheral sensors (physical sensors) to make an informed decision regarding the desired parameter. Having more number of sensors leads to improvement in reliability, resolution, coverage, and confidence; see [130].

Traditionally, information fusion is open-loop; we use feedback control to choose incentives to control how the sensors provide information. Hence we name the problem considered in this paper as *controlled information fusion*. The fusion is Bayesian and we are interested in designing the control laws for providing optimal incentives for social sensors that will result in accurate Bayesian estimates.

3.3.1 Formulation of Honest Information Elicitation

We consider the setup illustrated in Fig. 3.4. The fusion center controls the incentives given to the social sensors, and the social sensors share their decisions (quantized information on the underlying state) with the fusion center. The controlled information fusion problem is formulated as a partially observed Markov decision process (POMDP) to optimize the trade-off between the cost of information acquisition from the social sensors versus the usefulness of the information measured.

A: Controlled Fusion Social Learning Model

Let $k = 1, 2, \cdots$ denote the discrete time instants. It is assumed that each sensor decides once in a predetermined sequential order indexed by k. Let $x_0 (= x) \in \mathcal{X} = \{1, 2\}$ denote the state of nature, and is assumed to be a *random variable*[12] chosen at $k = 0$.

Let the probability mass function of the state x at time $k-1$ be denoted as

$$\pi_{k-1}(i) = \mathbb{P}(x = i | a_1, \ldots, a_{k-1}). \tag{3.20}$$

The state estimate (3.20) is computed from the decisions of the social sensors $a_1, \ldots, a_{k-1}$ and is termed as the *public belief*. Let the initial estimate be denoted as

[12] Sec.3.3.6 discusses the estimation problem when the state is changing according to a Markov chain.

$\pi_0 = (\pi_0(i), i \in \mathcal{X})$, where $\pi_0(i) = \mathbb{P}(x = i)$. Let the belief space, i.e, the set of distributions π over the state be denoted as

$$\Pi(2) \stackrel{\Delta}{=} \{\pi \in \mathbb{R}^2 : \pi(1) + \pi(2) = 1, 0 \leq \pi(i) \leq 1 \text{ for } i \in \{1, 2\}\}.$$

Social Sensor Dynamics: A social sensor, unlike a physical sensor, has its own dynamics since it learns from previous actions of other social sensors. It receives an observation on the underlying state, computes an estimate (private belief) using the information revealed by other sensors (their decisions), and takes an action to myopically maximize a reward function. This action/ decision is a quantization of the (private) belief, and is shared with the fusion center and other sensors.

1.) *Social Sensor's Private Observation*: Each social sensor k's obtains a noisy $y_k \in \mathcal{Y} = \{1, 2\}$ of the underlying state x with the observation likelihood distribution:

$$B_{ij} = \mathbb{P}(y_k = j | x = i). \tag{3.21}$$

The (discrete) observation likelihood distribution models the (limited) information gathering capabilities of the sensor.

2.) *Social Learning and Private Belief update*: Sensor k updates its private belief η_{y_k} by fusing observation y_k and the prior public belief π_{k-1}, via the following classical Bayesian update

$$\eta_{y_k} = \frac{B_{y_k} \pi_{k-1}}{\mathbf{1}' B_{y_k} \pi_{k-1}} \tag{3.22}$$

where B_{y_k} denotes the diagonal matrix $\begin{bmatrix} \mathbb{P}(y_k|x=1) & 0 \\ 0 & \mathbb{P}(y_k|x=2) \end{bmatrix}$ and $\mathbf{1}'$ denotes the 2-dimensional row vector of ones.

3.) *Social Sensor's Action*: Sensor k executes an action $a_k \in \mathcal{A} = \{1, 2\}$ myopically

to maximize a reward[13] function. The decision a_k of social sensor k is given by:

$$a_k = \arg\max_{a \in \mathcal{A}} r_a' \eta_{y_k}. \tag{3.23}$$

Here $r_a = [r(1,a),\ r(2,a)]$, with r_a' denoting the transpose of the reward vector. We consider

$$r(1,a) = \delta_a p_k + \Gamma_{1a},\ \ r(2,a) = \delta_a p_k + \Gamma_{2a},\ \ \text{with } \Gamma_{xa} = -\alpha_a \mathcal{I}(a \neq x) - \gamma_a. \tag{3.24}$$

Here $\delta_a \in [0,1]$, $\alpha_a, \gamma_a \in \mathbb{R}$ are the given parameters of the model and $\mathcal{I}$ denotes the indicator function. For an action $a \in \mathcal{A}$ of the social sensor, $\delta_a p$ indicates the effective incentive received by the social sensor; γ_a denotes the cost of taking the action; and α_a denotes the mis-representation or distortion weight [97].

Tie-breaking rule: When $r_a' \eta_{y_k} = r_{\bar{a}}' \eta_{y_k}, \forall \bar{a} \in \mathcal{A}/\{a\}$, $a_k \sim$ Uniform($\mathcal{A}$), i.e, an action from the set $\mathcal{A}$ is chosen with probability $\frac{1}{|\mathcal{A}|}$, where $|\mathcal{A}|$ denotes the cardinality of set $\mathcal{A}$. The uniform sampling tie-breaking rule ensures that the public belief (3.20) is still a martingale. This is required in the proof of Theorem 3.9.

Public Belief Dynamics: The fusion center shares sensor k's decision with the social sensors and the public belief (3.20) is updated (by the fusion center and subsequent sensors) according to the social learning Bayesian filter (see [75, 76]) as follows:

$$\pi_k = T^{\pi}(\pi_{k-1}, a_k) = \frac{R_{a_k}^{\pi_{k-1}} \pi_{k-1}}{\mathbf{1}' R_{a_k}^{\pi_{k-1}} \pi_{k-1}}. \tag{3.25}$$

Here, $R_{a_k}^{\pi_{k-1}} = \text{diag}(\mathbb{P}(a_k | x = i, \pi_{k-1}), i \in \mathcal{X})$ is the decision or action likelihood matrix (compare with the observation likelihood matrix B in (3.21)), where

$$R_{ia}^{\pi} = \mathbb{P}(a_k | x = i, \pi_{k-1}) = \sum_{y \in \mathcal{Y}} \mathbb{P}(a_k | y, \pi_{k-1}) \mathbb{P}(y | x = i),\ \ \mathbb{P}(a_k | y, \pi_{k-1}) = \begin{cases} 1 & \text{if } a_k = \arg\max_{a \in \mathcal{A}} r_a' \eta_{y_k}; \\ 0 & \text{otherwise.} \end{cases} \tag{3.26}$$

[13] Each sensor being an expected (and myopic) reward maximizer is rational [33]. This assumption implies that the social sensors have no altruistic concerns.

Note that $\pi_k \in \Pi(2)$.

Remark (**Information Cascade).** Note that the (decision) likelihood probability (3.26) is an explicit function of the prior (public belief) π_{k-1}. This is unlike a standard Bayesian update (like (3.22)), where the likelihood is independent of the prior. This unusual update of the social learning filter leads to herding behavior: In (3.26), if the action becomes independent of the observation, $R^{\pi}_{ia} = 1$ or 0. This in turn leads to information cascade, social learning stops as the public belief is frozen, as can be seen from (3.25). It can be shown that (Theorem 5.3.1, [75]) social learning stops in finite time.

Fusion Center Dynamics:

1.) *Information Fusion cost:* The fusion center minimizes the following cost of information fusion $c(p_k)$, with

$$c(p_k) = p_k - \Phi_s(k)\mathcal{I}(a_k = y_k|\pi_{k-1}). \tag{3.27}$$

Here $\mathcal{I}$ denotes the indicator function. The cost function should model the trade-off between incentives and truthful information disclosure[14]. One possible[15] cost function is (3.27). The information from different sensors is allowed to be weighed differently using $\Phi_s(k) \in (0, 1)$. Here the subscript s is used to denote the cost when only social learning is considered (see Sec.3.3.7 for the case when entropy cost, in addition to the effect of social learning, is considered). For simplicity, we assume the weights to be same for all sensors; i.e $\Phi_s(k) = \phi_s, \ \forall \ k$.

2.) *Information Fusion Incentive:* The fusion center incentivizes/compensates the social sensors for providing information about the underlying state. The fusion

[14]Acting according to self valuations ($a = y$) is in line with truthful information reporting in Peer Prediction literature; see [90]. We show in Sec.3.3.3 that $a = y$ corresponds to informative decisions. Here informativeness is in the sense of Blackwell [75]; see also Footnote 14.

[15]In Sec.3.3.7, we consider the information fusion cost that additionally has entropy of the state estimate.

center dynamically adapts these incentives over time as the sensors perform social learning: each sensor will have a different state estimate. Let $\mathcal{F}_k$ denote the history of past incentives and decisions $\{\pi_0, p_1, a_1, \cdots, p_{k-1}, a_k\}$ recorded by the fusion center and the social sensors. More technically,

$$\mathcal{F}_k := \sigma - \text{algebra generated by } (\pi_0, a_1, \ldots, a_k, p_1, \ldots, p_{k-1}). \tag{3.28}$$

The fusion center chooses the incentive as $p_{k+1} \in \mu_k(\mathcal{F}_k)$ for the sensor $k+1$ to provide information about its state via social learning. Here μ_k denotes a policy that associates the history $\mathcal{F}_k$ with an incentive p_{k+1}. Since $\mathcal{F}_k$ is increasing with time k (filtration), to implement a controller, it is useful to obtain a sufficient statistic that does not grow in dimension. The public belief π_k computed via the social learning filter (3.25) forms a sufficient statistic[16] for $\mathcal{F}_k$ and the incentive offered to social sensor $k+1$ is given as

$$p_{k+1} = \mu_k(\pi_k) \ \in [0, 1]. \tag{3.29}$$

The incentive is normalized to $[0, 1]$ without loss of generality.

B: Controlled Information Fusion Objective

Given the setup in Sec.3.3.1, the aim of the fusion center is to estimate the state $x_0(= x)$ by minimizing the cost of information acquisition (p). As discussed in (3.25), the fusion center performs Bayesian fusion of the information revealed by the social sensors.

Let $\bar{\boldsymbol{\mu}} = (\mu_0, \mu_1, \cdots)$ denote the sequence of policies employed by the fusion center at times $k = 0, 1, \cdots$. For each initial distribution π_0, the following cost is associated for the fusion center:

$$J_{\bar{\boldsymbol{\mu}}}(\pi) = \mathbb{E}_{\bar{\boldsymbol{\mu}}}\{\sum_{k=0}^{\infty} \rho^k c_{\mu_k}(p_k) | \pi_0 = \pi\}. \tag{3.30}$$

[16] See Sec.3.3.5 for justification.

Here p_k denotes the incentive, $\rho \in [0, 1)$ denotes an economic discount factor, μ_k denotes the decision policy (3.29) for the fusion center that maps the public belief π_k to an incentive $p_{k+1} \in [0, 1]$, $c_{\mu_k}(p_k)$ denotes the cost of information fusion incurred at time k, and $\mathbb{E}_{\bar{\boldsymbol{\mu}}}$ denotes the expectation conditioned on the policy sequence $\bar{\boldsymbol{\mu}}$.

The policy sequence $\bar{\boldsymbol{\mu}}$ can be restricted to the class of stationary (time invariant) policies $\boldsymbol{\mu} = (\mu, \mu, \cdots)$ for the infinite horizon discounted cost objective; see [75]. The fusion center aims to find the optimal stationary policy μ^* such that

$$J_{\mu^*}(\pi_0) = \inf_{\mu \in \boldsymbol{\mu}} J_\mu(\pi_0) \tag{3.31}$$

where $\boldsymbol{\mu}$ denotes the class of stationary policies.

Summary:
(3.25) are the dynamics and (3.30) is the optimization objective for the controlled information fusion problem considered in this work. The model parameters are the sensors' observation matrix B in (3.21) and the reward r_a in (3.24).

C: Stochastic Dynamic Programming Formulation

The optimal incentive policy μ^* in (3.31) and the corresponding optimal cost (value function) $V(\pi)$ satisfy the Bellman's stochastic dynamic programming equation [75]:

$$\begin{aligned} Q(\pi, p) &= c(p) + \rho \sum_{a \in \mathcal{A}} V(T^\pi(\pi, a))\sigma(\pi, a), \\ V(\pi) &= \min_{p \in [0,1]} Q(\pi, p), \quad J_{\mu^*}(\pi_0) = V(\pi_0), \quad \text{and} \quad \mu^*(\pi) = \arg\min_{p \in [0,1]} Q(\pi, p). \end{aligned} \tag{3.32}$$

where $T^\pi(\pi, a)$ is defined in (3.25) and $\sigma(\pi, a) = \mathbf{1}' R_a^\pi \pi$, and $c(p)$ is the information fusion cost defined in (3.27).

Even though Bellman's equation (3.32) specifies the optimal policy, it has two problems: (i) The state (belief) space $\Pi(2)$ is an uncountable set. Hence the dynamic programming

equation (3.32) does not translate into practical solution methodologies, as the optimal cost $V(\pi)$ needs to be evaluated at each $\pi \in \Pi(2)$.

(ii) The action (incentive) space for the information fusion center $p \in [0,1]$ is a continuum. It is well known [75] that even for a finite action case, computing the optimal policies is a computationally intractable PSPACE hard problem.

Discussion of model

1. *Social Sensor's Reward Function:* The nature of the results, specifically, the structural results (Theorem 1 and Theorem 3 in Sec.3.2.2); characterization of optimal incentive sequence (Theorem 2 in Sec.3.2.2); and the uniform bounds (Theorem 5 and Theorem 6 in Sec.3.3.3); is unaffected by the choice of the form of reward functions below.

   a.) (Resolution dependent reward): This form of reward function can be used to explicitly capture the effect or the influence of the observation distribution (resolution) matrix B of the social sensors on the actions. Let $r(x,y,a)$ denote the reward accrued if the sensor takes action a when the underlying state is x and the observation is y. The reward function is given as:

   $$r(x,a) = \sum_{y} r(x,y,a)B_{xy}, \text{ where } r(x,y,a) = \delta_a p - \alpha_a \mathcal{I}(a \neq x) - \beta_a \mathcal{I}(a \neq y) - \gamma_a. \tag{3.33}$$

   Here $\delta_a \in [0,1]$, $\alpha_a, \beta_a, \gamma_a \in \mathbb{R}$ are the given parameters of the model and $\mathcal{I}$ denotes the indicator function. For an action $a \in \mathcal{A}$ of the social sensor, $\delta_a p$ the effective incentive received (see discussion below) by the social sensor; γ_a denotes the cost of taking the action; α_a and β_a denote the misrepresentation or distortion weights.

b.) (Resolution independent reward): This form of the reward function is not explicitly dependent on the resolution of the social sensors, i.e,

$$r(x, a) = \delta_a p - \alpha_a \mathcal{I}(a \neq x) - \gamma_a. \tag{3.34}$$

c.) (Realization dependent reward): This form of reward function explicitly depends on the private observation or realization y_k for the social sensor k, i.e,

$$r(x, y_k, a) = \delta_a p - \alpha_a \mathcal{I}(a \neq x) - \beta_a \mathcal{I}(a \neq y_k) - \gamma_a. \tag{3.35}$$

d.) (General state-action reward): This form of reward function models a general state-action reward function, i.e,

$$r(x, a) = \delta_a p + \Gamma^y_{xa} \tag{3.36}$$

The parameter Γ^y_{xa} is any function of the resolution, realization, state, and action.

Motivation: The social sensor k receives a noisy observation y_k of the state x. The term $\beta_a I(a \neq y_k)$ models the distortion cost [97] induced by the sensor's realization in equation (3.35). For social sensor k, $I(a_k \neq y_k)$ is the binary distance function [97] of the distortion or mis-representation of the received information y_k as a_k. In case of (3.33), the term $\beta_a I(a \neq y)$ captures the inherent distortion that can result from the sensor's observation matrix B.

The information fusion center offers a single incentive $p_k \in [0, 1]$ to the social sensor k by using the information from the actions of the previous social sensors contained in the public belief π_{k-1} (see (3.29)). The weight δ_a helps to model asymmetric incentives for the different actions of the social sensor, and determines the effective incentive received by the social sensor for choosing different actions. The asymmetry is required to derive a feedback (public belief dependent) policy for the information fusion center to choose the future price. Symmetry ($\delta_{a=2}$ =

$\delta_{a=1}$) results in open loop or static prices (as the dependency cancels out) for the information fusion center. Since we are interested in dynamically changing the incentives to incorporate learning from the previous social sensors, we choose $\delta_{a=2} \neq \delta_{a=1}$.

2. *Information Fusion Cost:* The cost function for the fusion center is motivated by the revenue maximization problem with social learning literature [23, 99, 33, 24]:

$$\sum_{k=0}^{\infty} \rho^k (p_k - c) \mathcal{I}(a_k = \text{buy}). \tag{3.37}$$

Here (3.37) is the objective function of a monopoly that dynamically charges a price p_k for a product that costs c to manufacture, to a social sensor k that learns about the underlying value (state) of the product from the decisions of other social sensors. The monopoly's objective is to maximize the revenue collected. The price p_k is selected (using the optimal pricing policy) so as to influence or elicit the desired behavior (buy or not buy) from the social sensors.

A modification of (3.37) motivated by controlled information fusion applications in the presence of social learning is given by (3.27) and (3.30). Here p_k is the incentive offered by the fusion center and $\Phi_s(k) \in (0, 1)$ is the weight attached to the usefulness of the information acquired from sensor k. The objective of the information fusion is to maximize the number of sensors that act according to their observations, and estimate the underlying state. Since the sensors take into account the actions or decisions of the preceding sensors, fusion of informative decisions leads to improved estimate of the parameter, and hence improves the usefulness of information (in terms of reduction in the uncertainty of the Bayesian state estimate) fused by the fusion center and the successive sensors.

3.3.2 Main Results

We wish to determine conditions under which the optimal incentive policy has the following intuitive threshold structure: don't incentivize if the estimate $\pi < \pi^*$, and incentivize using an exactly specified incentive function otherwise. Some of the advantages of the threshold policy are: (i) To compute the threshold policy, one only needs to compute the single belief π^*; whereas a general policy requires PSPACE hard dynamic programming recursion offline. (ii) To implement a controller with a threshold policy, one only needs to encode π^* and the incentive function, so its practically useful.

Incentive Function: For future reference, we define the incentive function of the fusion center $\Delta(\eta_y) \in [0, 1]$ as

$$\Delta(\eta_y) = [l_1 \ -l_2] \frac{B_y \pi}{\mathbf{1}' B_y \pi} + l_3 \tag{3.38}$$

where η_y is the private belief update (3.22) with π_k replaced by π,

$$l_1 = \frac{\alpha_2}{\delta_2 - \delta_1}, \ l_2 = \frac{\alpha_1}{\delta_2 - \delta_1}, \ l_3 = \frac{\gamma_2 - \gamma_1}{\delta_2 - \delta_1}.$$

The incentive function (3.38) naturally arises[17] by reformulating (3.23). A set of parameters in the incentive function that ensure $\Delta(\eta_y) \in [0, 1]$ are $l_1 > 0$, $l_2 > 0$ and $l_3 > 0$. A *sufficient condition* is that $\alpha_1 > \alpha_2$, $\delta_2 > \delta_1$ and $\gamma_2 > \gamma_1$. For other forms of reward functions, the expression for $\Delta(\eta_y) \in [0, 1]$ and the conditions on the model parameters are suitably derived.

Model Assumptions: We now give sufficient conditions under which the optimal incentive policy (3.32) has a threshold structure.

(A1) The observation distribution $B_{xy} = \mathbb{P}(y|x)$ is TP2 (totally positive of order 2), i.e, the determinant of the matrix B is non-negative.

(A2) The reward vector r_a is supermodular, i.e, $r(1, 1) > r(2, 1)$ and $r(2, 2) > r(1, 2)$ for every $p \in [0, 1]$.

[17] See also the proof of Theorem 3.10.

(A1) is an assumption on the underlying stochastic model, and enables the comparison of the posteriors. The observation distribution being TP2 [75] implies that in higher states, the probability of receiving higher observations is higher than in lower states.
(A2) is required for the problem to be non-trivial. If it does not hold and $r(i,1) > r(i,2)$ for $i = 1,2$, then $a = 1$ always dominates $a = 2$; the sensors provide no useful information. (see Sec.3.3.8 for assumptions in non-binary environments) Theorem 3.8 below is our first main result. It provides a closed form expression for the optimal policy $\mu^*(\pi)$ of the controlled information fusion problem: the optimal policy has threshold structure. The choice over a continuum of actions is reduced to a choice between two exactly specified incentive policies. The optimal policy is not unique[18]. There exists a version of the optimal policy having the structure as in Theorem 3.8.

Theorem 3.8. Under (A1) and (A2), the optimal incentive policy defined in (3.31) is given explicitly as:

$$\mu^*(\pi) = \begin{cases} 0 & \text{if } \pi(2) \in [0, \pi_s^*(2)); \\ \Delta(\eta_{y=2}) & \text{if } \pi(2) \in [\pi_s^*(2), 1]. \end{cases} \tag{3.39}$$

Here the threshold state $\pi_s^*(2) \in (0,1)$ depends on the choice of $\phi_s \in (0,1)$ defined in (3.27), and the parameters in the incentive function $\Delta(\eta_{y=2})$ defined in (3.38).

The proof involves establishing the following: (i) show that due to the structure of the social learning filter in (3.25), the choice of incentives reduces from a continuum $[0,1]$ to a finite number at every belief; (ii) show that the incentive function $\Delta(\eta_y)$ is decreasing in π, hence the value function is monotone, and there exists a threshold π^*. According to Theorem 3.8, computing the optimal incentive policy is equivalent to finding the belief $\pi_s^*(2)$, below which it is optimal not to provide any incentive $p = 0$; and above which it is optimal to incentivize using $\Delta(\eta_{y=2})$ at every belief, to minimize the cost. Therefore,

[18] See the proof of Theorem 3.8.

the controlled information fusion problem reduces to a finite dimensional optimization problem of finding a threshold state π^*. Theorem 3.8 provides a closed form expression for the optimal policy of the controlled information fusion problem: the choice over a continuum of actions is reduced to a choice between two exactly specified policies: $\mu(\pi) = 0, \ \forall \ \pi$ and $\mu(\pi) = \Delta(\eta_{y=2}), \ \forall \ \pi$.

The practical usefulness of Theorem 3.8 stems from the following: (i) the search space of decision policies μ reduces from an infinite class of functions (over $\Pi(2)$) to those that switch once between the specified policies; (ii) at each instant (or belief) the fusion center only needs to decide between $p = \Delta(\eta_{y=2})$ and $p = 0$; (iii) the region in the belief space $\Pi(2)$ where it is optimal to incentivize using $\Delta(\eta_{y=2})$ is connected and convex.

Theorem 3.8 characterized the structure of the optimal incentive policy for controlled information fusion. A natural question is: How does the actual sample path of the optimal incentive sequence behave? Theorem 3.9 below gives a sample path characterization of optimal incentive policy implemented by the fusion center. It is shown that when the fusion center aims to minimize the expected payout for gathering truthful information to reduce the uncertainty in the Bayesian state estimate, the incentive sequence is a sub-martingale[19]; i.e, it increases on average[20] over time.

Theorem 3.9. Consider the information fusion problem with optimal policy $\mu^*(\pi)$ in (3.39). Under (A1), the optimal incentive sequence $p_k = \mu^*(\pi_{k-1})$ is a sub-martingale.

The proof involves establishing the following: (i) the incentive function $\Delta(\eta_{y=2})$ is convex in π; (ii) the optimal incentive policy in (3.39) is such that the incentives are increasing on average over time, using the notion of predictable sequences [48]. Typically in stochastic control problems, it is difficult to characterize the optimal control sequence;

[19] See Appendix for definition.

[20] Here average is over different iterations of the estimation process. For example, each round of labelling/classification in Crowdsourcing can be seen as one iteration.

one can only characterize the optimal control policy. Theorem 3.9 is interesting because we can characterize the optimal sequence of incentives as a sub-martingale. According to Theorem 3.9, the optimal incentive policy of the fusion center is such that the sample path of the incentive sequence displays an increasing trend, i.e, the incentives increase on average over time.

The usefulness of Theorem 3.9 stems from the following: (i) it gives a sample path characterization of the optimal incentive policy implemented by the fusion center; (ii) the sub-martingale property assures that the average incentives should always increase over time. This is useful in assessing the reliability of the fusion center.

The increase in incentives over time can be attributed to the fact that the senors polled for information at a later instant have more accurate estimate of the state due to learning from predecessors, and hence require higher compensation to reveal the same.

3.3.3 Consistency of Controlled Information Fusion

An elementary application of the martingale convergence theorem [48] shows that the social learning protocol (3.25) results in social sensors forming an information cascade; that is, after some time n^*, all sensors choose the same action and social learning stops (see Theorem 5.3.1, [75]). Therefore, the true state can never be estimated using social learning, indeed, the belief will not converge to the true state asymptotically.

In this section, we show that by dynamically controlling the incentives over time, the fusion center can indeed learn the true state. However, this comes at the price of employing a sub-optimal incentive policy. We further provide uniform bounds on the additional cost incurred for consistency[21]. When it is sufficient to know the state with a

[21]Let the true state be $x = \theta$. The pair (θ, π_k) is consistent, if π_k converges to a point mass at θ in probability.

degree of confidence, policies that guarantee state estimation in finite time are discussed. We also provide uniform bounds on the budget saved as a result of estimating the state only upto a degree of confidence.

Controlled Information Fusion

Fig.3.5 shows the bi-directional interaction between the fusion center and the social sensor. The incentives chosen by the fusion center affects the reward function of the social sensors, and hence affects the decisions chosen. The decisions chosen in turn affect the estimate of the state (3.20) for the fusion center as in (3.25). Recall that social learning terminates after a finite horizon (see remark on Information cascade after (3.26)). Theorem 3.10 below shows how to control the incentives to the social sensors to delay herding and information cascades, and hence estimate the state asymptotically. In particular, it is shown how the fusion center can *control the incentives* such that the fusion of Bayesian estimates is consistent.

We will express[22] the belief space $\Pi(2)$ as a disjoint union of three connected regions to describe the sensors' decision dynamics as a function of the incentive p: a region $\mathcal{P}_1^p$ - where action $a = 2$ is optimal; a region $\mathcal{P}_3^p$ - where action $a = 1$ is optimal; a region $\mathcal{P}_2^p$ - where action $a = y$ is optimal. From (3.23), the decision of the social sensor depends on the private belief η_y and the reward r_a (defined in (3.24)). Therefore, define:

$$\begin{aligned}
\mathcal{P}_1^p &= \{\pi \in \Pi(2) : (r_1 - r_2)'\eta_{y=1} \leq 0\} \\
\mathcal{P}_2^p &= \{\pi \in \Pi(2) : (r_1 - r_2)'\eta_{y=1} > 0 \cap (r_1 - r_2)'\eta_{y=2} \leq 0\} \\
\mathcal{P}_3^p &= \{\pi \in \Pi(2) : (r_1 - r_2)'\eta_{y=2} > 0\}
\end{aligned} \tag{3.40}$$

where r_a for $a = \{1, 2\}$ are the social sensors' rewards and $\mathcal{P}^p$ models the explicit depen-

[22]This is possible because of (A1) and (A2); see [74].

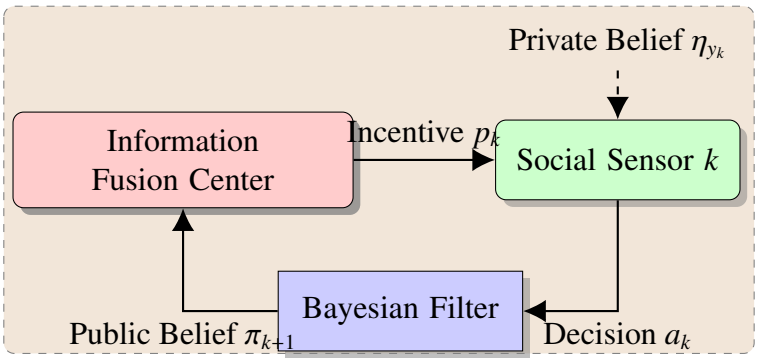

Figure 3.5: Bi-directional interaction between the information fusion center and the social sensor. The fusion center provides an incentive p_k to the social sensor, which has a private belief η_{y_k} after observation y_k. The social sensor takes a decision a_k and this quantized information on the underlying state is used to update the public belief π_{k+1} using a social learning Bayesian filter (3.25). The incentive p_k at time k directly modifies the reward function of the social sensor, and hence affects the state estimate π_{k+1} at time $k+1$.

dence of the width of the regions on the incentive parameter p through r_a, $\eta_{y=1}$ and $\eta_{y=2}$ denote the private belief updates after $y=1$ and $y=2$ respectively. The region $\mathcal{P}_1^p \cup \mathcal{P}_3^p$ is the *herding* region and $\mathcal{P}_2^p$ is the *social learning* region for any $p \in [0,1]$.

Theorem 3.10. Under (A1) and (A2), the following relation holds between the incentive p_k and the public belief π_{k+1}:

$$\pi_{k+1} \in \begin{cases} \mathcal{P}_3^p & \text{iff } p_k \in [0, \Delta(\eta_{y_k=2})); \\ \mathcal{P}_2^p & \text{iff } p_k \in [\Delta(\eta_{y_k=2}), \Delta(\eta_{y_k=1})); \\ \mathcal{P}_1^p & \text{iff } p_k \in [\Delta(\eta_{y_k=1}), 1]. \end{cases}$$

where the regions $\mathcal{P}_i^p$ for $i=1,2,3$ are defined in (3.40), and $\Delta(\eta_y)$ is as in (3.38).

According to Theorem 3.10, relation between the incentive p_k at time k and the state estimate (public belief π_k) at the next instant $k+1$ is such that, when p_k belongs to the

intervals defined by the private beliefs (in the incentive function $\Delta(\eta_{y_k})$), the widths of the herding and social learning regions change (see Fig.3.6) so that the public belief (π_{k+1}) belongs to the desired $\mathcal{P}_i^p$. Fig.3.6 shows the variation of the width of the regions with respect to the incentive parameter p. Theorem 3.10 characterizes the sensitivity of the regions $\mathcal{P}_1^p, \mathcal{P}_2^p, \mathcal{P}_3^p$ with respect to the incentive $p \in [0, 1]$, and Corollary 3.3.1 below shows how to stop the information cascade so that social learning can proceed indefinitely so that the state estimate converges to the true state.

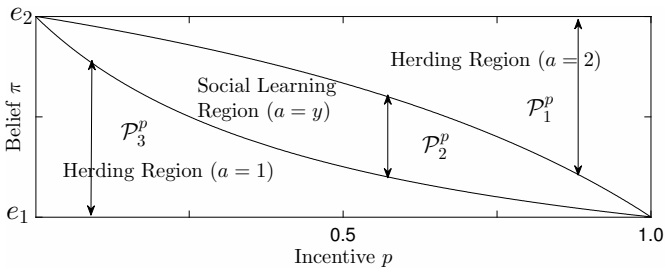

Figure 3.6: Herding ($\mathcal{P}_1^p \cup \mathcal{P}_3^p$) and social learning ($\mathcal{P}_2^p$) regions with respect to the incentive parameter p. It is seen that when the incentives are small (close to 0), the sensors herd on low quality actions ($a = 1$); and when the incentives are high (close to 1), the sensors herd on high quality actions ($a = 2$); however, only the actions in the social learning region are informative or reflect the sensors' true valuation.

Corollary 3.3.1. Let $p_k = \Delta(\eta_{y_k=2})$ for $k = 1, 2, \ldots$. The fusion of Bayesian estimates is consistent, i.e, the fusion center learns the true state asymptotically.

We know that the fusion center can force the state estimates to be in the social learning region by choosing incentives in the range $p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1}))$, see Fig.3.6. From (3.40), Lemma 3.4 and Theorem 3.5 in the Appendix, the social sensors' decision likelihood matrices R_a^π (as in (3.25)) in regions $\mathcal{P}_1^p, \mathcal{P}_2^p$, and $\mathcal{P}_3^p$ for any $p \in [0, 1]$ are

$$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \text{ and } \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

respectively. In the herding region $\mathcal{P}_1^p \cup \mathcal{P}_3^p$, the decision of the social sensor is independent of the public belief and the public belief (3.25) is frozen. In the social learning region $\mathcal{P}_2^p$, the sensors take informative[23] actions; i.e, each sensor acts according to its observation/valuation. In the social learning region, sensors take informative actions $a = y$; or $R_a^\pi = B$. The observations are conditionally independent given the true state. Therefore, by suitably controlling the incentives, the fusion center fuses information that is i.i.d on the true state. It is well known [45, 128] that fusion of Bayesian estimates is consistent (convergence in probability); i.e, for a point mass at the true state θ denoted as $g(\theta)$,

$$\lim_{k\to\infty} \mathbb{P}(|\pi_k - g(\theta)| > \epsilon) = 0 \; \forall \; \epsilon > 0.$$

In other words, the fusion center *can* learn the true state asymptotically by choosing the incentives as $p_k = \Delta(\eta_{y_k=2})$ for $k = 1, 2, \ldots$.

Cost of consistency for the fusion center

When the incentive policy is the optimal threshold policy (3.39), the fusion of Bayesian estimates computed from the social sensors' decisions (3.25) is not consistent. This is because, the optimal incentive policy for the fusion center is such that below a certain threshold it is optimal to not incentivize. From Theorem 3.10, when the fusion center stops incentivizing $p = \mu^*(\pi) = 0$, the public belief is in the herding region $\mathcal{P}_3^p$. In the herding region, social learning ceases and there is no improvement in uncertainty – mean square error between the state estimate and the true parameter remains at a fixed non-zero value. If, however, the fusion center chooses a sub-optimal policy (3.42), it

[23] Informativeness is in the sense of Blackwell; see [75]. For any two observation matrices B_1 and B_2, B_1 is more informative than B_2 in the Blackwell sense ($B_1 \succ_B B_2$) if $B_2 = B_1\Gamma$, for any stochastic matrix Γ. When the sensors act according to their observations, $\pi \in \mathcal{P}_2^p$, and the decision likelihood matrix in (3.25) $R_S^\pi = B$; and when the sensors don't act according to the observations (they herd), $\pi \in \mathcal{P}_3^p$, the decision likelihood matrix $R_H^\pi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. We have for $\Gamma = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, $R_H^\pi = R_S^\pi \Gamma \Rightarrow R_S^\pi \succ_B R_H^\pi$.

will incur additional cost for the incentives; but the fusion of estimates computed from the social sensors' decisions (3.25) will be consistent (Corollary 3.3.1). Theorem 3.11 below provides uniform bounds on the additional cost incurred by the fusion center for employing a sub-optimal incentive policy that results in consistent information fusion. Consider the objective function for the fusion center:

$$W_{\mu_c}(\pi) = \mathbb{E}_{\mu_c}\{\sum_{k=0}^{\infty} \rho^k c_{\mu_c}(p_k)|\pi_0 = \pi\} \tag{3.41}$$

where $W_{\mu_c}(\pi)$ denotes the cost incurred by employing the sub-optimal policy (compare with (3.39))

$$\mu_c(\pi) = \left\{\Delta(\eta_{y=2}) \; \forall \; \pi(2) \in [0,1]\right\}. \tag{3.42}$$

Theorem 3.11. Let (A1) hold. The additional cost (on average) incurred by the fusion center for employing the sub-optimal policy $\mu_c(\pi)$ in (3.42) instead of the optimal policy $\mu^*(\pi)$ in (3.39) is bounded as:

$$\sup_{\pi} |W_{\mu_c}(\pi) - J_{\mu^*}(\pi)| \leq 2\frac{(1-\phi_s)}{1-\rho} \tag{3.43}$$

where $J_{\mu^*}(\pi)$ is the optimal cost (3.31).

The proof uses the fact that both $W_{\mu_c}(\pi)$ and $J_{\mu^*}(\pi)$ are decreasing in π established using similar arguments as in Theorem 3.8, and $W_{\mu_c}(\pi) - J_{\mu^*}(\pi) \geq 0 \; \forall \; \pi$. Theorem 3.11 characterizes the trade-off between consistency and cost of information acquisition. It says that when the fusion center employs a sub-optimal policy, the average additional cost incurred is bounded above by the weight ϕ_s in the information fusion cost (3.27), discount factor ρ that captures the degree of impatience of the fusion center.

The usefulness of Theorem 3.11 stems from the following: (i) It gives an upper bound on the additional discounted cost incurred when the fusion center chooses the incentives such that the fusion of Bayesian estimates computed as in (3.25) is consistent. (ii) It helps in choosing the weight ϕ_s and the discount factor ρ for the fusion center.

3.3.4 Finite time bounds for the fusion center

In Sec.3.3.3, it was shown that by employing a sub-optimal policy the fusion center can estimate the true state asymptotically. However, it is often enough to know the state with a degree of confidence. In this section, we obtain uniform bounds on the budget saved by estimating the state upto a degree of confidence.

The degree of confidence characterizes regions in the belief space $\Pi(2)$ that can be used to estimate the states. For a degree of confidence $\vartheta \in (0,1)$, any belief in the confidence region $\pi(2) \in [0,\vartheta]$ is identified with state $x = 1$, and any belief in the confidence region $\pi(2) \in [1-\vartheta, 1]$ is identified with state $x = 2$. For example, when the public belief (posterior) is such that $\pi(2) \in [0.9, 1]$, then the fusion center is (atleast) 90% confident that the state $x = 2$, and if $\vartheta < 0.1$, the state is estimated as $x = 2$. For a degree of confidence $\vartheta \in (0,1)$, consider using the following policy

$$\mu_f(\pi) = \begin{cases} 0 & \text{if } \pi(2) \in [0,\vartheta]; \\ \Delta(\eta_{y_k}) & \text{if } \pi(2) \in (\vartheta, 1-\vartheta); \\ 0 & \text{if } \pi(2) \in [1-\vartheta, 1]. \end{cases} \tag{3.44}$$

It can be shown using martingale convergence theorem [48] that when using the policy (3.44), the public belief hits one of the two confidence regions in finite time[24].

The following theorem provides a bound on the budget saved by employing the policy in (3.44) instead of the policy (3.42). Let $\pi_\vartheta = \begin{bmatrix} \vartheta \\ 1-\vartheta \end{bmatrix}$ and $\eta_\vartheta = \frac{B_{y=2}\pi_\vartheta}{\mathbf{1}' B_{y=2}\pi_\vartheta}$.

Theorem 3.12. Let (A1) hold. For a degree of confidence ϑ, the budget saved by the fusion center by employing the policy $\mu_f(\pi)$ in (3.44) instead of the policy $\mu_c(\pi)$ in (3.42) is bounded as:

$$\sup_{\pi} |W_{\mu_c}(\pi) - W_{\mu_f}(\pi)| \leq 2\frac{(1-\phi_s)}{1-\rho} + \frac{|\Delta(\eta_\vartheta) - \phi_s|}{1-\rho}. \tag{3.45}$$

[24]The arguments are similar to those used to establish information cascades occur in finite time in [19] and Theorem 5.3.1, [75].

where ρ is the discount factor.

The proof follows using arguments similar to Theorem 5 in this work. Theorem 3.12 provides an uniform bound on the budget saved by employing the policy $\mu_f(\pi)$ in (3.44) instead of $\mu_c(\pi)$ in (3.42). A bound on the budget saved with respect to the optimal policy $\mu^*(\pi)$ can be obtained from Theorem 3.12 and Theorem 3.11 using the triangle inequality of the norm,

$$|J_{\mu^*}(\pi) - W_{\mu_f}(\pi)| \leq |W_{\mu_c}(\pi) - J_{\mu^*}(\pi)| + |W_{\mu_c}(\pi) - W_{\mu_f}(\pi)|.$$

In Theorem 3.12, the fact (Lemma 3.15) that $\Delta(\eta_{y=2})$ is decreasing in π, and $|\varepsilon| \geq \varepsilon$ is utilized in deriving the bounds.

Summary: Theorem 3.10 (together with Corollary 3.3.1) showed how the fusion center can employ a sub-optimal incentive policy such that information fusion with social sensors is consistent, and Theorem 3.11 gave an uniform bound on the average additional cost incurred for employing the sub-optimal policy $\mu_c(\pi)$ in (3.42). Theorem 3.12 provided an uniform bound on the budget saved by employing a finite time state estimation policy $\mu_f(\pi)$ in (3.44) instead of the optimal policy $\mu^*(\pi)$ in (3.39).

3.3.5 Strategic behaviour in Social Sensors

The information fusion center polls the social sensors in a pre-determined order and they decide what information to reveal, i.e, it was assumed that the sensors do not hide their signals and are not strategic. However, the rewards can be suitably designed so that the sensors reveal information when polled. In this section, we show how to design the reward functions to prevent the social sensors from being strategic. This implies that the social sensors have no forward-looking tendencies and reward function of the social

sensors has no externalities, and the public belief (3.20) forms a sufficient statistic for the history of past actions and incentives.

Under an additional minor restriction[25] on the reward parameters, it is shown below that the social sensors have no incentive to delay or hide their signals.

Social sensors do not display contrarian behavior:
The optimal policy for the fusion center dictates that it either incentivize or not incentivize, see Theorem 3.8. When the fusion center is offering incentives ($\Delta(\eta_{y=2})$), from Theorem 3.10, it is seen that it is optimal for the social sensors to act according to their observations. As the social sensors are assumed to be Bayes rational, they have no incentive to deviate. When the fusion center is not incentivizing, the sensors always herd.

Social sensors are not strategic: Let $\mathcal{R}_H$ and $\mathcal{R}_S$ denote the regions where the fusion center does not incentivize ($\mu(\pi) = 0$) and incentivizes ($\mu(\pi) = \Delta(\eta_{y=2})$) respectively. A social sensor deciding at time k considers the following scenarios:

a.) $\pi_k \in \mathcal{R}_S$ and $\pi_{k+1} \in \mathcal{R}_H$. In other words if the sensor delays revealing information and the belief update after the next $(k+1)$ sensors' decision belongs to the region where there is no incentivization.

$p_{k+1} = 0$, so the social sensor k would be better off revealing at time k.

b.) $\pi_k,\ \pi_{k+1} \in \mathcal{R}_S$ and $\pi_{k+1}(2) < \pi_k(2)$.

Consider the rewards for the social sensor from (3.24),

$$r_1 = [\delta_1 p + \Gamma_{11} \quad \delta_1 p + \Gamma_{21}]$$
$$r_2 = [\delta_2 p + \Gamma_{12} \quad \delta_2 p + \Gamma_{22}] \tag{3.46}$$

[25]This is independent of the actual form of the rewards when the rewards in both states are non-zero.

Assume $\Gamma_{ij} > 0$ for all i, j without loss of generality. Note that the reward vector r_a is also required to be super-modular for any p , so $\Gamma_{11} > \Gamma_{21}$ and $\Gamma_{22} > \Gamma_{12}$. Let $T(\pi, y_k) = \frac{B_{y_k}\pi}{\mathbf{1}' B_{y_k}\pi}$ denote the private belief of sensor k. There are two possible observations for the social sensor k, $y_k = 1, 2$. We will establish the result for $y_k = 1$, and the result follows immediately for $y_k = 2$.

Let $\bar{r}_a = [\delta_a p_{k+1} + \Gamma_{1a} \;\; \delta_a p_{k+1} + \Gamma_{2a}]$.

Proposition 3.3.1. Let the observation of sensor k be $y_k = 1$. There is a discount factor $D \in (0, 1]$ such that $r_1' T(\pi_k, y_k = 1) \geq D\, \bar{r}_1' T(\pi_{k+1}, y_k = 1)$.

Proof: From the definition of First-order stochastic dominance and TP2 on B, we have the following[26]

$$r_1' T(\pi_k, y_k = 1) \leq \bar{r}_1' T(\pi_{k+1}, y_k = 1)$$

$$\therefore r_1' T(\pi_k, y_k = 1) > D\, \bar{r}_1' T(\pi_{k+1}, y_k = 1),$$

$$\text{where } D = \frac{r_1' T(\pi_k, y_k = 1)}{\bar{r}_1' T(\pi_{k+1}, y_k = 1)} - \epsilon, \text{for } \epsilon > 0.$$

Considering the largest possible deviation[27] $\pi_k(2) = 1$ and $\pi_{k+1}(2) = 0$, it is easily seen that the smallest value for $D = \frac{\Gamma_{21}}{\delta_1 + \Gamma_{11}} - \epsilon < 1$. □

Discussion: The social sensors are not more forward looking than D from Proposition 3.3.1. By suitably choosing the reward parameters, we can obtain $D = 1$. This implies that the social sensors have no incentive to deviate when $y_k = 1$.

c.) $\pi_k,\ \pi_{k+1} \in \mathcal{R}_S$ and $\pi_{k+1}(2) > \pi_k(2)$. By using similar arguments as in Proposition 3.3.1, we obtain the discount factor $D = \frac{\Gamma_{12}}{\delta_2 + \Gamma_{22}} - \epsilon < 1$. By suitably choosing the reward parameters, we can obtain $D = 1$. This implies that the social sensors have no incentive to deviate when $y_k = 2$. Also, the result follows immediately for $y_k = 1$.

[26] Note that for any $a, b > 0$, $a < b \Rightarrow a > (\frac{a}{b} - \epsilon)b$ for any $\epsilon > 0$.

[27] Note that $\pi_k,\ \pi_{k+1} \in \mathcal{R}_S$. Clearly, this is included in $\pi_k(2), \pi_{k+1}(2) \in [0, 1]$.

It was shown that when the reward parameters are chosen so that $D = 1$, myopically maximizing the expected reward is a Markov perfect equilibrium.

3.3.6 Controlled Information Fusion with Dynamic States

So far, we considered the problem of incentivized information fusion for estimating the random variable $x \in \mathcal{X}$. In this section, we consider the information fusion to estimate the state of a Markov chain x_k for $k = 0, 1, 2, \cdots$ with social sensors. The dynamic states might correspond to, for example, a change in the product/ service quality on AirBnb or Amazon.

Let the state x_k evolve as a Markov chain on the space $\mathcal{X}$ with a transition probability matrix P and an initial distribution π_0. Below we briefly highlight the changes in the social learning model in Sec.3.3.1 for the case of dynamic states.
The private belief update in (3.22) for the social sensors taking the possible state change into account is given as

$$\eta_{y_k} = \frac{B_{y_k} P' \pi_{k-1}}{\mathbf{1}' B_{y_k} P' \pi_{k-1}} \tag{3.47}$$

The public belief update in (3.25) taking the possible state change into account is given as

$$\pi_k = T^{\pi}(\pi_{k-1}, a_k) = \frac{R_{a_k}^{\pi_{k-1}} P' \pi_{k-1}}{\mathbf{1}' R_{a_k}^{\pi_{k-1}} P' \pi_{k-1}}. \tag{3.48}$$

The optimal incentive policy in case of a random variable $\mu^*(\pi)$ in Theorem 3.8 is near optimal for the case of dynamic states, when transitions out of the current state is allowed only with a small probability. This is shown in Theorem 3.13 below.

Let $\mu^*(\pi)$ denote the optimal policy for estimating/ localizing the random variable ($P = I$); and $\mu_\epsilon^*(\pi)$ denote the optimal policy for estimating/ tracking the state of a

Markov chain with $P = \begin{bmatrix} 1-\epsilon_1 & \epsilon_1 \\ \epsilon_2 & 1-\epsilon_2 \end{bmatrix}$, where $\epsilon_1, \epsilon_2 > 0$.

Theorem 3.13. Let $\rho \in [0, 1)$ denote the economic discount factor. Let $V_{\mu^*}(\pi)$ and $V_{\mu^*_\epsilon}(\pi)$ denote the optimal costs incurred by employing the optimal policy $\mu^*(\pi)$ and $\mu^*_\epsilon(\pi)$ respectively. The following holds:

$$V_{\mu^*}(\pi) - V_{\mu^*_\epsilon}(\pi) \leq \frac{2\rho(1-\phi_s)(\epsilon_1+\epsilon_2)}{(1-\rho)^2} \times \max\{|B_{21}-B_{11}|, |B_{22}-B_{12}|\}. \tag{3.49}$$

The proof follows from Theorem 2 in [113]. Theorem 3.13 says that the policy $\mu^*(\pi)$ incurs a total cost $V_{\mu^*}(\pi)$ that is within $O(\epsilon_1+\epsilon_2)$ of the total cost $V_{\mu^*_\epsilon}(\pi)$. When $\epsilon_1, \epsilon_2 << 1$, the policy $\mu^*(\pi)$ for the state localization problem ($P = I$) is near optimal for the state tracking problem ($P \neq I$).

Characterizing the nature of the optimal incentive sequence (as in Sec.3.2.2) in case of a random variable relied on the crucial fact that the belief is a martingale unconditional on the state. However, when the states are changing, the public belief (3.48) is not a martingale (see [33]). This implies that, even though, $\mu^*(\pi)$ is near optimal, the incentive sequence that results from the fusion center employing $\mu^*(\pi)$ need not show an increasing trend on average.

3.3.7 Numerical Results

Sec.3.3.7 below illustrates a controlled information fusion with quadratic cost unlike (3.27). It is shown that a multi-threshold incentive policy is optimal for the fusion center. Sec.3.3.7 illustrates the sensitivity of the optimal threshold (3.39) to the parameters ϕ_s (the weight in (3.27)) and ρ (discount factor in the objective (3.30)) that are chosen by the fusion center. Sec.3.3.7 illustrates the relation between the information

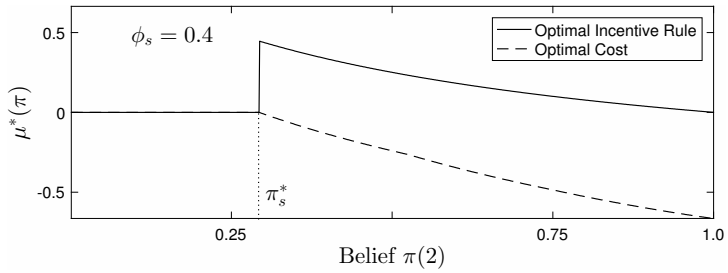

Figure 3.7: Optimal Policy for social learning weight $\phi_s = 0.4$.

gathering capabilities of the sensor (observation matrix B in (3.21)) and the average incentives provided by the fusion center. Sec.3.3.8 discusses the formulation and a numerical simulation for the controlled information fusion in non-binary environments.

Bellman's equation (3.32) is solved by discretizing the state space $\Pi(2)$. The optimal incentive policy and the optimal cost for the fusion center are computed by constructing a uniform grid of 1000 points for $\pi(2) \in [0, 1]$ and then implementing the policy and value iteration algorithm [75] for a duration of $N = 100$. Usefulness of information vs

$\alpha_1 = 0.288$	$\alpha_2 = 0.278$	$\beta_1 = 0.11$
$\beta_2 = 0.1$	$\gamma_1 = 0.1$	$\gamma_2 = 0.414$

Table 3.1: For $\delta_1 = 0.3$, $\delta_2 = 0.95$, the following parameters were obtained as a solution of $\Delta(e_1) = 1$ and $\Delta(e_2) = 0$ for the reward vector (3.33) parameters with the observation matrix B=$\begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$.

Incentivizing trade-off for the fusion center. It can be seen that $\pi^*(2)$ is decreasing with ϕ_s – a higher weight will necessitate incentivizing sooner. According to Theorem 3.11, higher ϕ_s implies that the additional cost for employing a sub-optimal policy is smaller; in other words, π_s^* is smaller. The parameters of the incentive function (3.38) are given in Table 3.1 and the discount factor $\rho = 0.4$. Here ϕ_s denotes the weight in the information fusion cost (3.27).

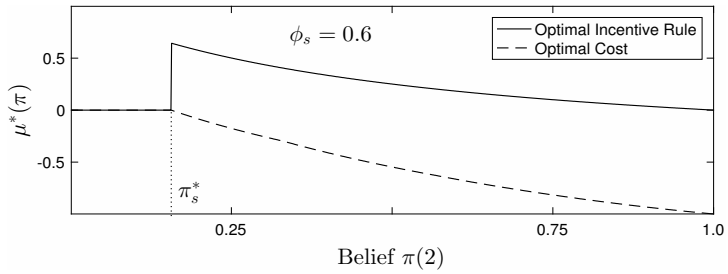

Figure 3.8: Optimal Policy for social learning weight $\phi_s = 0.6$.

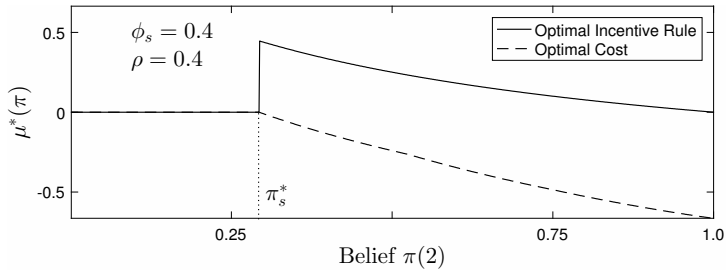

Figure 3.9: Optimal Policy for discount factor $\rho = 0.4$.

Optimal cost vs Discount factor: It is seen that a higher discount factor leads to smaller (expected) costs for higher states. This indicates that it is beneficial for the fusion center to attach more importance to future costs as as it should also take into account the benefit from sensors performing social learning. The parameters of the incentive function (3.38) are specified in Table 3.1 and the weight $\phi_s = 0.4$. Here ρ denotes the discount factor in the objective (3.30) and ϕ_s denotes the weight in the information fusion cost (3.27).

Multi-threshold Incentive Policies

This subsection illustrates numerically the nature of the optimal incentive policies for formulations of the information cost more general than (3.27), in particular we consider the *entropy* cost. We will see that the optimal incentive policy has a multi-threshold structure.

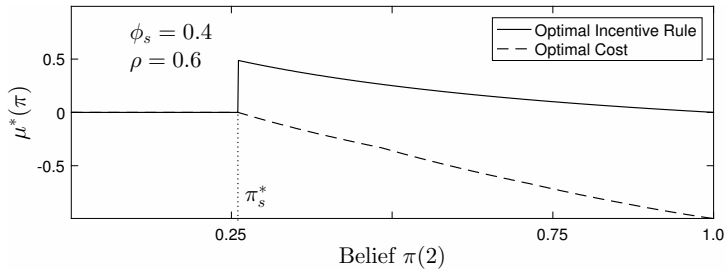

Figure 3.10: Optimal Policy for discount factor $\rho = 0.6$.

Expenditure & Entropy Cost for Information Fusion: Suppose the fusion center aims to minimize the expenditure to receive truthful accounts of the information gathered by the social sensors in addition to minimizing the entropy of the state estimate, i.e,

$$c(p) = p + \psi_e(\pi)C_e(\pi) - \phi_e \mathcal{I}(a = y|\pi) \tag{3.50}$$

where $\phi_e \in (0, 1)$ denotes the scalar weight, p denotes the expenditure, ψ_e denotes the importance of the entropy cost, and $C_e(\pi) = -\sum_{i=1}^{2} \pi(i)\log_2\pi(i)$ for $\pi(i) \in (0, 1)$ and $C_e(\pi) \stackrel{\Delta}{=} 0$ for $\pi(i) = \{0, 1\}$. Figure below shows the optimal cost and optimal policy for the fusion center when it considers entropy of the state estimate in addition to the expenditure in the information fusion cost (3.27). It can be seen that the optimal policy has a multi-threshold structure, and the optimal cost is discontinuous. A discontinuous cost implies a slight change in the initial conditions will lead to significantly different costs. Optimal policy being multi-threshold is unusual: it implies that if it is optimal to incentivize at a particular belief, it need not be optimal to do the same when the belief is larger. Multi-threshold incentive policy with the entropy cost. The regions in the belief space $\Pi(2)$ where it is optimal to not incentivize $\mu^*(\pi) = 0$ is no more connected and convex. Having a connected region in the belief space where it is optimal not to incentivize has implications on the confidence of the fusion center in implementing the incentive policy: once its optimal to incentivize at a certain belief, it need not be optimal to continue incentivizing when the belief is larger, i.e, when it is more certain about the estimate of the state. The optimal cost is discontinuous in Fig.7b, and this implies that a

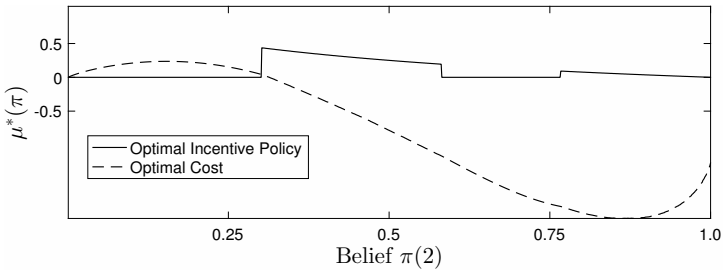

Figure 3.11: The parameters are in Table 3.1 with $\phi_e = 0.25$, discount factor $\rho = 0.8$, and $\psi_e(\pi) = 0.1 - \pi^2(2)$. Here $\psi_e(\pi)$ captures the requirement of higher weight when the belief is smaller.

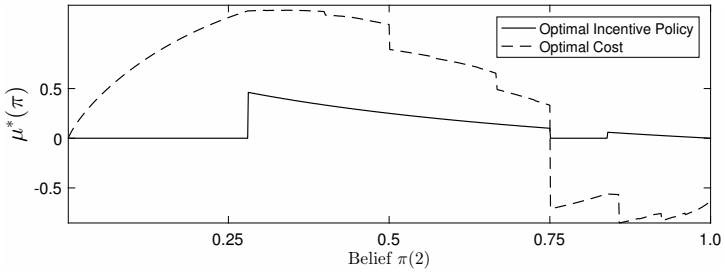

Figure 3.12: Discontinuous optimal cost. The parameters are in Table 3.1 with $\phi_e = 0.4$, discount factor $\rho = 0.6$, and $\psi_e(\pi) = 0.6 \times \mathcal{I}(\pi(2) < 0.75) - 0.35 \times \mathcal{I}(\pi(2) > 0.75)$. Here $\psi_e(\pi)$ captures the requirement of higher weight when the belief is smaller.

slight change in the initial conditions will lead to a significantly different cost.

Sensitivity of Optimal Incentive Policy

The following numerical results along with Theorem 3.11 provide a rationale for choosing the parameters: ϕ_s – the weight in the information fusion cost (3.27) and ρ – the discount factor in the fusion center's objective (3.30).

(i) *Usefulness of Information vs Incentivizing*:

We illustrate the trade-off between usefulness of information and incentivizing in the information fusion cost (3.27), and see how it affects the threshold π_s^* in (3.39). Figure below shows the affect of increasing the weight ϕ_s when the remaining parameters are

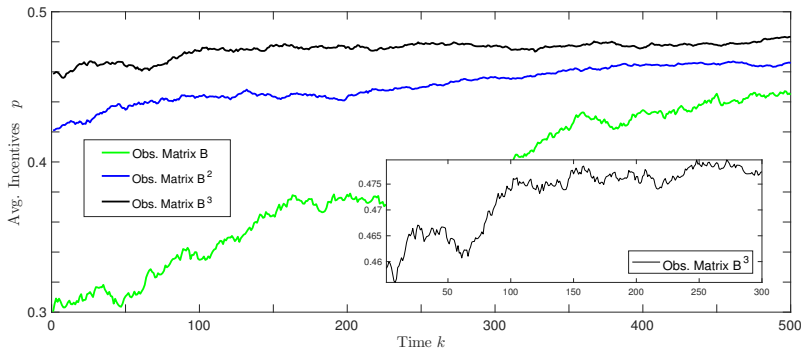

Figure 3.13: The figure shows the incentives averaged over independent sample paths for the fusion center over time for observation matrices B, B^2 and B^3. The observation matrices are ordered in the decreasing order of informativeness (see Footnote 8). The parameters are specified in Tables I & II. The weight $\phi_s = 0.4$ in the information fusion cost (3.27) and the discount factor $\rho = 0.6$. It can be seen that the range (or the slope) of the average incentives over the time horizon is highest for the case of observation matrix B. The average incentives display an increasing trend. The zoomed in subfigure shows the increasing trend in case of observation matrix B^3. It can be seen that average incentives offered in case of B^3 is higher than B^2 which in turn is higher than B.

the same. It can be seen that π_s^* is decreasing with ϕ_s. From Theorem 3.11, higher ϕ_s implies that the additional cost for employing a sub-optimal policy is smaller; in other words, $\pi_s^*(2)$ is smaller.

(ii) *Optimal cost vs Discount factor*:

We illustrate the relation between total cost incurred by the fusion center for different discount factors ρ in the objective function (3.30). The discount factor models the degree of impatience of the fusion center, as the cost incurred at time k is $\rho^k c(p_k)$. A smaller discount factor indicates that the fusion center pays more attention to the current costs than future costs. It is seen from the figure below that a higher discount factor leads to smaller (expected) costs for higher states. This indicates that it is beneficial for the fusion center to attach more importance to future costs as it should also take into account the benefit from sensors performing social learning.

Obs. matrix B^2	$\alpha_1 = 0.3132$	$\alpha_2 = 0.3032$	$\beta_1 = 0.11$
	$\beta_2 = 0.1$	$\gamma_1 = 0.1$	$\gamma_2 = 0.414$
Obs. matrix B^3	$\alpha_1 = 0.3233$	$\alpha_2 = 0.3133$	$\beta_1 = 0.11$
	$\beta_2 = 0.1$	$\gamma_1 = 0.1$	$\gamma_2 = 0.414$

Table 3.2: The reward vector (3.33) parameters for B^2 and B^3. For $\delta_1 = 0.3$, $\delta_2 = 0.95$, the following parameters were obtained as a solution of $\Delta(e_1) = 1$ and $\Delta(e_2) = 0$ for the reward vector (3.33) parameters with observation matrix $B = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$.

Sample Path of Optimal Incentives

This subsection illustrates the sample path properties of the optimal incentive sequence over time (which was characterized in Theorem 3.9 to be a sub-martingale). Fig.3.13 shows the average incentives provided to the social sensors over time. The fusion center employs the optimal incentive policy (3.39) and fuses the information revealed by social sensors in a Bayesian way (3.25). Each sample path has a duration of $N = 500$, i.e, sequential information fusion from 500 social sensors. The figure shows the average over 100 independent such sample paths for three different observation likelihood matrices (3.21). We consider the following observation likelihood matrices for illustrating the relation between the information gathering capabilities of the sensor (3.21) and the average incentives provided by the fusion center: B, B^2, and B^3. We know that B is more informative than B^2, which is in turn more informative than B^3, in the Blackwell sense [75].

Parameters:

The parameters of the incentive function (3.38) using the resolution dependent reward (3.33) for B^2 and B^3 are specified in Table 3.2. In Fig.3.13, it can be seen that the range (or the slope) of the average incentives over the time horizon is highest for the case of

observation matrix B (compared to B^2 and B^3). It can be seen from Fig.3.13 that the average incentives display an increasing trend.

3.3.8 Controlled Information Fusion in non-binary environments

In this section, we briefly discuss the formulation for multiple states. Partial results on social learning with multiple states and 2 actions appears in [74]. In the controlled fusion problem considered in this work, the social sensors reveal the observation to the fusion center. This requires that the cardinality of $\mathcal{A}$ and $\mathcal{Y}$ be equal. Due to the complexity of analyzing the structural results for the optimal policy in case of multiple actions and states, we only describe the formulation and illustrate the incentive policy using a numerical simulation for a $\mathcal{X} = \mathcal{A} = \mathcal{Y} = \{1, 2, 3\}$. When $|\mathcal{X}| = 3$, the public belief is in the belief space

$$\Pi(3) = \stackrel{\Delta}{=}\{\pi \in \mathbb{R}^2 : \sum_i \pi(i) = 1, 0 \leq \pi(i) \leq 1 \text{ for } i \in \{1, 2, 3\}\}.$$

The number of regions in the space $\Pi(3)$ that need be considered for analyzing the structural results of the optimal incentive policy are 5 (see (3.52) below) as opposed to 3 in (3.40).

Model Assumptions:

(A'1) The observation distribution $B_{xy} = \mathbb{P}(y|x)$ is TP2 (totally positive of order 2), i.e, all second order minors of matrix B are non-negative.

(A'2) The reward vector r_a is supermodular, i.e, $r_{a+1} - r_a$ is an increasing vector for $a = \{1, 2\}$ and every $p \in [0, 1]$.

The social sensors' decision $a(\pi, y) = \arg\max r_a' \eta_y$ is increasing in π and y under

(A'1) and (A'2); see [75]. This can be used to establish the single crossing condition,

$$\{\pi \in \Pi(3) : (r_a - r_{a+1})'\eta_y \leq 0\} \subseteq \{\pi \in \Pi(3) : (r_a - r_{a+1})'\eta_{y+1} \leq 0\}. \tag{3.51}$$

We now can define the following regions in the belief simplex $\Pi(3)$ (compare with (3.40)):

$$\begin{aligned}
&\mathcal{P}_1^p = \{\pi \in \Pi(3) : (r_1 - r_3)'\eta_{y=1} \cap (r_2 - r_3)'\eta_{y=1} \leq 0\},\\
&\mathcal{P}_2^p = \{\pi \in \Pi(3) : (r_1 - r_3)'\eta_{y=2} \leq 0 \cap (r_2 - r_3)'\eta_{y=2} \leq 0 \ \cap \ (r_1 - r_2)'\eta_{y=1} \leq 0\},\\
&\mathcal{P}_3^p = \{\pi \in \Pi(3) : (r_1 - r_3)'\eta_{y=3} \leq 0 \cap (r_2 - r_3)'\eta_{y=3} \leq 0 \ \cap \ (r_1 - r_2)'\eta_{y=2} \leq 0 \cap (r_2 - r_3)'\eta_{y=2} > 0\\
&\cap (r_1 - r_2)'\eta_{y=1} > 0 \cap \{(r_1 - r_3)'\eta_{y=1} > 0\},\\
&\mathcal{P}_4^p = \{\pi \in \Pi(3) : (r_1 - r_2)'\eta_{y=3} \leq 0 \cap (r_2 - r_3)'\eta_{y=3} > 0 \ \cap \ (r_1 - r_2)'\eta_{y=2} > 0 \cap (r_1 - r_3)'\eta_{y=2} > 0\},\\
&\mathcal{P}_5^p = \{\pi \in \Pi(3) : (r_1 - r_2)'\eta_{y=3} > 0 \cap (r_1 - r_3)'\eta_{y=3} > 0\}.
\end{aligned} \tag{3.52}$$

The value function for the fusion center is given by:

$$\begin{aligned}
V(\pi) &= \min\{c(p) + \rho \sum_a \sum_{j=1}^{5} V(T^j(\pi, a))\sigma(\pi, a)\mathcal{I}(\pi \in \mathcal{P}_j^p)\},\\
V(\pi) &= \min_{p\in[0,1]} \Big\{p - \phi_s \mathcal{I}(\pi \in \mathcal{P}_3^p) + \rho \sum_a \sum_{j=1}^{5} V(T^j(\pi, a))\sigma(\pi, a)\mathcal{I}(\pi \in \mathcal{P}_j^p)\Big\}.
\end{aligned} \tag{3.53}$$

Here $T^j(\pi, a) = \frac{R_a^j \pi}{\mathbf{1}' R_a^j \pi}$, with $R^j = BM^j$ for $j = 1, 2, \cdots, 5$.

Fig.3.14 shows the optimal incentive policy for a 3 state, observation, and action model. Lemma 3.4 can be used to find the matrices M^j for $j = 1, 2, \cdots, 5$. The observation distribution for the controlled fusion problem for 3 states and actions is chosen as:

$$B = \begin{bmatrix} 0.7479 & 0.1986 & 0.0536 \\ 0.6023 & 0.2543 & 0.1434 \\ 0.2785 & 0.2459 & 0.4756 \end{bmatrix}.$$

The value iteration algorithm based on (3.53) was run for a horizon $N = 100$.

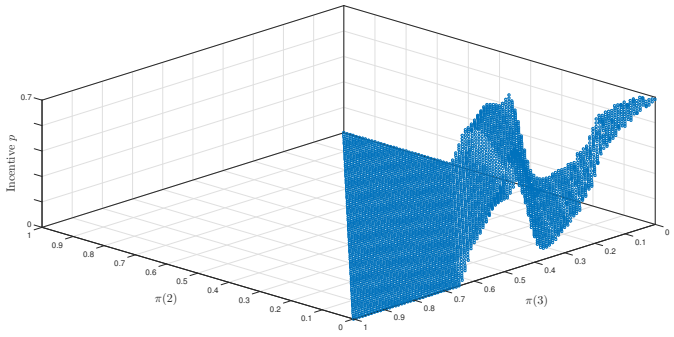

Figure 3.14: Optimal Incentive Policy for social learning weight $\phi_s = 0.6$, $\rho = 0.8, \beta_1 = 0.6771, \beta_2 = 0.5465, \beta_3 = 0.7113$, $\delta_1 = 0.3, \delta_2 = 0.4, \delta_3 = 0.5$, $\gamma_1 = 0.5, \gamma_2 = 0.3, \gamma_3 = 0.2$, and $\alpha_a = 0$ for all $a \in \{1, 2, 3\}$. The belief space $\Pi(3)$ was discretized into a grid of 5151 belief points using Fruedenthal triangulation [75]. The incentive $p \in [0, 1]$ was discretized into 50 values.

3.3.9 Conclusions

Unlike data fusion involving physical sensors for tracking targets, we considered information fusion with social sensors, which provide reviews on social media review platforms such as Amazon, Yelp, and Airbnb. Our main objective is to control the information fusion by dynamically providing incentives to the social sensors. We can draw the following conclusions:

1. The time path of the optimal incentive sequence is a sub-martingale – it increases over time. This is intuitive in the sense that, the future customers learn from their predecessors the value of the information they have about the underlying state of nature. So the fusion center needs to compensate more in order to make them reveal the truthful experiences.

2. The fusion center can aggregate the information about the underlying state, or estimate the true state, asymptotically.

3. The time path of the optimal incentive sequence is near optimal even if the states

are changing, provided the out-of-state transition probabilities are small.

4. Our results for two states and observations provide substantial insight into the nature of the complexity of controlled information fusion with social sensors, and highlight the means to derive structural results for the optimal policy in a multi-state case.

3.3.10 Appendix: Proofs

Proof of Theorem 3.8:

We first show that due to the structure of the social learning filter in (3.25), the choice of incentives reduces from a continuum $[0, 1]$ to a finite number at every belief. Next, we show that the incentive function $\Delta(\eta_y)$ is decreasing in π for any y.

Theorem 3.14. Let $\Delta(\eta_{y=1})$ and $\Delta(\eta_{y=2})$ be two possible incentives at belief π. Under (A1) and (A2), the Q function in (3.32) can be simplified as:

$$Q(\pi, p) = \begin{cases} p + \rho V(\pi) & \text{if } p \in [0, \Delta(\eta_{y=2})); \\ p - \phi_s + \rho \mathbb{E}V(\pi) & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})); \\ p + \rho V(\pi) & \text{if } p \in [\Delta(\eta_{y=1}), 1]. \end{cases} \tag{3.54}$$

and $V(\pi) = \min Q(\pi, p)$. Here, $\mathbb{E}V(\pi) = \mathbf{1}' B^{\pi}_{y=1} \pi \times V(\eta_{y=1}) + \mathbf{1}' B^{\pi}_{y=2} \pi \times V(\eta_{y=2})$.

Proof of Theorem 3.14:

From Lemma 3.10 and Theorem 3.5, we have

$$R^{\pi} = \begin{cases} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} & \text{if } p \in [0, \Delta(\eta_{y=2})); \\ \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})); \\ \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} & \text{if } p \in [\Delta(\eta_{y=1}), 1]. \end{cases} \tag{3.55}$$

From (3.55), it is clear that the sensors' decision

$$a = \begin{cases} 1 & \text{if } p \in [0, \Delta(\eta_{y=2})); \\ y & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})); \\ 2 & \text{if } p \in [\Delta(\eta_{y=1}), 1]. \end{cases} \tag{3.56}$$

Therefore,

$$\sum_{a \in \mathcal{A}} V(T^{\pi}(\pi, a))\sigma(\pi, a) = \begin{cases} V(\pi) & \text{if } p \in [0, \Delta(\eta_{y=2})); \\ \mathbb{E}V(\pi) & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})); \\ V(\pi) & \text{if } p \in [\Delta(\eta_{y=1}), 1]. \end{cases} \tag{3.57}$$

where $\mathbb{E}V(\pi) = \mathbf{1}' B^{\pi}_{y=1} \pi \times V(\eta_{y=1}) + \mathbf{1}' B^{\pi}_{y=2} \pi \times V(\eta_{y=2})$. □Theorem 3.14 represents the Q function (3.32) over the range $[0, 1]$ into *three* regions. The following corollary highlights why such a partition is useful.

Corollary 3.3.2. At every public belief $\pi \in \Pi(2)$, it is sufficient to choose one of the three incentives $\{0, \Delta(\eta_{y=2}), \Delta(\eta_{y=1})\}$.

Proof. From Theorem 3.14, the instantaneous reward is a linear function in p and

$$\arg\min_{p \in [0, \Delta(\eta_{y=2}))} Q(\pi, p) = 0, \quad \arg\min_{p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1}))} Q(\pi, p) = \Delta(\eta_{y=2}), \quad \arg\min_{p \in [\Delta(\eta_{y=1}), 1]} Q(\pi, p) = \Delta(\eta_{y=1}).$$

These hold as for any value of p in each of the three regions, the corresponding continuation payoff is the same from Theorem 3.14. □

Lemma 3.15. The incentive function $\Delta(\eta_y)$ is decreasing in π for every y.

Proof. The incentive function is given as (3.38), where $l_1, l_2, l_3 > 0$. With $\pi = [1 - \pi(2),\ \pi(2)]'$, differentiating w.r.t $\pi(2)$,

$$\frac{d(\Delta(\eta_y))}{d\pi(2)} = -(l_1 + l_2)B_{1y}B_{2y} < 0 \quad \square$$

Proof of Theorem 3.8:

From Corollary 3.3.2 the value function (3.32) is:

$$V(\pi) = \min\{\rho V(\pi), \Delta(\eta_{y=2}) - \phi_s + \rho\mathbb{E}V(\pi), \Delta(\eta_{y=1}) + \rho V(\pi)\}.$$

$$\Rightarrow V(\pi) = \min\{0, \Delta(\eta_{y=2}) - \phi_s + \rho\mathbb{E}V(\pi)\}, \tag{3.58}$$

as $\Delta(\eta_{y=1}) \geq 0$.

By using the value iteration algorithm [75] on (3.58), we have

$$V_{n+1}(\pi) = \min\{0, \Delta(\eta_{y=2}) - \phi_s + \rho\mathbb{E}V_n(\pi)\} \tag{3.59}$$

with $V_0(\pi) = 0\ \forall\ \pi$.

From Lemma 3.15, the incentive function is decreasing. From the definition of First-Order Stochastic Dominance (see Appendix A), and Lemma A.4, we have $\mathbb{E}V_n(\pi)$ is decreasing in π. Therefore, $V_{n+1}(\pi)$ and hence $V(\pi)$ is decreasing in π.

Let $V(0)$ and $V(1)$ denote the values for $\pi = \left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$. It is seen by substitution that $\mathbb{E}V(0) = V(0)$ and $\mathbb{E}V(1) = V(1)$. By definition, we know that $\Delta(\eta_y) \in [0, 1]$. Using[28] Lemma 3.15, let $\Delta(e_1) > \phi_s$ and $\Delta(e_2) < \phi_s$. The value function for the fusion center is given by (3.58). We have the following:

[28]Note that after normalization $\Delta(e_1) = 1$ and $\Delta(e_2) = 0$.

1. For $V(\pi) = \Delta(\eta_{y=2}) - \phi_s + \rho\mathbb{E}V(\pi)$, $V(0) = \frac{\Delta(e_1)-\phi_s}{(1-\rho)} > 0$, and $V(1) = \frac{\Delta(e_2)-\phi_s}{(1-\rho)} < 0$.
2. For $V(\pi) = 0$, $V(0) = V(1) = 0$.

The value function $V(\pi)$ in (3.58) is decreasing with a positive value at e_1 and a negative value e_2, so must be zero at some point(s). Let $\Sigma = \{\pi(2)|0 = \Delta(\eta_{y=2}) - \phi_s + \rho\mathbb{E}V(\pi)\}$. Since the value function $V(\pi)$ is monotone in π, the set Σ is convex.

Choosing $\pi_s^*(2) = \{\hat{\pi}(2)|\hat{\pi}(2) > \pi(2)\ \forall\ \pi(2) \in \Sigma\}$, the result follows. □

Proof of Theorem 3.9:

We will first establish the property of the incentive function $\Delta(\eta_y)$ as the optimal policy depends on it.

Lemma 3.16. Under (A1), $\Delta(\eta_{y=1})$ is concave in π, and $\Delta(\eta_{y=2})$ is convex in π.

Proof of Lemma 3.16:

The incentive function $\Delta(\eta_{y=2})$ is given in (3.38). A differentiable function $f : [0, 1] \to [0, 1]$ is convex if

$$f(w_1) \geq f(w_2) + f'(w_2)(w_1 - w_2), \text{for all } w_1, w_2 \in [0, 1]. \tag{3.60}$$

A function f is concave if $-f$ is convex.

From (3.60) with $w_1 = \pi_1(2)$ and $w_2 = \pi_2(2)$, and using Lemma A.4, it is verified that the function $\Delta(\eta_{y=2})$ is convex in π. Similarly, it can be shown that $\Delta(\eta_{y=1})$ is concave in π. □

Proof of Theorem 3.9:

Consider the sub-optimal policy $\hat{\mu}(\pi)$ given as

$$\hat{\mu}(\pi) = \begin{cases} \Delta(\eta_{y=2}) - \epsilon & \text{if } \pi(2) \in [0, \pi_*(2)); \\ \Delta(\eta_{y=2}) & \text{if } \pi(2) \in [\pi_*(2), 1]. \end{cases}$$

Here $\epsilon > 0$ and $\pi_*(2) \in [0, 1]$. Let $W_k = \hat{\mu}(\pi_{k-1})$.

From Lemma 3.16, $\Delta(\eta_{y=2})$ is convex in π. Let $u^S(\pi_{k+1}) = \Delta(\eta_{y_k=2})$ denote the price at

time $k+1$. So $u^S(\pi)$ is convex in π.

We know that the public belief π_k is a martingale ([33]), i.e, $\mathbb{E}[\pi_{k+1}|\mathcal{F}_k] = \pi_k$. For $\epsilon \to 0$,

$$\mathbb{E}[W_{k+1}|\mathcal{F}_k] = \mathbb{E}[u^S(\pi_{k+1})|\mathcal{F}_k] \geq u^S(\mathbb{E}[\pi_{k+1}|\mathcal{F}_k]) \geq u^S(\pi_k) \geq W_k$$

by Jensen's inequality and martingale property of the public belief. Therefore $W_k(= \hat{\mu}(\pi_{k-1}))$ is a sub-martingale.

Consider a function $\bar{\mu}(\pi)$ given by

$$\bar{\mu}(\pi) = \begin{cases} 0 & \text{if } \pi(2) \in [0, \pi^*(2)); \\ 1 & \text{if } \pi(2) \in [\pi^*(2), 1]. \end{cases}$$

Let $H_k = \bar{\mu}(\pi_{k-1})$. From Theorem A.5, $(H.W)_k$ is a sub-martingale. But $(H.W)_k = p_k$. Therefore, the optimal incentive sequence $p_k = \mu^*(\pi_{k-1})$ is a sub-martingale, $\mathbb{E}[p_{k+1}|\mathcal{F}_k] \geq p_k$, i.e, it increases on average over time. □

Proof of Theorem 3.10:

We'll prove that $\pi \in \mathcal{P}_2^p$ iff $p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1}))$. Other cases are proved similarly.

We can write

$$r_1 = [(\delta_1 p - \gamma_1) \;\; (\delta_1 p - \alpha_1 - \gamma_1)], r_2 = [(\delta_2 p - \alpha_2 - \gamma_2) \;\; (\delta_2 p - \gamma_2)]. \tag{3.61}$$

By definition,

$$\mathcal{P}_2^p = \{\pi \in \Pi(2) : (r_1 - r_2)'\eta_{y=1} > 0 \;\cap\; (r_1 - r_2)'\eta_{y=2} \leq 0\}.$$

We have,

$$(r_1 - r_2)'\eta_{y=1} > 0 \Leftrightarrow p < \frac{1}{\delta_2 - \delta_1}\left\{[\alpha_2 \;\; - \alpha_1]\eta_{y=1} + (\gamma_2 - \gamma_1)\right\} = \Delta(\eta_{y=1}).$$
$$(r_1 - r_2)'\eta_{y=2} \leq 0 \Leftrightarrow p \geq \frac{1}{\delta_2 - \delta_1}\left\{[\alpha_2 \;\; - \alpha_1]\eta_{y=2} + (\gamma_2 - \gamma_1)\right\} = \Delta(\eta_{y=2}).$$

Proof of Theorem 3.11:

Define the following region in the belief space $\Pi(2)$:

$$\mathcal{H} = \{\pi | \pi(2) \leq \pi^*(2)\}. \tag{3.62}$$

Here $\mathcal{H}$ denotes the region where the optimal policy in (3.39) is such that $\mu^*(\pi) = 0$. For any sub-optimal policy μ_c and the corresponding cost $W_{\mu_c}(\pi)$, it is clear that $W_{\mu_c}(\pi) - J_{\mu^*}(\pi) \geq 0 \; \forall \; \pi$. Also, $W_{\mu_c}(e_2) = J_{\mu^*}(e_2)$. Let $\mathcal{I}$ denote the indicator function. We have

$$W_{\mu_c}(\pi) - J_{\mu^*}(\pi) = \mathcal{I}(\pi \in \mathcal{H})\{W_{\mu_c}(\pi) - J_{\mu^*}(\pi)\} + \mathcal{I}(\pi \notin \mathcal{H})\{W_{\mu_c}(\pi) - J_{\mu^*}(\pi)\}$$
$$\Rightarrow \sup_{\pi} |W_{\mu_c}(\pi) - J_{\mu^*}(\pi)| \leq \Big\{ \sup_{\pi} \mathcal{I}(\pi \in \mathcal{H})\{W_{\mu_c}(\pi) - J_{\mu^*}(\pi)\}\Big\} + \Big\{ \sup_{\pi} \mathcal{I}(\pi \notin \mathcal{H})\{W_{\mu_c}(\pi) - J_{\mu^*}(\pi)\}\Big\}. \tag{3.63}$$

where $\mathcal{H}$ is defined in (3.62). From Theorem 3.8, we know that $J_{\mu^*}(\pi) = V(\pi)$ is monotone (non-increasing) in π. Similar arguments can be used to establish that $W_{\mu_c}(\pi)$ is monotone (non-increasing) in π. Therefore, we have[29] for (3.63)

$$\sup_{\pi} |W_{\mu_c}(\pi) - J_{\mu^*}(\pi)| \leq 2\Big\{ \sup_{\pi} \mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\Big\} \tag{3.64}$$

as $J_{\mu^*}(\pi) = 0 \; \forall \; \pi \in \mathcal{H}$ from (3.58) and Theorem 3.8. The set $\mathcal{H}$ defined in (3.62) is compact by definition. For the discount factor $\rho \in [0, 1)$ and bounded instantaneous costs, the cumulative discounted cost is bounded [75]. Therefore in (3.64),

$$\sup_{\pi}\{\mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\} = \max_{\pi}\{\mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\}$$

and $\tilde{\pi} = \arg\max_{\pi}\{\mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\}$. We have for $\pi_0 = \tilde{\pi}$,

$$\begin{aligned}
\max_{\pi}\{\mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\} &= \mathbb{E}_{\mu_c}\Big\{ \sum_{k=0}^{\infty} \rho^k \{c_{\mu_c}(p_k)\} \Big| \pi_0 = \tilde{\pi}\Big\} \leq \mathbb{E}_{\mu_c}\Big\{ \sum_{k=0}^{\infty} \rho^k \max_{\Delta(\eta_{y=2}):\pi \in \mathcal{H}} c_{\mu_c}(p_k)\Big\} \\
&= (1 - \phi_s)\mathbb{E}\Big\{ \sum_{k=0}^{\infty} \rho^k \Big\} \\
&= \frac{(1 - \phi_s)}{1 - \rho}. \quad \square
\end{aligned}$$

[29] Equation (3.64) follows from

$$\mathbb{E}_{\mu_c}\Big\{ \sum_{k=0}^{\infty} \rho^k \{c_{\mu_c}(p_k)\}\Big\} > \mathbb{E}_{\mu_c}\Big\{ \sum_{k=0}^{\infty} \rho^k \{\mathcal{I}(\pi_k \in \mathcal{H}) c_{\mu_c}(p_k)\}\Big\}.$$

CHAPTER 4

STOCHASTIC CONTROL WITH SOCIAL INFLUENCE: FRAMEWORK-III

Framework-III: Consider the following specifications: (i) the individual decision makers interact and influence each other's opinions/ actions over a social network, (ii) the optimal policy of the operator is determined as the solution that achieves the stochastic control objective, which is to estimate the state with minimum incurred cost.

So far, we have discussed people-centric stochastic control where, the learning was modeled using Bayesian social learning. Each sensor was assumed to act once after observing all its' predecessors. However, when the sensors are learning from only from their neighbors over a social network and are acting repeatedly, Bayesian social learning is intractable [66, 55]. Therefore, we consider an influence network, where the social sensors are situated at different levels in the influence hierarchy. The opinions are influenced in one direction, and the sensors can be polled to provide the information that they posses. In other words, the learning model is relaxed to an influence model for tractability.

We first describe the problems considered in the literature that fit into Framework-III. We then illustrate an application of Framework-III in adaptive polling on hierarchical social influence networks[1].

4.1 Relevant Problems in Literature

Framework-III is applicable in a wide range of situations. Below, we illustrate a few.

Information aggregation over social networks. Twitter was used in disaster recov-

[1] Bhatt, S., & Krishnamurthy, V. Adaptive Polling in Hierarchical Social Networks using Blackwell Dominance. Accepted at IEEE Transactions on Signal and Information Processing over Networks, 2019.

ery [116], real-time information related to the Boston Marathon bombing [132], to monitor gas availability in New York City during and shortly after Hurricane Sandy [132], and to monitor the US National Football League (NFL) games in real-time [135].

Polling on social networks. Polling has numerous applications such as forecasting the outcome of an election[126, 120], estimating the fraction of individuals infected with a disease [54], and predicting the success of a particular product. [114] analyzes all US presidential electoral college results from 1952 – 2008 where both intention and expectation polling were conducted and shows a remarkable result: In 77 cases where expectation and intent polling pointed to different winners, expectation polling was accurate 78% of the time! Unlike [114], we consider a Bayesian approach, and also consider the hierarchical influence structure along with the time-varying nature of the state (adaptive polling). [77] analyzes a Bayesian approach to intent and expectation polling and illustrates how the posterior distribution of the leading candidate in the poll can be estimated based on incestuous estimates (each node summarizes the belief of its neighbors, which in turn are influenced by the nodes belief). Unlike [77], we consider hierarchical influence structure and feedback control, in the sense that current estimate dictates where and how to poll in the hierarchical network. [42] investigates the role played by the network structure in polling by considering the trade-off between number of polled individuals and the bias introduced due to the network structure. [42] concludes that the estimators that consider the network structure into account are considerably more efficient than standard polling estimators. We take the (influence) structure of the network into account, but unlike [42], propose adaptive versions of the polling algorithms. [83] presents a dynamic Bayesian forecasting method that systematically combines information from historical forecasting models in real time with results from the large number of state-level opinion surveys that are released publicly during the campaign. Similar to [83] we consider a dynamic polling method, but unlike [83] also take the influence

structure of the social network into account.

4.2 Adaptive Polling in Hierarchical Social Influence Networks

Suppose a hedge fund wants to mobilize the Twitter chatter related to a particular stock or a company they want to invest in. The chatter has information that might help make money for the hedge fund. It is not economical to store and process the entire network data when the desired feature related to the company or stock evolves over time. How should the hedge fund sequentially poll the Twitter network to estimate the desired feature as it evolves over time, with minimum computational and storage cost?

4.2.1 Formulation of Adaptive Polling

We consider the typical framework for information diffusion and formation of opinions in a social network. The underlying state (true sentiment underlying social media message, popularity of a product/political party, quality of commercial product) evolves over time stochastically [116, 135, 34, 132, 26]. This underlying state is observed by the individuals in the social network through tweets, political commentary blogs and videos, or reviews on social media. Using the available information and interaction with neighbours, individuals form opinions about the underlying state.

How should the pollster poll the hierarchical social network to estimate the state while minimizing the polling cost (measurement cost and uncertainty in the Bayesian state estimate)? We formulate this adaptive polling problem as a partially observed Markov decision process (POMDP).

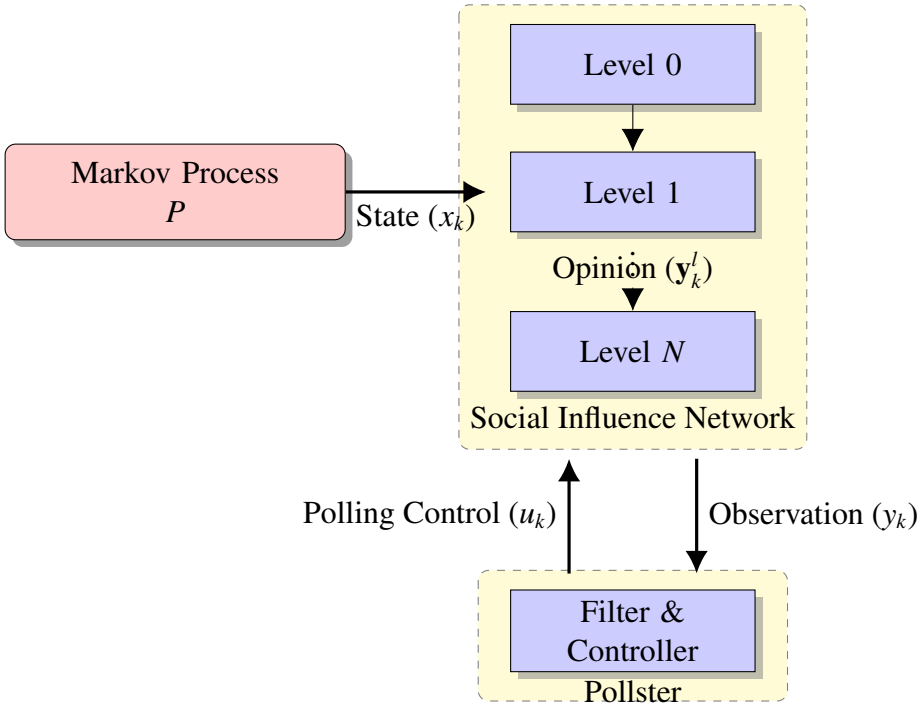

Figure 4.1: The figure shows a simple hierarchical influence network where the individuals are grouped into $N+1$ levels Level 0, Level $1, \cdots$, Level N in a hierarchical fashion. Each level influences the opinion of the level below it. The underlying state of nature x_k determines the opinion. A pollster samples observations y_k from the nodes having opinions $\mathbf{y}_k^l$, runs a local filter to compute the state estimate, and chooses a control to affect the (future) polling action. It is assumed that the pollster knows the number of hierarchical levels in the network and the corresponding node associations. The aim of the pollster is to estimate the underlying state by adapting its polling strategy to incur minimum polling cost.

A: Model of Adaptive Polling

The population is classified into $N + 1$ levels depending on the hierarchical influence as shown in Fig.4.1. Many social networks have a hierarchical influence structure; see for example [84, 5, 118, 22, 25]. The population is sampled sequentially by a pollster to gather the information on the underlying state.

The POMDP for adaptive polling is specified by

$$\theta = (\mathcal{X}, \mathcal{Y}, \mathbb{Y}, \mathcal{U}, C, P, O(u), \rho), \tag{4.1}$$

where $\mathcal{X}$ denotes the state space, $\mathcal{Y}$ denotes the observation space, $\mathbb{Y}$ denotes the opinion space, $\mathcal{U}$ denotes the control/ action space, C denotes the state-action cost matrix, P is the state transition matrix, $O(u)$ is the control dependent observation distribution/ likelihood matrix and $\rho \in [0, 1)$ is the economic discount factor. We now describe the above 8 components of the model (4.1):

1. *State:* Let $x_k \in \mathcal{X} = \{1, 2, \cdots, X\}$ denote a Markov chain evolving at discrete time instants $k = 0, 1, \cdots$ on a finite state space. As mentioned previously, the state models the time evolving ground truth (sentiment, popularity, quality) quantized into a finite number of levels.
2. *Transition matrix:* Let P denote the time-homogeneous transition probability matrix of the Markov chain x_k with elements

$$P_{ij} = \mathbb{P}(x_{k+1} = j | x_k = i), \quad i, j \in \mathcal{X}, \ \forall \ k. \tag{4.2}$$

3. *Pollster's control / actions:* $\mathcal{U} = \{1, 2, \cdots, U\}$ denotes the set of possible controls (actions), with $u_k \in \mathcal{U}$ denoting the action chosen at time k. For example, the action can denote the choice of the hierarchical level the pollster seeks the opinion.
4. *Polling cost:* Let $C(x_k, u_k)$ denote the instantaneous cost incurred by the pollster for taking action u_k when in state x_k. This models the measurement cost and quality (accuracy) of the polling algorithm. For example, conducting surveys and opinion polls incurs a measurement cost and the type of poll conducted affects the quality of the information gathered.
5. *Pollster's observation distribution:* Let $\mathcal{Y}$ denote a finite set of observations with $y_k \in \mathcal{Y}$ representing the observations of the underlying state $x_k \in \mathcal{X}$. The obser-

vations $y_k \in \mathcal{Y}$ for the pollster model the information on the state gathered via views/ sentiments expressed by the nodes or individuals in the hierarchical social network (see Fig.4.1). Let $O(u)$ denote the observation probability matrix with elements

$$O_{ij}(u) = \mathbb{P}(y_{k+1} = j | x_{k+1} = i, u_k = u), \quad i \in \mathcal{X}, j \in \mathcal{Y}, \ \forall k. \tag{4.3}$$

At each time k, the pollster receives an observation y_k on the underlying state x_k after taking an action u_k. The observation matrix/ distribution $O(u)$ models the likelihood of the observations $y_k \in \mathcal{Y}$ given the state $x_k \in \mathcal{X}$, and is different for different polling algorithms.

The observations obtained by the pollster are the *opinions* about the state provided by the nodes. We now discuss the opinion dynamics, the corresponding opinion distribution, and how the observation distribution can be expressed in terms of the opinion distributions:

i.) ***Opinion dynamics***: Let $\mathbf{y}_k^l \in \mathbb{Y}$ denote the opinion of nodes at level l of the hierarchical network (see Fig.4.1). Here $|\mathbb{Y}| = |\mathcal{X}|$. The opinion dynamics in Fig.4.1 proceeds according to the following protocol for $k = 0, 1, \cdots$

   i. The state x_k evolves on time scale k.

   ii. Opinions $\mathbf{y}_k^l$, for $l = 1, 2, \cdots, N$, are formed at the Level l at the fast time-scale $\bar{k} = k + l\delta$, where $0 < \delta \ll 1$.

   iii. At time $k + 1$, state transitions to x_{k+1}.

   We assume that $N\delta \ll 1$, where the number of levels in Fig. 4.1 is $N + 1$. This implies that the state x_k is evolving over a slower time-scale than the time-scale over which the opinions are formed across the network given in Fig.4.1.

ii.) ***Opinion distribution***: The opinions at different levels in the hierarchical social influence network are formed via information diffusion as fol-

lows [133, 86]: opinion at the highest level $\mathbf{y}_k^0$ is directly influenced by the state x_k. Opinion $\mathbf{y}_k^l$, $l \geq 0$ influences $\mathbf{y}_k^{l+1}$ (see Fig.4.1). This is modeled probabilistically as $\mathbb{P}(\mathbf{y}_k^{l+1} = j|\mathbf{y}_k^l = i)$.

Let the opinion distribution at Level 0 be given by the stochastic matrix B having elements

$$B_{ij} = \mathbb{P}(\mathbf{y}_k^0 = j|x_k = i), \quad i \in \mathcal{X}, j \in \mathbb{Y}, \forall k. \tag{4.4}$$

The opinions at levels $l \in \{1, \cdots, N\}$ in the hierarchical network are directly influenced by the preceding levels (see Fig.4.1). The opinions at levels $l \in \{1, \cdots, N\}$ are given by $(B_l)_{ij} = \mathbb{P}(\mathbf{y}_k^l = j|x_k = i)$ for $i \in \mathcal{X}$, $j \in \mathbb{Y}$, and $\forall k$. The opinions at Level l are determined by the opinion distribution via the following decomposition

$$(B_l)_{ij} = \sum_{m \in \mathcal{Y}} \mathbb{P}(\mathbf{y}_k^l = j|\mathbf{y}_k^{l-1} = m)\mathbb{P}(\mathbf{y}_k^{l-1} = m|x_k = i). \tag{4.5}$$

For tractability, assume[2] that the confusion matrix between successive levels is modeled using the same time-homogeneous opinion distribution B in (4.4). So the opinions at levels $l \in \{0, 1, \cdots, N\}$ have an opinion distribution

$$B_l = B^{l+1}, \tag{4.6}$$

where B_l denotes the opinion distribution at level l.

iii.) ***Observation distribution via Opinion distribution***: Since the observations for the pollster, to update the estimate of the state, are the opinions from the nodes, the observation distribution (4.3) is directly related to the opinion distribution (4.4). For example, in case of adaptive intent polling (Sec.4.2.3

[2]This is a modeling assumption, and Example 1 in Sec.4.2.7 shows how to estimate such a structure using a modified EM algorithm. For the case where it is known a priori that the distribution (confusion matrix) between the hierarchical levels are different, results in Sec.4.2.5 on Approximate Blackwell Dominance can be used to obtain the policies and performance bounds on the proposed polling algorithms in Sec.4.2.3 and Sec.4.2.4.

below), $O(u) = Bf_u(B)$, where f_u is any matrix polynomial, where the probabilities with which the nodes at different levels in the hierarchical social network are polled are proportional to the co-efficients of the polynomial f_u; and in case of adaptive expectation polling (Sec.4.2.4 below), $O(u) = B_l^{l_u/l}$, where the nodes at level l are polled to provide information on the nodes at level l_u.

B: Polling Objective

The actions taken by the pollster influences the noisy state-observations via the selection of the observation distribution. The goal of the pollster is to choose an action, based on the history of past actions and observations, that minimizes the expected costs incurred over time. We consider the following infinite horizon discounted cost for specifying the objective [75, Chapter 7]:

$$J_\mu(\pi_0;\theta) = \mathbb{E}_\mu\{\sum_{k=0}^{\infty}\rho^k C(x_k, u_k = \mu(\mathcal{I}_k))|\pi_0\}. \tag{4.7}$$

Here $J_\mu(\pi_0;\theta)$ denotes the expected cumulative cost with respect to the stationary (time-independent) policy μ, $\mathcal{I}_k = \{\pi_0, u_0, y_1, \cdots, u_{k-1}, y_{k-1}\}$ denotes the history of past actions and observations and $\pi_0 = (\pi_0(i), i \in \mathcal{X})$, where $\pi_0(i) = \mathbb{P}(x_0 = i)$ is the initial probability distribution over the state space. The objective of the pollster is to find the optimal stationary policy μ^* such that

$$J_{\mu^*}(\pi_0;\theta) = \inf_{\mu\in\boldsymbol{\mu}} J_\mu(\pi_0;\theta) \tag{4.8}$$

where $\boldsymbol{\mu}$ denotes the class of stationary policies.

Discounting in Polling:

The parameter $\rho \in [0, 1)$ is an economic discount factor that determines the way the polling cost is counted towards the *polling value* defined in (3.31), and affects the optimal policy μ^*. A discounting of $\rho = 0$ implies that the pollster is myopic, in that it

only minimizes the instantaneous polling cost $C(x, u)$ without considering future polling costs. A discounting of $\rho > 0$ implies that the pollster geometrically weighs the polling costs incurred in the future.

Summary:

We have formulated the adaptive polling problem as a POMDP with parametrization θ defined in (4.1). The infinite horizon objective (4.7) is for notational convenience. For a finite horizon formulation - the optimal policy is non-stationary - but all subsequent results continue to hold. Our main result will exploit Blackwell dominance to construct a myopic upper bound to the optimal policy μ^* in (3.31).

Discussion of Model

1. We assume that the POMDP model θ in (4.1) is known. This implies that the number of hierarchical levels and nodes-level associations are known to the pollster. Otherwise the problem becomes an adaptive stochastic control problem which is intractable to solve. Note that Blackwell dominance is a class type result - even if the observation probabilities (4.3) are not known exactly, as long as the Blackwell dominance condition is satisfied, the main result (Theorem 4.1 below) holds.
2. The opinion dynamics are such that the entire network holds the view on the state x_k at time k, between times k and $k + 1$. This modeling assumption has two implications: (i) It enables the decomposition of the opinion distributions (4.5), (ii) It implies that during the sampling instant, the information gathered by the pollster, from anywhere in the hierarchical network, pertains to the same underlying state.
3. Opinions from higher levels (see Fig.4.1) are more informative (Blackwell sense) and hence the information acquisition is costlier compared to lower levels. This is

motivated by study in [30], which shows that information acquisition from more informative distributions (in the sense of Blackwell) is more costly. The intuition is that nodes at a higher level pay more attention to acquire information to form an opinion, and hence require commensurate compensation to divulge that information to the pollster. The assumption on the cost $C(x, u)$ captures this intuition.

4. *Example*: Consider estimating the subject clarity (quality of teaching) as $x(\text{Subject Clarity}) = \{\text{Unclear}, \text{Clear}\}$ in live interactive audience participation real world platforms like *Poll Everywhere*. The pollster's observations and opinions formed by the nodes at level l could be modeled as $y = \mathbf{y}^l = \{\text{Choice.1}, \text{Choice.2}\}$. The pollster incurs a measurement cost for obtaining information from the participants, and a cost for the accuracy or the uncertainty reduction in the state estimate. In this example application, polling opinions from all the participants incurs the same measurement cost, but opinions from the knowledgeable participants, who influence and lead the discussions, will result in a larger reduction in uncertainty in the state estimate. The objective of the pollster is to estimate the quality of teaching by polling observations from the participant pool while incurring the least polling cost on average.

C: Stochastic Dynamic Programming for Adaptive Polling

In this section, we formulate the solution of (4.8) as a stochastic dynamic programming problem over the $X-$ dimensional unit simplex $\Pi(X) = \{\pi : \pi(i) \in [0, 1], \ \sum_{i=1}^{X} \pi(i) = 1\}$ of posterior probabilities (beliefs).

1. *Belief State Formulation:* Let π_k denote the belief at time k and the i^{th} element $\pi_k(i)$ is:

$$\pi_k(i) = \mathbb{P}(x_k = i|\mathcal{I}_k) \tag{4.9}$$

where $x = \{1, 2, 3, \cdots, X\}$ denotes the state space and $\mathcal{I}_k = \{\pi_0, u_0, y_1, \cdots, u_{k-1}, y_{k-1}\}$ denotes the history of past actions and observations. The belief (3.20) is computed from the opinions gathered by the pollster, and is a sufficient statistic [75] for the history of actions and opinions $\{u_1, y_1, \ldots, u_{k-1}, y_{k-1}\}$. The dynamics of the POMDP is given by the Bayesian filtering update

$$\pi_k = T(\pi_{k-1}, y_k, u_k), \text{ where } T(\pi, y, u) = \frac{O_y(u)P'\pi}{\mathbf{1}'O_y(u)P'\pi} \tag{4.10}$$

and $O_y(u) = \text{diag}(O_{1y}(u), \cdots, O_{Xy}(u))$. Here $\mathbf{1}$ is the column vector of 1s and P' denotes the matrix transpose.

As is well known in POMDPs, instantaneous cost $C(\pi_k, u_k)$ in terms of the belief π_k given by

$$C(\pi_k, u_k) = \sum_i C(x_k = i, u_k)\pi_k(i), \tag{4.11}$$

where π_k is the belief at time k. The costs $C(\pi, u)$ in (4.11) capture the cost of measurement and the uncertainty or error in the state estimate. In order to capture the uncertainty in the estimate it is necessary to use a non-linear cost on the belief state– this results in a non-standard POMDP (classical POMDPs use a linear cost in the belief but cannot capture uncertainty in the estimate). The non-linearity is required so that the costs are zero at the vertices of the belief space $\Pi(X)$ (reflecting perfect state estimation) and largest at the centroid of the belief space (most uncertain estimate). Any non-linear cost can be used in the formulation of the polling problems. In this work, we consider the following non-linear costs [75, Chapter 8] – entropy and state-estimation error – to illustrate the different formulations.

Associated with a stationary polling policy μ and initial belief $\pi_0 \in \Pi(X)$, the objective (4.7) can be re-expressed as:

$$J_\mu(\pi_0; \theta) = \mathbb{E}_\mu\{\sum_{k=0}^{\infty} \rho^k C(\pi_k, u_k = \mu(\pi_{k-1}))|\pi_0\}. \tag{4.12}$$

Our aim is to find the optimal stationary polling policy $\mu^* : \Pi(X) \to \mathcal{U}$ defined in (3.31).

2. *Stochastic Dynamic Programming:* Obtaining the optimal policy μ^* in (4.8) is equivalent to solving Bellman's stochastic dynamic programming equation [75, Chapter 7]:

$$\begin{aligned} \mu^*(\pi) &= \arg\min_{u \in \mathcal{U}} Q(\pi, u) \\ J_{\mu^*}(\pi; \theta) = V(\pi) &= \min_{u \in \mathcal{U}} Q(\pi, u), \text{where} \\ Q(\pi, u) &= C(\pi, u) + \rho \sum_{y \in \mathcal{Y}} V(T(\pi, y, u)) \sigma(\pi, y, u). \end{aligned} \tag{4.13}$$

Here $\sigma(\pi, y, u) = \mathbf{1}' O_y(u) P' \pi$ is the measure on the observation alphabet $\mathcal{Y}$, $V(\pi)$ is the value function denoting the minimum expected cost, $T(\pi, y, u)$ is the Bayesian filtering update, $Q(\pi, u)$ is the state-action value function ($Q-$ function), $C(\pi, u)$ is the state-action cost in terms of the belief state (4.11) and ρ is the discount factor. Since the belief space $\Pi(X)$ is a continuum, Bellman's equation (4.13) does not translate into practical solution methodologies as $V(\pi)$ needs to be evaluated at each $\pi \in \Pi(X)$. The computation of optimal policy of the POMDP is P-SPACE hard [100]. Also, the costs $C(\pi, u)$ capture the cost of measurement and the uncertainty or error in the state estimate, and hence are non-linear in the belief. In order to capture the uncertainty in the estimate it is necessary to use a non-linear cost on the belief state. The non-linearity is required so that the costs are zero at the vertices of the belief space $\Pi(X)$ (reflecting perfect state estimation) and largest at the centroid of the belief space (most uncertain estimate). This results in a non-standard[3] POMDP. This motivates the construction of *optimal upper bound policy* $\bar{\mu}(\pi)$ to the optimal policy $\mu^*(\pi)$ that is inexpensive to compute. In the remainder of the section, we construct such upper bound policies in terms of eas-

[3]POMDP solvers can only handle POMDPs with linear costs, see [75, Chapter 7].

ily computable myopic policies using *Blackwell dominance*, for adaptive polling problems formulated as POMDPs.

4.2.2 Meta-theorems for Adaptive Polling

Given the POMDP model for adaptive polling (4.1) and the polling objective, the aim of this section is to describe the key meta theorems that will be used in subsequent sections to develop efficient adaptive polling algorithms. In this section, we provide two main theorems using Blackwell dominance: (i) A structural result using Blackwell dominance for the adaptive polling POMDP, i.e characterization of achievable performance without brute force computations but using mathematical analysis, (ii) An information theoretic consequence of Blackwell dominance, namely, Rényi divergence, and why this is useful for the pollster.

Optimality of Myopic Polling Policies

The aim of the pollster is to estimate the (underlying) evolving state x_k by incurring minimum cumulative cost (4.12). The pollster employs the control $u_k = \mu^*(\pi_{k-1})$ to obtain opinions ($y_k \in \mathcal{Y}$) from the nodes, and then updates the belief $\pi_{k-1} \to \pi_k$ about the underlying state $x_k \in \mathcal{X}$ using (4.10).

Define the myopic policy $\bar{\mu}(\pi)$ as

$$\bar{\mu}(\pi) = \arg\min_{u \in \mathcal{U}} C(\pi, u) \tag{4.14}$$

Theorem 4.1 below provides sufficient conditions on the observation distribution of the pollster $O(u)$ so that a myopic polling policy (4.14) upper bounds the optimal polling policy in (3.31).

Theorem 4.1 (Optimality of Myopic Policies via Blackwell Dominance)**.** Consider the adaptive polling problem formulated in Sec.4.2.1 as a POMDP. Assume that the cost $C(\pi, u)$ is concave in π. Suppose $O(u) \succeq_B O(u+1)\ \forall u \in \mathcal{U}$. Then $\bar{\mu}(\pi)$ is an upper bound to the optimal polling policy $\mu^*(\pi)$, i.e, $\mu^*(\pi) \leq \bar{\mu}(\pi)$ for all $\pi \in \Pi$. In particular, for belief states where $\bar{\mu}(\pi) = 1$, the myopic policy coincides with the optimal policy $\mu^*(\pi)$.

The concavity of the costs $C(\pi, u)$ for each polling action u implies that the value function $V(\pi)$ is concave [75, Theorem 8.4.1] on $\Pi(X)$. This together with Jensen's inequality is used to establish Theorem 4.1. Theorem 4.1 is a well known structural result for POMDPs [75, Chapter 14, Sec.14.7], and it says the following:

i.) The instantaneous costs satisfying $C(\pi, 1) < C(\pi, u)$ for $u = 2, \cdots, U$ does not trivially imply that the myopic policy $\bar{\mu}(\pi)$ in (4.14) coincides with the optimal policy $\mu(\pi)$, since the optimal policy applies to the cumulative cost function involving an infinite horizon trajectory of the dynamical system. But when $O(u) \succeq_B O(u+1)\ \forall\ u \in \mathcal{U}$ and the costs are concave, the myopic policy coincides and forms a provably optimal upper bound to the computationally intractable optimal policy. The concavity of the costs imply that extremes are better than averages, reflecting perfect state estimation; and captures the effect of increasing marginal utility.

ii.) The trivial sub-optimal policy $\hat{\mu}(\pi) = 1\ \forall \pi \in \Pi(X)$ is also an upper bound to the optimal policy - but a useless bound because $\mu^*(\pi) = 1 \implies \hat{\mu}(\pi) = 1$. In comparison, the upper bound constructed via Theorem 4.1 (Blackwell dominance) says that $\bar{\mu}(\pi) = 1 \implies \mu^*(\pi) = 1$, which is a much more useful construction. Thus, the myopic polling policy forms a provably optimal upper bound to the computationally intractable optimal policy.

Blackwell Dominance and Rényi Divergence Interpretation

In this section, we will discuss the information theoretic consequences of Blackwell dominance. While Blackwell Dominance helps compute inexpensive policies that provably upper bound the computationally intractable optimal policy, the information theoretic consequences guide the choice of observation channels (likelihoods) for the pollster.

Rényi Divergence is a generalization of the Kullback-Leibler divergence [41], and it measures the dissimilarity between two distributions. Theorem 4.2 below shows the relation between Rényi Divergence and Blackwell dominance.
With a slight abuse of notation in (4.3), let $O_i(u)$ denote the i^{th} row of the observation likelihood matrix $O(u)$. In words, $O_i(u)$ is the distribution over the observation alphabet $\mathcal{Y}$ conditional on the state $x = i$.

Definition. (Rényi Divergence) For an observation likelihood $O(u)$, the Rényi Divergence of order $\alpha \in [0, 1)$ for $i, j \in \mathcal{X}$ is defined as

$$D_\alpha(O_i(u)||O_j(u)) = \frac{1}{\alpha - 1} \log \sum_{y \in \mathcal{Y}} O_{iy}^{\alpha}(u) O_{jy}^{\alpha-1}(u). \tag{4.15}$$

Theorem 4.2 (Ordering of Rényi Divergence)**.** If the observation distribution for the pollster satisfy $O(u) \succeq_B O(u + 1)\ \forall u \in \mathcal{U}$, then for any $i, j \in \mathcal{X}$:

$$D_\alpha(O_i(u)||O_j(u)) \geq D_\alpha(O_i(u + 1)||O_j(u + 1))\ \forall u \in \mathcal{U}. \tag{4.16}$$

Theorem 4.2 says that when $O(u) \succeq_B O(u + 1)$, conditional on the state, the observation distributions are more dissimilar in case of $O(u)$. Here, more the dissimilarity, better the pollster is able to distinguish the states. In other words, Theorem 4.2 provides a ranking of channel structures in terms of their ability to distinguish the states. In Sec.4.2.3 and Sec.4.2.4, we discuss the ordering of Rényi Divergence for more general channels that

arise in hierarchical social influence networks, where the information theoretic consequences guide the choice of the observation distributions for the pollster.

4.2.3 Main Result. Adaptive Intent Polling Algorithm

In intent polling [114], to decide between two states, the sampled individuals are asked "who will you vote for?". In this section, we develop an adaptive version of intent polling [114]: the resulting algorithm (Algorithm 1 below) is designed for hierarchical social networks with time-varying state of nature.

We present novel sufficient conditions for Blackwell dominance in the context of adaptive intent polling. These conditions enable the application of Theorem 4.1 to determine myopic policies that upper bound the optimal adaptive intent polling policy. These myopic polices are used for polling in Algorithm 1, which is inexpensive to implement.

In the adaptive intent polling (Algorithm 1), the pollster adapts the intent polling policies, namely, the probabilities with which the nodes at different levels in the hierarchical social influence network are polled. This affects the observation distribution $O(u)$, and hence the state estimate (see Fig.4.1).

Formulation of Intent Polling Costs

The instantaneous cost in the adaptive intent polling problem consists of two components– the measurement cost and the entropy cost (uncertainty in the state estimate):

a.) *Measurement Cost*: Let $u \in \{1, 2, \cdots, U\}$ model the choice of distributions

Algorithm 1: Adaptive Intent Polling for Pollster

1 **Polling Policy:** Compute the myopic adaptive intent polling policy $\bar{\mu}_I : \Pi(X) \to \mathcal{U}$ that maps beliefs to polling actions.

2 For an initial belief π_0, **Loop** $k = 1, 2, \cdots$:

3 **Polling Action:** Polling action $u_k = \bar{\mu}_I(\pi_{k-1})$ is a choice of a distribution $\beta^{(u_k)}$, where $\beta^{(u_k)} = (\beta_0^{(u_k)}, \beta_1^{(u_k)}, \cdots \beta_N^{(u_k)})$ and $\sum_i \beta_i^{(u_k)} = 1$ and $N + 1$ is the number of levels in the network. Poll a node at level l with a probability $\beta_l^{(u_k)}$ and ask the following question to obtain the observation y_k:

"*What does it (a node at level l) think the state is?*"

4 **State-estimation:** Estimate the state π_k using the Bayesian filtering update (4.10) with observation distribution $O(u_k) = B f_{u_k}(B)$, where B is the opinion distribution (4.4) and $f_{u_k}(z)$ is a Hurwitz polynomial.

5 **Polling Cost:** Incur an intent polling cost $C(\pi_k, u_k) = S(\beta^{(u_k)}) + \eta_e(\pi_k, u_k)$ that is composed of measurement and entropy costs respectively.

6 **End**

(polling actions) $\beta^{(u)}$, where $\beta^{(u)} = (\beta_0^{(u)}, \beta_1^{(u)}, \cdots \beta_N^{(u)})$ and $\sum_i \beta_i^{(u)} = 1$. Here $\beta_l^{(u)}$ for $l = 0, 1, 2 \cdots, N$ denotes the probability of selecting a node from level l, having an opinion distribution B^{l+1}. Let $s(l)$ denote the measurement cost from level l. Since nodes at higher levels in the hierarchy (small l) provide more informative (in the Blackwell sense) observations, higher costs are associated with obtaining observations from these levels [30], i.e, $s(l) \geq s(l+1)$, and the average measurement cost for employing the polling algorithm $\beta^{(u)}$ is $S(\beta^{(u)}) = \sum_{l=0}^{N} \beta_l^{(u)} s(l)$.

b.) *Entropy Cost*: The entropy cost models the uncertainty in the state estimate π

in (3.20), and is given as

$$\eta_e(\pi, u) = -\gamma_1(u) \sum_{i=1}^{2} \pi(i) \log_2 \pi(i) + \gamma_2(u)$$

for $\pi_k(i) \in (0, 1)$ and $\eta_e \overset{\Delta}{=} 0$ for $\pi_k(i) = \{0, 1\}$. Here $\gamma_1, \gamma_2 > 0$ are user defined scalar weights.

Since more informative opinions lead to larger reduction in uncertainty, $\gamma_1(u) > \gamma_1(u+1)$ and $\gamma_2(u+1) > \gamma_2(u)$.

The net instantaneous cost $C(\pi, u)$ incurred by the pollster in case of adaptive intent polling is thus given as:

$$C(\pi, u) = S(\beta^{(u)}) + \eta_e(\pi, u). \tag{4.17}$$

The cost (4.17) expressed in terms of the belief state π captures the fact that a control with higher measurement cost should result in a smaller entropy (more reduction in uncertainty) cost and vice versa.

Matrix polynomials and Blackwell Dominance

The aim of this section is to provide a rationale for choosing the intent polling actions (distributions) $\beta^{(u)},\ u \in \mathcal{U}$, with $f_u(z) = \sum_{l=0}^{N} \beta_l^{(u)} z^l$ denoting the polynomial associated with action distributions $\beta^{(u)}$. It is shown that when $f_u(z)$ is Hurwitz stable, there exists a Blackwell dominance relation between the observation distributions (4.3) for the pollster. Let $\mathcal{P}_N = \{h | h(z) = \sum_{i=0}^{N} \beta_i z^i,\ \sum_{i=0}^{N} \beta_i = 1,\ \beta_i \geq 0\}$ denote the collection of all polynomials with co-efficients that are a convex combination.

Proposition 4.2.1. Let Q be a stochastic matrix. For $n > m$, let $p(z) \in \mathcal{P}_n$ and $q(z) \in \mathcal{P}_m$ be two polynomials such that all the roots of $q(z)$ are roots of $p(z)$. If $q(z)$ and $p(z)$ are Hurwitz, then $q(Q) \succeq_B p(Q)$.

Proposition 4.2.1 provides an useful partial order of the observation distributions to choose a polling action. In adaptive intent polling (Theorem 4.3), the degree of the polynomial is the same as the number of levels in the hierarchy (Fig.4.1). A polling action in adaptive intent polling corresponds to choosing the (normalized) co-efficients of a polynomial, and these coefficients are the probabilities of polling from the various levels of the social network.

According to Proposition 4.2.1, if the polynomials are Hurwitz and have common factors, a Blackwell dominance relation exists between their corresponding matrix polynomials. If, however, $p(z) \in \mathcal{P}_n$ is not a Hurwitz polynomial, then $q(Q) \succeq_B p(Q)$ only if the polynomial $q(z) \in \mathcal{P}_m$ is the single quadratic factor ($m = 2$) corresponding to any conjugate pair of zeros of $p(z)$ having smallest argument in magnitude; see [15].

Myopic Policies for Adaptive Intent Polling

Our main result on adaptive intent polling is Theorem 4.3 below. It shows that when it cheaper for the pollster to (myopically) listen to the polynomial channel that provides largest reduction in uncertainty on the state, it is indeed optimal to do that. Polynomial channels are parallel cascaded channels that model the communication medium between the pollster and the nodes of a social network having a hierarchical influence structure as in Fig.4.1, when the pollster polls all levels of the hierarchical network as in intent polling.

Let $f_u(z) = \sum_{l=0}^{N} \beta_l^{(u)} z^l$ denote the polynomial corresponding to the polling policy $\beta^{(u)}$. For an opinion distribution B (defined in (4.4)), let the matrix polynomials be $f_u(B)\ \forall u \in \mathcal{U}$.

Theorem 4.3 (Adaptive Intent Polling)**.** Consider the adaptive intent polling problem with costs specified in (4.17). Let the observation distribution for the pollster be $O(u) =$

$Bf_u(B)$ $\forall u \in \mathcal{U}$. Assume that the polynomial $f_U(z) \in \mathcal{P}_N$ is Hurwitz[4].

(a) Then, $O(u) \succeq_B O(u+1)$ $\forall u \in \mathcal{U}$.

(b) By Theorem 4.1, the myopic intent polling policy $\bar{\mu}_I(\pi)$ forms an upper bound to the optimal intent polling policy $\mu_I^*(\pi)$, i.e, $\mu_I^*(\pi) \leq \bar{\mu}_I(\pi)$ for all $\pi \in \Pi$. In particular, for belief states where $\bar{\mu}_I(\pi) = 1$, the myopic policy coincides with the optimal policy $\mu_I^*(\pi)$.

The instantaneous cost for adaptive intent polling (4.17) is concave in π by definition. The proof of Theorem 4.3 follows from Proposition 4.2.1 and Theorem 4.1. The adaptive intent polling algorithm employed by the pollster determines how the opinions are gathered, and the opinions are distributed as $O(u)$ for the pollster. For an opinion distribution B, the observation distribution of the pollster in case of adaptive intent polling is given as $O(u) = Bf_u(B)$, where $f_u(B) = \sum_{l=0}^{N} \beta_l^{(u)} B^l$ and nodes at level l are sampled with probability $\beta_l^{(u)}$. The matrix polynomial $f_u(B)$ has an identity observation likelihood for the co-efficient $\beta_0^{(u)}$. This motivates the choice $O(u) = Bf_u(B)$ $\forall u \in \mathcal{U}$.

Proposition 4.2.1 provides a justification for the polynomial $f_U(z)$ to be Hurwitz. If $f_U(z)$ is Hurwitz, then a way to compute $f_g(z)$ for $g \in \{U-1, \cdots, 2, 1\}$ is by successive long-division of $f_U(z)$ by linear or quadratic factors of $f_U(z)$. If we know that the polynomial $f_U(z)$ is Hurwitz, then the polynomial $f_{U-1}(z)$ obtained by removing any linear or quadratic factor from $f_U(z)$ is also Hurwitz. From Proposition 4.2.1, we know that if two polynomials are Hurwitz, there is a Blackwell dominance relation between them. From Theorem 4.3, the observation distribution for the pollster is ordered as

$$O(U-1) = B \cdot f_{U-1}(B) \succeq_B B \cdot f_U(B) = O(U).$$

A similar procedure can be carried out to obtain observation distributions for $u = U-2, \cdots, 1$ as long as the number of levels $N+1$ is greater than U in the hierarchical social network, as there are $N+1$ roots for a polynomial of degree $N+1$.

[4]A polynomial f is Hurwitz if all its zeroes lie in the open left half-plane of the complex plane, and all its co-efficients have the same sign.

Information Theoretic Interpretation

The aim of this section is to provide an interesting link between Hurwitz stability and channel capacity in terms of Blackwell dominance. Let $I(\mathcal{X};\mathcal{Y}^{(u)})$ denote the mutual information of channel $f_u(B)$ and $C^{(u)}$ denote the capacity. Let $f_u^i(B)$ denote the i^{th} row of the matrix polynomial $f_u(B)$.

Proposition 4.2.2. If the channel error probabilities (likelihoods) for the pollster satisfy $f_u(B) \geq_B f_{u+1}(B)\ \forall u \in \mathcal{U}$, then

a.) Shannon Capacity Ordering: $C^{(u)} \geq C^{(u+1)}\ \forall u \in \mathcal{U}$.

b.) Rényi Divergence Ordering:

$$D_\alpha(f_u^i(B)\|f_u^j(B)) \geq D_\alpha(f_{u+1}^i(B)\|f_{u+1}^j(B))$$

for all $u \in \mathcal{U}$ and for all $i, j \in \mathcal{X}$.

The proof of Proposition 4.2.2 follows from Theorem A.6 and Theorem 4.2. From Proposition 4.2.2, the Hurwitz polynomial channels are ordered such that the channel that is a sub channel of the other results in a larger reduction in uncertainty on the state.

Together with Proposition 4.2.1, Proposition 4.2.2 provides an interesting link between Hurwitz (stable) polynomials and channel capacity. From Proposition 4.2.1, those polling actions that result in Hurwitz (stable) polynomials allow decomposition of channels into sub channels that have higher capacity from Proposition 4.2.2.

4.2.4 Main Result. Adaptive Expectation Polling

In expectation polling [114], to decide between two states, the sampled individuals are asked “who will your friends vote for?". In a hierarchical network, this can be seen as

asking “who will your more influential friends vote for?". In this section, we develop an adaptive version of expectation polling [114]: the resulting algorithm (Algorithm 2 below) is designed for hierarchical social influence networks with time-varying state of nature.

We present novel sufficient conditions for Blackwell dominance in the context of adaptive expectation polling. These conditions involving ultrametric matrices enable the application of Theorem 4.1 to determine myopic policies that upper bound the optimal adaptive expectation polling policy. The myopic policies are used for polling in Algorithm 2, which is inexpensive to implement.

Algorithm 2 below is a more sophisticated version of standard expectation polling, for multiple states and hierarchical social networks.

In the adaptive expectation polling (Algorithm 2), the pollster controls the observation distribution $O(u)$ by choosing different levels to gather the opinion, and this in turn affects the estimate of the state (see Fig.4.1).

Formulation of Expectation Polling Costs

The instantaneous cost in adaptive expectation polling consists of two components– the measurement cost and the uncertainty in the state estimate:

a.) *Measurement Cost*: Let $u \in \{1, 2, \cdots, U\}$ model the choice of levels. In adaptive expectation polling, unlike adaptive intent polling, not all levels are polled. The pollster selects a level l and asks the nodes at level l to provide information about the other levels. Let $S(u)$ denote the measurement cost for action u. Since more informative opinions are costlier to obtain [30], from Theorem 4.4(i)

below, $S(u) \geq S(u+1)\ \forall u \in \mathcal{U}$.

b.) *State-Estimation error*: The state-estimation error incurred in choosing action u is modelled as

$$\eta_2(\bar{x}, u) = w_u \|\bar{x} - \pi\|_2. \tag{4.18}$$

The scalar $w_u > 0$ allows the costs associated with different controls/ or the levels to be weighed differently. In (4.18), π denotes the posterior distribution updated according to (4.10) and $\bar{x} \in \{e_1, e_2, \cdots, e_X\}$, where e_i is the unit indicator vector. Note that this is an alternate representation of the state space $\mathcal{X}$. In (4.18) using the law of iterated expectation [75, Chapter 8, Sec.8.4.2],[72, Lemma 3.2], $\eta_2(\pi, u)$ can be expressed in terms of the belief π as follows[5]:

$$\eta_2(\pi, u) = w_u \left(1 - \pi'\pi\right) \tag{4.19}$$

Since more informative opinions lead to smaller state-estimation error, from Theorem 4.4(i) below, $w_{u+1} > w_u$.

The net instantaneous cost $C(\pi, u)$ in (4.11) incurred by the pollster in case of adaptive expectation polling is thus given as:

$$C(\pi, u) = S(u) + \eta_2(\pi, u) \tag{4.20}$$

The cost (4.20) expressed in terms of the belief state π models the fact that asking the nodes at level i to provide information on the opinions of nodes at levels $j(< i)$ is costly, but more informative – smaller state estimation error.

[5]

$$J_\mu(\pi_0) = \mathbb{E}_\mu\{\sum_{k=0}^{\infty} \rho^k \mathbb{E}\{C(\bar{x}_k, u_k)\,|\mathcal{I}_k\}|\pi_0\}.$$

$$J_\mu(\pi_0) = \mathbb{E}_\mu\{\sum_{k=0}^{\infty} \rho^k \sum_{i=1}^{X} C(e_i, u_k)\pi_k(i)\,|\pi_0\} \text{ from (3.20).}$$

The instantaneous cost is thus given as $\sum_{i=1}^{X} C(e_i, u_k)\pi_k(i)$, where $C(e_i, u_k) = S(u_k) + w_{u_k}\|e_i - \pi_k\|_2$. By noting that $\|e_i - \pi_k\|_2 = \sum_{m=1}^{X} |e_i(m) - \pi_k(m)|^2$, the equation (4.19) follows from (4.18) by simple algebraic manipulation.

Algorithm 2: Adaptive Expectation Polling for Pollster

1 **Polling Policy:** Compute the myopic adaptive expectation polling policy $\bar{\mu}_E : \Pi(X) \to \mathcal{U}$ that maps beliefs to polling actions.

2 For an initial belief π_0, **Loop** $k = 1, 2, \cdots$:

3 **Polling Action:** Polling action $u_k = \bar{\mu}_E(\pi_{k-1})$ is a choice of level in the network. Poll a node at level l and ask the following question to obtain the observation y_k:

"*what does a node at level l think the nodes at level $j(< l)$ would report the state as?*"

4 **State-estimation:** Estimate the state using the Bayesian filtering update (4.10) with observation distribution $O(u_k) = B_l^{l_{u_k}/l}$, where B_l is the opinion distribution (4.6) and B in (4.4) is an ultrametric matrix. Here nodes at level l are polled to provide information of the nodes at level l_{u_k}.

5 **Polling Cost:** Incur an expectation polling cost $C(\pi_k, u_k) = S(\beta^{(u_k)}) + \eta_2(\pi_k, u_k)$ that is composed of measurement and state estimation error respectively.

6 **End**

Fractional Exponents of Stochastic Matrices and Blackwell Dominance

The aim of this section is to provide a rationale for choosing the expectation polling actions, which correspond to sampling nodes at a particular level and soliciting information from other levels (see Fig.4.1). It is shown fractional matrix powers can model requesting information from hidden levels in a hierarchical social influence network. When the opinion distribution is an ultrametric matrix, there is a relation between the fractional matrix powers and Blackwell dominance.

Definition. (Ultrametric Matrix [62]) A square stochastic matrix Q is *ultrametric* if

1. Q is symmetric.
2. $Q_{ij} \geq \min\{Q_{ik}, B_{kj}\}$ for all i, j, k.
3. $Q_{ii} > \min Q_{ik}$ for all $k \neq i$.

For any ultrametric matrix Q, the K^{th} primary root, $Q^{1/K}$, is also stochastic for any positive integer K; see [62].

Proposition 4.2.3. For any ultrametric matrix Q, the following hold for any positive integer j:

a) $Q^{j/K} \succeq_B Q^j$.

b) $Q^{j/K} \succeq_B Q^{(j+1)/K} \ldots \succeq_B Q^{(j+K-1)/K}$

c) $Q^{j/(K+1)} \succeq_B Q^{j/(K)}$.

d) $Q \succeq_B Q^{j/K}$, for all $j > K$.

Clearly, any integer power of a stochastic matrix is a stochastic matrix. Proposition 4.2.3 says that fractional power of certain stochastic matrices, namely ultrametric, are also stochastic. In adaptive expectation polling (Theorem 4.4), polling actions correspond to choosing different levels in the hierarchy (Fig.4.1) and soliciting opinions of nodes at other levels. In Proposition 4.2.3, $Q^{j+1/K+1}$ can be used to interpret the notion of node at level K providing information on nodes' opinions at level j, and hence provides a way to order the likelihoods corresponding to different polling actions. According to Proposition 4.2.3, when the opinion distribution B in (4.4) is ultrametric, there exists a Blackwell dominance relation between the observation distributions of the pollster.

Myopic policies for Adaptive Expectation Polling

Our main result in adaptive expectation polling is Theorem 4.4 below. It shows that when it is cheaper for the pollster to (myopically) listen to the ultrametric channel that provides the most information on the state, it is optimal to do so. Ultrametric channels are (hidden) cascaded channels that model the communication medium between the pollster and the nodes of a social network having a hierarchical influence structure as in Fig.4.1, when the pollster seeks opinions formed at the hidden levels from the levels that are easily accessible.

Theorem 4.4 (Adaptive Expectation Polling)**.** Consider the adaptive expectation polling problem with costs specified in (4.20). Assume that the opinion distribution B (defined in (4.4)) is ultrametric. Let the observation distributions for the pollster be $O(u) = B_l^{l_u/l}\ \forall u \in \mathcal{U}$.

(a) For the choice of levels $l_u > l_{u+1}$, we have

$$O(u) \succeq_B O(u+1)\ \forall u \in \mathcal{U}.$$

(b) By Theorem 4.1, the myopic expectation polling policy $\bar{\mu}_E(\pi)$ forms an upper bound to the optimal expectation polling policy $\mu_E^*(\pi)$, i.e, $\mu_E^*(\pi) \leq \bar{\mu}_E(\pi)$ for all $\pi \in \Pi$. In particular, for belief states where $\bar{\mu}_E(\pi) = 1$, the myopic policy coincides with the optimal policy $\mu_E^*(\pi)$.

The instantaneous cost for adaptive expectation polling (4.20) is concave in π by definition. The proof of Theorem 4.4 follows from Proposition 4.2.3 below and Theorem 4.1. The expectation polling algorithm employed by the pollster determines how the opinions are gathered, and the opinions are distributed as $O(u)$ for the pollster. Proposition 4.2.3 below provides a justification for the opinion distribution B to be ultrametric. Note that B_l denotes the opinion distribution at level l, i.e, $B_l = B^{l+1}$ from Fig.4.1. For any $K > 0$,

clearly $B_K^{j+1/K+1} = B_j$. This motivates the choice of the observation distribution of the pollster in case of adaptive expectation polling as $O(u) = B_l^{l_u/l}$, where nodes at level l are polled to provide information of the nodes at level l_u. It is easiest (see Sec.4.2.7) to poll nodes at level N, so a convenient choice is $O(u) = B_{N+1}^{l_u/N+1}$.

Information Theoretic Interpretation

The aim of this section is to provide a link between ultrametric channels (hidden channels) and Shannon capacity in terms of Blackwell dominance. Let $I(\mathcal{X}; \mathcal{Y}^{(l_u)})$ denote the mutual information of the ultrametric channel $Q^{l_u/K}$ and $C^{(l_u)}$ denote the capacity defined in (A.5). Let $Q_i^{l_u/K}$ denotes the i^{th} row of the channel $Q^{l_u/K}$.

Proposition 4.2.4. If the channel error probabilities (likelihoods) for the pollster satisfy $Q^{l_u/K} \geq_B Q^{l_v/K}$ for any $K > 0$, we have

i.) Shannon Capacity Ordering: $C^{(l_u)} \geq C^{(l_v)}$ for $l_u > l_v$.

ii.) Rényi Divergence Ordering:

$$D_\alpha(Q_i^{l_u/K} \| Q_j^{l_u/K}) \geq D_\alpha(Q_i^{l_v/K} \| Q_j^{l_v/K})$$

for all $u \in \mathcal{U}$ and for all $i, j \in \mathcal{X}$.

The proof of Proposition 4.2.4 follows from Theorem A.6 and Theorem 4.2. Proposition 4.2.4 provides an ordering of Rényi Divergence and Shannon capacity between ultrametric channels $Q^{l_u/K}$, $K > 0$, $\forall u \in \mathcal{U}$. From Proposition 4.2.4, the ultrametric channels are ordered such that the information of nodes at Level 0, for example, revealed by the nodes at Level $N(\neq 0)$ result in a larger reduction in uncertainty on the state, than opinions from nodes at Level $N + 1(\neq 0)$.

4.2.5 Approximate Blackwell Dominance

So far we have discussed sufficient conditions for Blackwell dominance; when these conditions hold, the optimal adaptive polling policy is provably upper bounded by a myopic policy. *A natural question is: Can efficient polling methods be developed when Blackwell dominance does not hold exactly?*

This section discusses approximate Blackwell dominance and its applications in a novel polling method called adaptive neighborhood expectation polling. The main idea involves Le Cam deficiency.

Le Cam Deficiency

Given a collection of matrices, it is important to check whether there exists a Blackwell dominance relation, as Theorem 4.1 can used to compute inexpensive policies. In this section, an approximation procedure using *Le Cam deficiency* is provided. *Le Cam deficiency* enables to calculate the closest matrix that is Blackwell comparable.

Definition. (Le Cam deficiency) For any two stochastic matrices W and H, the *Le Cam deficiency* is

$$\delta(W, H) \stackrel{\Delta}{=} \inf_{R \in \mathcal{M}} \|W - HR\|_\infty, \tag{4.21}$$

where $\mathcal{M}$ denotes the set of all stochastic matrices and $\|\cdot\|_\infty$ denotes the induced norm.

The inf in (4.21) is achieved – this can be shown using *Le Cam randomization* criterion [123]. The Le Cam deficiency is an approximation measure that quantifies the loss when using one observation distribution instead of the other. There is no loss if there exists a mechanism able to convert the observations from one distribution to the other.

Algorithm 3: Approximate Blackwell Dominance

Let $\mathcal{M}$ denotes the set of all stochastic matrices.

Initialize: $O(1) = \hat{O}(1)$

For $u \in \{1, 2, \cdots, U-1\}$**, do**:

$R^*_{u+1} = \arg\min_{R \in \mathcal{M}} \|O(u+1) - \hat{O}(u)R\|_\infty$

$\hat{O}(u+1) = \hat{O}(u)R^*_{u+1}$

end

Output: $\hat{O}(u)$ for $u \in \mathcal{U}$.

(4.21) can be solved as a convex optimization problem using *CVXOPT* toolbox in Python or *CVX* in Matlab. Solving (4.21) yields observation distributions that are Blackwell comparable.

Consider a POMDP model $\theta = (\mathcal{X}, \mathcal{Y}, P, O(u), C, \rho)$, where $O(u)$ for $u = \{1, 2, \cdots, U\}$ are observation matrices that are not Blackwell comparable. Consider an approximation $\gamma = (\mathcal{X}, \mathcal{Y}, P, O(1), \hat{O}(\hat{u}), C, \rho)$, where $\hat{u} = \mathcal{U}/\{1\}$ and the observations distributions are such that

$$O(1) \succeq_B \hat{O}(2) \cdots \succeq_B \hat{O}(U). \tag{4.22}$$

Algorithm 3 details a procedure to compute observation distributions that share a Blackwell dominance relation (4.22).

Applications of Approximate Blackwell Dominance

Algorithm 3 can be used to design POMDPs for adaptive polling that have observation distributions that are not Blackwell comparable – for example, when the polling distributions in case of adaptive intent polling are not Hurwitz, when the opinion distributions

are not ultrametric in case of adaptive expectation polling, when the pollster has a choice between different polling algorithms over the polling horizon, etc.

1. *Adaptive Neighborhood Expectation Polling:* Here each polled node gathers the opinion from other nodes at the same level on each state and reports the opinion fraction to the pollster. The question asked by the pollster in case of adaptive NEP polling is

   "*what does a node at level l think the fraction*

   in favor of different states is, at level l?"

   This polling algorithm is a more sophisticated version of Neighborhood Expectation Polling (NEP) [95]. NEP is a polling algorithm to decide between two states where the pollster asks the following question [95]: "what is a nodes' estimate of the fraction of votes for a particular candidate?".

   In the case of adaptive NEP polling, the pollster controls the observation distribution $O(u)$ by choosing different levels to gather the information in the form of fractions, and this in turn affects the estimate of the state (see Fig.4.1).

   Remark: In case of adaptive NEP polling, the nodes report opinion fractions to the pollster. If instead, the nodes report probabilities with $\mathcal{Y} = [0,1]^{|\mathcal{X}|}$, there is a possibility that the pollster receives biased information. There is a disjunction effect – the beliefs about the state change when aggregated differently. This is the well known *Simpson's Paradox*; see [21].

   The adaptive NEP polling algorithm employed by the pollster determines how the opinions are gathered, and the observations for the pollster are tuples reported by the nodes that indicate the fraction in favor of each state. Channels specified by multinomial distributions model the likelihood of opinion counts in favor of different states from different nodes at the same level. Let $\mathcal{N} \in \{1, 2, \cdots, \mathbb{N}\}$ denote the

number of nodes accessible (friends with) to nodes at each level in the hierarchical social influence network. This models the possibility of different individuals or nodes having different friends with $\mathbb{N}$ denoting a finite maximum number. Let the observation alphabet for the pollster be $\mathcal{Y} = \{(\frac{n_1}{N}, \frac{n_2}{N}, \cdots, \frac{n_X}{N}) \; \forall \; \mathcal{N} : \boldsymbol{n}_i \in \mathbb{Z}_+, \; \sum_i \boldsymbol{n}_i = \mathcal{N}\}$, where $\mathbb{Z}_+$ denotes the set of non-negative integers. Let $O(l)$ denote the opinion fraction that the pollster receives from level l, and has elements

$$(O(l))_{ij} = \mathbb{P}(y_{k+1}^l = j | x_{k+1} = i, \mathcal{N}_j), i \in \mathcal{X}, j \in \mathcal{Y}.$$

Here,

$$j = (\frac{\boldsymbol{n}_1^{(j)}}{\mathcal{N}_j}, \frac{\boldsymbol{n}_2^{(j)}}{\mathcal{N}_j}, \cdots, \frac{\boldsymbol{n}_X^{(j)}}{\mathcal{N}_j}), \; \mathcal{N}_j \in \{1, 2, \cdots, \mathbb{N}\}, \sum_h \boldsymbol{n}_h^{(j)} = \mathcal{N}_j.$$

$$\mathbb{P}\Big(y_{k+1}^l = j | x_{k+1} = i, \mathcal{N}_j\Big) = \frac{\mathcal{N}_j!}{\boldsymbol{n}_1^{(j)}! \times \cdots \times \boldsymbol{n}_X^{(j)}!} \prod_{h=1}^{X} (B_l)_{ih}^{\boldsymbol{n}_h^{(j)}}. \tag{4.23}$$

Here $\mathcal{N}_j$ and $\boldsymbol{n}_i^{(j)}$ indicate the total and the number in favor of $x = i$ reported and B_l denotes the opinion distribution (4.6) at level l. The likelihood in (4.23) is the well known *multinomial distribution*.

The observation distributions (4.23) are not necessarily Blackwell ordered, but it is intuitive that the opinion fractions in (4.23) from nodes at level i are more informative than opinion fractions from nodes at level $j(> i)$ in Fig.4.1 owing to obvious Blackwell dominance relation of opinion distributions B_l for $l = i, j$ in (4.6). However, Algorithm 3 can be used to obtain approximate Blackwell dominance of observation distributions (4.23).

2. *Adaptive Polling with Choice:* In this section, we establish that expectation polling from the lowest level (least informative) and seeking opinions about the highest level is better (more informative) than intent polling (here, the pollster seeks information from all levels). Depending on the availability of access to different levels for the pollster, it can switch between polling algorithms.

For example, when using intent polling on an organizational network (implicitly hierarchical in nature), the executive levels might become inaccessible during IPOs or financial crisis. Then, the pollster can switch to listening the inside information from the lower levels (expectation polling), to estimate the underlying state of nature.

Let the opinion distribution B (defined in (4.4)) be ultrametric and $f_2(z) \in \mathcal{P}_N$ be *any* polynomial. Let the true POMDP model be $\theta = (\mathcal{X}, \mathcal{Y}, \mathbb{Y}, P, O(1), O(2), C)$ and the approximation be $\gamma = (\mathcal{X}, \mathcal{Y}, \mathbb{Y}, P, O(1), \hat{O}(2), C)$. Let $\mu(\cdot; \gamma)$ denote the policy parameterized by the approximate model γ.

Proposition (Adaptive Expectation v/s Intent)**.** Let $O(1) = B_{N+1}^{l_1/N+1}$, and $O(2) = Bf_2(B)$ for some l_1 and f_2, denote the observation distributions in case of adaptive expectation polling and adaptive intent polling respectively.
(i) The approximate Blackwell ordering using Algorithm 3 is

$$O(1) \succeq_B \hat{O}(2).$$

(ii) The myopic polling policy $\bar{\mu}(\pi; \gamma)$ is an upper bound to the optimal polling policy $\mu^*(\pi; \gamma)$, i.e, $\mu^*(\pi; \gamma) \leq \bar{\mu}(\pi; \gamma)$ for all $\pi \in \Pi$.

For $u = 1$, the pollster chooses expectation polling and hence listens to an ultrametric channel, and for $u = 2$, the pollster chooses intent polling and hence listens to a polynomial channel. As $O(2) = Bf_2(B)$, we have $B \succeq_B O(2)$. Note that since $O(1) = B_{N+1}^{l_1/N+1}$, when $l_1 = 1$ (nodes at Level N are polled to provide opinion of nodes at Level 0), $O(1) = B_{N+1}^{1/N+1} = B \succeq_B O(2)$. This implies that expectation polling is more informative than intent polling.
For $l_1 > 1$, there is no apparent comparison of ultrametric and polynomial channels. However, Algorithm 3 can be used to design POMDPs for adaptive polling for arbitrary l_1 and f_2.

4.2.6 Performance Bounds and Ordinal Sensitivity

In Sec.4.2.5, we discussed an approximation procedure to compute a POMDP model for an adaptive polling problem that has a Blackwell dominance structure and is close (Le Cam sense) to the true POMDP. Sec.4.2.6 provides performance bounds on the comparison of POMDPs for adaptive polling.

Sec.4.2.6 provides the ordinal sensitivity in polling, i.e, an ordering of the cumulative costs with respect to the variation in opinion distributions B (defined in (4.4)).

Performance Bounds on Adaptive Polling

Let $\theta = (\mathcal{X}, \mathcal{Y}, \mathbb{Y}, P, O(u), C, \rho)$ denote the given POMDP model for adaptive polling and $\gamma = (\mathcal{X}, \mathcal{Y}, \mathbb{Y}, P, \hat{O}(u), C, \rho)$ denote the POMDP model for adaptive polling having a Blackwell dominance relation between the observation distributions. Let $J_{\mu^*(\gamma)}(\pi; \theta)$ and $J_{\mu^*(\gamma)}(\pi; \gamma)$ be defined as in (4.12), and denote the cumulative costs incurred by the two models θ and γ respectively, when using the polling policy $\mu^*(\gamma)$. Let $J_{\mu^*(\theta)}(\pi; \theta)$ and $J_{\mu^*(\theta)}(\pi; \gamma)$ be defined as in (4.12), and denote the cumulative costs incurred by the two models θ and γ respectively, when using the polling policy $\mu^*(\theta)$. Theorem 4.5 below provides a bound on the deviations from the optimal cost and policy performance of the POMDP models for adaptive polling.

Theorem 4.5. Consider two POMDP models $\theta = (\mathcal{X}, \mathcal{Y}, \mathbb{Y}, P, O(u), C, \rho)$ and $\gamma = (\mathcal{X}, \mathcal{Y}, \mathbb{Y}, P, \hat{O}(u), C, \rho)$ for adaptive polling. Then for the mis-specified model and mis-specified policy, the following sensitivity bounds hold:

$$\text{Mis-specified Model: } \sup_{\pi \in \Pi} |J_{\mu^*(\gamma)}(\pi; \gamma) - J_{\mu^*(\gamma)}(\pi; \theta)| \leq G\|\gamma - \theta\|. \tag{4.24}$$

$$\text{Mis-specified Policy: } J_{\mu^*(\gamma)}(\pi; \theta) \leq J_{\mu^*(\theta)}(\pi; \theta) + 2G\|\gamma - \theta\|. \tag{4.25}$$

Here $G = \max_{i \in \mathcal{X}, u} \frac{C(e_i, u)}{1-\rho}$ and e_i denotes the indicator vector with a '1' in the i^{th} position, and

$$\|\gamma - \theta\| = \max_u \max_i \sum_y \sum_j P_{ij} |O_{jy}(u) - \hat{O}_{jy}(u)|.$$

Theorem 4.5 provides uniform bounds on the additional cost incurred for using parameters that are Blackwell comparable in place of the given parameters of the POMDP for adaptive polling. The proof follows from arguments similar to [75, Theorem 14.9.1], and is omitted.

So far it was assumed that the pollster has complete knowledge of the node-level associations. However, if a set of nodes are misclassified to a different level by the pollster, then the pollster is essentially updating the belief using different observation distributions. Theorem 4.5 can be used to compute the performance bounds for this misclassification as well.

Ordering of Hierarchical Social Influence Networks

So far we have discussed two types of polling algorithms on a single hierarchical social influence network. In this section, we briefly discuss how to order hierarchical influence networks that differ in the opinion distributions B, according to the expected polling cost. Theorem 4.6 below shows that some networks are inherently more expensive to poll than others; it defines a partial order over networks that results in an ordering of the cost of polling.

Let the POMDP model of the hierarchical influence network $\mathbb{H}_i$ for $i = 1, 2, \cdots$ be θ_i, where the tuple $\theta_i = (\mathcal{X}, \mathcal{Y}, \mathbb{Y}, P, O^{(i)}, C)$. Let $\mu_i^*(\pi; \theta_i)$ denote the optimal polling policy on each of the network, and let $J_{\mu_i^*(\theta_i)}(\pi; \theta_i)$ denote the corresponding optimal cumulative cost.

Theorem 4.6 (Ordinal sensitivity in Polling)**.** Consider two hierarchical networks $\mathbb{H}_1$ and $\mathbb{H}_2$. Let the POMDPs for adaptive polling of each hierarchical network have the observation distributions that satisfy $O^{(1)} \succeq_B O^{(2)}$. Then

$$J_{\mu_1^*(\theta_1)}(\pi; \theta_1) \leq J_{\mu_2^*(\theta_2)}(\pi; \theta_2). \tag{4.26}$$

Here $O^{(1)} \succeq_B O^{(2)}$ denotes $O^{(1)}(u) \succeq_B O^{(2)}(u)\ \forall\ u \in \mathcal{U}$.

The proof of Theorem 4.6 follows from arguments similar to Theorem 14.8.1 in [75], and is omitted. Since the observation likelihood for the pollster ($O^{(i)}\ \forall i$) depends on the opinion distribution (4.6), Theorem 4.6 provides a way to compare the cumulative costs of hierarchical influence networks with different opinion distributions. The result is useful, in that, a hierarchical influence network that has more informative opinion distribution at every level compared to another hierarchical influence network is cheaper to poll on average as the nodes provide more informative opinions.

4.2.7 Performance Evaluation

The main results of this chapter involve using Blackwell dominance to construct myopic policies that provably upper bound the optimal adaptive polling policy. In this section, the performance of this myopic upper bound is illustrated using numerical examples for adaptive polling. As discussed in Sec.4.2.1, the discount factor ρ determines the way the polling cost is counted towards the polling value $J_{\mu^*}(\pi_0)$ defined in (3.31) when using the optimal policy $\mu^*(\pi)$. Since the computationally inexpensive myopic policy is used for polling in Algorithm 1 and Algorithm 2, instead of the optimal policy $\mu^*(\pi)$, the performance loss and sensitivity (both defined below) in terms of the polling value $J_{\mu^*}(\pi_0)$ is evaluated for different values of the discount factor.

Let $J_{\bar{\mu}}(\pi_0)$ denote the discounted costs associated with the myopic policy $\bar{\mu}(\pi)$. We consider the following two measures for measuring the effectiveness of the myopic polling policy:

(i) The percentage loss in optimality due to using the myopic policy $\bar{\mu}$ instead of optimal policy μ^* is

$$\mathcal{L}_1 = \frac{J_{\bar{\mu}}(\pi_0) - J_{\mu^*}(\pi_0)}{J_{\mu^*}(\pi_0)}. \tag{4.27}$$

In (4.27), the total average cost is evaluated using 1000 Monte carlo simulations over a horizon of 100 time units. The optimal cost $J_{\mu^*}(\pi_0)$ is calculated as in (3.31).

(ii) Let Π_1^s represent the set of belief states for which $C(\pi, 1) < C(\pi, u)\ \forall u = 2, \cdots, U$. So on the set Π_1^s, the myopic policy coincides with the optimal policy $\mu^*(\pi)$. What is the performance loss outside the set Π_1^s? Define the following discounted cost

$$\tilde{J}_{\mu^*}(\pi_0) = \mathbb{E}\left\{\sum_{k=1}^{\infty} \rho^{k-1} \tilde{C}\left(\pi_k, \mu^*(\pi_k)\right)\right\}, \text{ where } \rho \in [0, 1),$$

$$\tilde{C}\left(\pi, \mu^*(\pi)\right) = \begin{cases} C\left(\pi, 1\right) & \pi \in \Pi_1^s \\ C(\pi, 1) + w_2 \eta_2(\pi, 2) & \pi \notin \Pi_1^s \end{cases}$$

Clearly a lower bound for the percentage loss in optimality due to using the myopic policy $\bar{\mu}$ instead of optimal policy μ^* is

$$\mathcal{L}_2 = \frac{J_{\bar{\mu}}(\pi_0) - \tilde{J}_{\mu^*}(\pi_0)}{\tilde{J}_{\mu^*}(\pi_0)}. \tag{4.28}$$

In (4.28), the cumulative discounted cost is evaluated using 1000 Monte carlo simulations over a horizon of 100 time units.

Here μ^* is the optimal policy of the non-standard (non-linear cost) POMDP, and is solved using POMDP algorithms in [75, Chapter 8, Sec.8.4.4].

4.2.8 Conclusions

We considered the problem of adaptive (stochastic feedback control based) polling in hierarchical social networks, and the problem was formulated as a partially observed Markov decision process (POMDP). POMDPs are intractable to solve. The key idea was to exploit Blackwell dominance to construct myopic bounds that provably upper bound the optimal polling policy. The following conclusions were drawn:

1. Blackwell dominance was extended to the case of polynomial observation likelihoods (channels) described by matrix polynomials. This can be used to develop an adaptive intent polling algorithm that is inexpensive to implement.
2. Blackwell dominance was extended to the case of ultrametric observation likelihoods (channels) described by fractional matrix powers. This can be used to develop an adaptive expectation polling algorithm that is inexpensive to implement.
3. Information theoretic consequences of Blackwell dominance, namely Rényi Divergence and Shannon capacity, order the observation channels by their ability to distinguish the states, and hence can guide the choice of observation distributions for the pollster.
4. It was shown that Le Cam deficiency can be used to derive general inexpensive polling algorithms, as Blackwell dominance is only a partial order.

4.2.9 Appendix A: Proofs

Proof of Theorem 4.1:

Denote by $y^{(u)}$ as the observations recorded when using action u. Then $O(u+1) = O(u)R$

implies the following

$$\mathbb{P}\left(y^{(u+1)}|x\right) = \sum_{y^{(u)}} \mathbb{P}\left(y^{(u+1)}|y^{(u)}\right)\mathbb{P}\left(y^{(u)}|x\right) \tag{4.29}$$

For notational convenience, let $T(\pi, y, u)$ be written as $T(\pi, y^{(u)} = y)$. Observe that,

$$T\left(\pi, y^{(u+1)} = y\right) = \frac{O_{u+1}(y)P'\pi}{\sigma\left(\pi, y^{(u+1)} = y\right)} = \sum_r \Lambda(r)T(\pi, y^{(u)} = r) \tag{4.30}$$

where $\Lambda(r)$ is a probability mass function w.r.t r and defined as

$$\Lambda(r) = \mathbb{P}\left(y^{(u+1)} = y|y^{(u)} = r\right)\frac{\sigma\left(\pi, y^{(u)} = r\right)}{\sigma\left(\pi, y^{(u+1)} = y\right)} \tag{4.31}$$

The following inequality follows from the concavity of $V(\pi)$ and (4.31)

$$\begin{aligned} V\left(T\left(\pi, y^{(u+1)} = y\right)\right) &= V\left(\sum_r \Lambda(r)T(\pi, y^{(u)} = r)\right) \\ V\left(T\left(\pi, y^{(u+1)} = y\right)\right) &\geq \sum_r \Lambda(r)V\left(T(\pi, y^{(u)} = r)\right) \end{aligned} \tag{4.32}$$

Following completes the proof of Theorem 4.1 using (4.32).

$$\begin{aligned} \sum_y \sigma(\pi, y^{(u+1)} = y)V\left(T\left(\pi, y^{(u+1)} = y\right)\right) &\geq \sum_y \sum_r \Lambda(r)V\left(T(\pi, y^{(u)} = r)\right)\sigma(\pi, y^{(u+1)} = y) \\ &= \sum_r V\left(T\left(\pi, y^{(u)} = r\right)\right)\sigma\left(\pi, y^{(u)} = r\right) \end{aligned} \tag{4.33}$$

$\therefore C(\pi, 1) \leq C(\pi, u)\ \forall u \Rightarrow \mu^*(\pi) = 1 \Rightarrow \mu^*(\pi) \leq \bar{\mu}(\pi)$. □

Proof of Theorem 4.2:

Let $O(u) \geq_B O(u+1)$ for $u \in \mathcal{U}$. From the definition of Rényi Divergence (4.15) we have [94]:

$$D_\alpha(O_i(u+1)||O_j(u+1)) \leq \min\left\{(1-\alpha)D(O_i(u+1)||O_j(u+1)), \alpha D(O_j(u+1)||O_i(u+1))\right\}. \tag{4.34}$$

We know that [115]:

$$O(u) \;\geq_B\; O(u+1) \;\Rightarrow\; D(O_i(u)\|O_j(u)) \;\geq\; D(O_i(u+1)\|O_j(u+1)), \quad (4.35)$$

for all $i, j \in \mathcal{X}$. From (4.34) and (4.35), the result follows. □

Proof of Proposition 4.2.1:

It is given that $p(z) \in \mathcal{P}_n$ and $q(z) \in \mathcal{P}_m$, with $n > m$. Clearly, $f(Q)$ and $g(Q)$ are stochastic matrices. Further, if the quotient polynomial $h(z) = \frac{f(z)}{g(z)} \in \mathcal{P}_{(n-m)}$, then it is easily seen that $g(Q) \geq_B f(Q)$.

Since the polynomials $p(z)$ and $q(z)$ are Hurwitz, the quotient polynomial $h(z) = \frac{p(z)}{q(z)} = \sum_{i=0}^{(n-m)} \alpha_i z^i$ has positive co-efficients; i.e $\alpha_i > 0$. It suffices to prove that $h(z) \in \mathcal{P}_{(n-m)}$. It is clear that $p(1) = q(1) = 1$, which implies that $h(1) = 1$; i.e, $\sum_{i=0}^{(n-m)} \alpha_i = 1$. □

Proof of Proposition 4.2.3:

We will only prove Theorem 4.2.3b and Theorem 4.2.3c.

For Theorem 4.2.3b, we have $Q^{(j+J)/K} = Q^{j/K} \times Q^{J/K}$. Therefore $Q^{j/K} \geq_B Q^{(j+J)/K}$.

For Theorem 4.2.3c, we have $Q^{j/K} = Q^{j/K+1} \times Q^{j/K(K+1)}$. Therefore $Q^{j/K} \geq_B Q^{j/K+1}$. □

4.2.10 Appendix B: EM Algorithm with Ultrametric Constraints

The parameters of the POMDP are computed using a sequence of observations obtained from level N in Fig.4.1. Specifically, we describe a modified version of the EM algorithm [43] is used to compute the maximum likelihood estimate of the tuple (P, B_{N+1}), where B_{N+1} is restricted to the space of ultrametric stochastic matrices. The opinion probability matrices at all other levels are computed by taking fractional exponents of B_{N+1}. In the modified EM algorithm, computing B_{N+1} requires maximizing an auxiliary likelihood function (of observation sequences) subject to ultrametric constraints on

B_{N+1}. However, the space of ultrametric stochastic matrices is non-convex because of constraint $B_{N+1}(i, j) \geq \min\{B_{N+1}(i, k), B_{N+1}(k, j)\}$ and thus computationally intractable. The following reformulation based on the Big-M method in linear programming [56] is used to deal with the non-convex constraint. For all $i, j, k \in \mathcal{X}, i \neq j \neq k$:

$$B_{N+1}(i, j) \geq B_{N+1}(i, k) + M(1 - \kappa), \tag{4.36}$$

$$B_{N+1}(i, j) \geq B_{N+1}(k, j) + M\kappa, \tag{4.37}$$

$$B_{N+1}(k, j) \geq B_{N+1}(i, k) + M(1 - \kappa), \tag{4.38}$$

$$B_{N+1}(i, k) \geq B_{N+1}(k, j) + M\kappa, \tag{4.39}$$

$$\kappa \geq 0, \tag{4.40}$$

$$-\kappa \geq -1, \tag{4.41}$$

for some large positive value M. The resulting observation likelihood B_{N+1} is a stochastic and ultrametric matrix.

CHAPTER 5

CONCLUDING REMARKS

We provided a POMDP framework for sequential decision making with social sensors. The key unifying theme of this thesis was to obtain structural results for people-centric stochastic control problems, i.e, derive mathematical insights about the nature of interaction between a controller and the social network. The main conclusions drawn from the thesis are summarized below:

1. People can act as sensors providing useful source of information. Models in social psychology and microeconomics are an useful idealization for studying the behavior of people in social and economic situations.
2. The process of influence and learning in people is in general in-efficient, i.e, the social network might not aggregate the information on the parameters that drive the public choices or opinions. However, the controller can be suitably designed to intervene in the decision making process of the people to aggregate the information efficiently.
3. The interaction between a controller and the network of social sensors (people) is highly non-trivial (as opposed to physical sensors), and it exhibits several unusual results:
   (a) The stopping sets in Bayesian quickest change detection problems using risk-averse social sensors are non-convex. This has an effect on the confidence of the controller implementing the change detection.
   (b) The time-path of the controller interventions with the social network, viewed as a stochastic process, is either a sub- or super-martingale. This can be used to check the reliability of the controller performing the interventions–

the optimal time sequence of interventions should increase or decrease on average, and not both simultaneously.

(c) The information sensing medium (or channels) between the controller and the social network exhibits stochastic ordering relations that can be exploited to devise inexpensive sensing schemes. The information theoretic consequences can be exploited to derive the error rates and can in turn inform the choice of the sensing scheme.

The people-centric decision analysis approach using POMDP provides a testable framework to analyze sequential decision making problems under uncertainty, when social network is used as an information network. Since the models used are well studied and empirically validated, the mathematical insights obtained can be straightforwardly used to inform decision making in real-world applications.

Future Directions

1. This work mainly considered ideas from microeconomics to model the decision making behavior in people in social and economic situations. However, these models are not sufficient to capture the entirety of human decision making. Models from behavioral economics [63, 127, 28, 64, 65] have recently gained popularity for being able to better explain human decision making. Also, the way individuals form an opinion is modeled in behavioral economics using communicative rationality [58] and motivated reasoning [79]. It would be natural extension of the framework to incorporate these models from behavioral economics.
2. This work considered decision making with complete knowledge of the model parameters, which were assumed to be estimated using statistical signal process-

ing techniques. It will of interest to derive policies for the controller using a model-free approach using reinforcement learning techniques [124] by incorporating additional knowledge of human decision making behavior.

APPENDIX A

PRELIMINARIES

Let $\Pi(X) \stackrel{\Delta}{=} \{\pi \in \mathbb{R}^X : \mathbf{1}_X'\pi = 1, 0 \leq \pi(i) \leq 1 \text{ for all } i \in \mathcal{X}\}$ denote the belief space or the $X-1$ dimensional unit simplex.

Definition 1. *MLR Ordering* [92] ($\geq_r$): Let $\pi_1, \pi_2 \in \Pi(X)$ be any two belief state vectors. Then $\pi_1 \geq_r \pi_2$ if

$$\pi_1(i)\pi_2(j) \leq \pi_2(i)\pi_1(j),\ i < j, i, j \in \{1, \ldots, X\}.$$

Definition 2. *First-Order Stochastic Dominance* ($\geq_s$): Let $\pi_1, \pi_2 \in \Pi(X)$ be any two belief state vectors. Then $\pi_1 \geq_s \pi_2$ if

$$\sum_{i=j}^{X} \pi_1(i) \geq \sum_{i=j}^{X} \pi_2(i) \text{ for } j \in \{1, \ldots, X\}.$$

Lemma A.1. [92] $\pi_2 \geq_s \pi_1$ iff for all $v \in \mathcal{V}$, $v'\pi_2 \leq v'\pi_1$, where $\mathcal{V}$ denotes the space of X- dimensional vectors v, with non-increasing components, i.e, $v_1 \geq v_2 \geq \ldots v_X$.

Lemma A.2. [92] $\pi_2 \geq_s \pi_1$ iff for all $v \in \mathcal{V}$, $v'\pi_2 \geq v'\pi_1$, where $\mathcal{V}$ denotes the space of X- dimensional vectors v, with non-decreasing components, i.e, $v_1 \leq v_2 \leq \ldots v_X$.

Definition 3. *Submodular function* [125]: A function $f : \Pi(X) \times \{1, 2\} \to \mathbb{R}$ is submodular if $f(\pi, u) - f(\pi, \bar{u}) \leq f(\bar{\pi}, u) - f(\bar{\pi}, \bar{u})$, for $\bar{u} \leq u, \pi \geq_r \bar{\pi}$.

Definition 4. *Single Crossing Condition* [125]: A function $g : \mathcal{Y} \times \mathcal{A} \to \mathbb{R}$ satisfies a single crossing condition in (y, a) if

$$g(y, a) - g(y, \bar{a}) \geq 0 \Rightarrow g(\bar{y}, a) - g(\bar{y}, \bar{a}) \geq 0$$

for $\bar{a} > a$ and $\bar{y} > y$. For any such function g ,

$$a^*(y) = \underset{a}{\text{argmin}}\ g(y, a) \text{ is increasing in } y. \tag{A.1}$$

Theorem A.3. [125] If $f : \Pi(X) \times \{1,2\} \to \mathbb{R}$ is sub-modular, then there exists a $u^*(\pi) = \underset{u\in\{1,2\}}{\operatorname{argmin}} f(\pi, u)$ satisfying,

$$\bar{\pi} \geq_r \pi \Rightarrow u^*(\pi) \leq u^*(\bar{\pi})$$

Definition 5. (Martingale [48]): Let $\mathcal{F}_k$ denote the sigma algebra. A sequence $\{X_k\}$ such that $\mathbb{E}[|X_k|] < \infty$ is a martingale (with respect to $\mathcal{F}_k$) if

$$\mathbb{E}[X_{k+1}|\mathcal{F}_k] = X_k, \text{ for all } k.$$

If $\mathbb{E}[X_{k+1}|\mathcal{F}_k] \geq X_k, \text{ for all } k.$, the sequence $\{X_k\}$ is a *sub-martingale*.

Definition 6. ([48]) A sequence H_k is said to be a predictable sequence if $H_k \in \mathcal{F}_{k-1}$.

In words, H_k may be predicted with certainty using the information available at time $k-1$.

Lemma A.4 ([75])**.** Under (A1), we have $\sigma(\pi_1, a) \geq_s \sigma(\pi_2, a)$, where $\sigma(\pi, a) = \begin{bmatrix} \mathbf{1}' B^{\pi}_{y=1}\pi \\ \mathbf{1}' B^{\pi}_{y=2}\pi \end{bmatrix}$.

Theorem A.5 ([48])**.** Let W_k be a sub-martingale. If $H_k \geq 0$ is predictable and each H_k is bounded, then $(H.W)_k$ is a sub-martingale.

Theorem A.5 corresponds to Theorem 5.2.5 in [48].

Definition 7 (Blackwell Dominance [20, 75])**.** A stochastic[1] matrix $B(1) \in \mathbb{P}(\mathcal{Y}^{(1)}|\mathcal{X})$ Blackwell dominates (more informative) another stochastic matrix $B(2) \in \mathbb{P}(\mathcal{Y}^{(2)}|\mathcal{X})$ written as $B(1) \geq_B B(2)$, if

$$B(2) = B(1)R, \text{ for any stochastic matrix } R. \tag{A.2}$$

[1] A $\mathcal{X} \times \mathcal{Y}$ matrix B is (row) stochastic if $\sum_j B_{ij} = 1$ for all $i \in \mathcal{X}$, $j \in \mathcal{Y}$, and $B_{ij} \in [0,1]$.

Blackwell dominance also has an information theoretic consequence: Consider the classic Discrete Memoryless Channel (DMC) [41] with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ denoted as $\mathbb{P}(\mathcal{Y}|\mathcal{X})$. Let $I(\mathcal{X};\mathcal{Y})$ denote the mutual information of the DMC. The post-processing of channel $B(1)$ in (A.2) is written as $\mathcal{X} \to \mathcal{Y}^{(1)} \to \mathcal{Y}^{(2)}$. Then from Data Processing Inequality [41], it follows that

$$B(1) \succeq_B B(2) \Rightarrow I(\mathcal{X};\mathcal{Y}^{(1)}) \geq I(\mathcal{X};\mathcal{Y}^{(2)}). \tag{A.3}$$

Theorem A.6 below provides a relation between Blackwell Dominance and Shannon capacity.

Theorem A.6 ([39, 108, 109])**.** For any two conditional distributions $B(1) \in \mathbb{P}(\mathcal{Y}^{(1)}|\mathcal{X})$ and $B(2) \in \mathbb{P}(\mathcal{Y}^{(2)}|\mathcal{X})$,

$$B(1) \succeq_B B(2) \Rightarrow C^{(1)} \geq C^{(2)}, \tag{A.4}$$

where the Shannon capacity $C^{(i)}$ of a DMC is defined as

$$C^{(i)} = \sup_{p_{\mathcal{X}}(x)} I(\mathcal{X};\mathcal{Y}^{(i)}), \quad i = 1, 2. \tag{A.5}$$

Here $p_{\mathcal{X}}(x)$ is the marginal distribution over the input alphabet $\mathcal{X}$.

APPENDIX B

TRACKING INFECTION DIFFUSION IN SOCIAL NETWORKS

Below we discuss the main results in the following paper.

- Krishnamurthy, V., **Bhatt, S.** and Pedersen, T. *Tracking infection diffusion in social networks: Filtering algorithms and threshold bounds.* IEEE Transactions on Signal and Information Processing over Networks, 3(2), pp.298-315, 2017.

This paper deals with the statistical signal processing over graphs for tracking infection diffusion in social networks. Infection (or Information) diffusion is modeled using the Susceptible-Infected-Susceptible (SIS) model. Mean field approximation is employed to approximate the discrete valued infection dynamics by a deterministic ordinary differential equation, thereby yielding a generative model for the infection diffusion. The infection is shown to follow polynomial dynamics and is estimated using an exact non-linear Bayesian filter. We compute posterior Cramér-Rao bounds to obtain the fundamental limits of the filter which depend on the structure of the network. Considering the time-varying nature of the real world networks, a filtering algorithm for estimating the degree distribution is investigated using generative models for real world networks.

APPENDIX C

MULTIPLE STOPPING TIME POMDP

Below we discuss the main results in the following paper.

- Krishnamurthy, V., Aprem, A. and **Bhatt, S.** *Multiple stopping time POMDPs: Structural results & application in interactive advertising on social media.* Automatica, 95, pp.385-398, 2018.

This paper considers a multiple stopping time problem for a Markov chain observed in noise, where a decision maker chooses at most L stopping times to maximize a cumulative objective. We formulate the problem as a Partially Observed Markov Decision Process (POMDP) and derive structural results for the optimal multiple stopping policy. The main results are as follows: i) The optimal multiple stopping policy is shown to be characterized by threshold curves Γ_l, for $l = 1, \cdots, L$, in the unit simplex of Bayesian Posteriors. ii) The stopping sets S^l (defined by the threshold curves Γ_l) are shown to exhibit the following nested structure $\mathcal{S}^{l-1} \subset \mathcal{S}^l$. iii) The optimal cumulative reward is shown to be monotone with respect to the copositive ordering of the transition matrix. iv) A stochastic gradient algorithm is provided for estimating linear threshold policies by exploiting the structural results. These linear threshold policies approximate the threshold curves Γ_l, and share the monotone structure of the optimal multiple stopping policy.

APPENDIX D

POLICY GRADIENT USING WEAK DERIVATIVES FOR REINFORCEMENT LEARNING

Below we discuss the main results in the following paper.

- **Bhatt, S.**, Koppel, A., and Krishnamurthy, V. *Policy Gradient using Weak Derivatives for Reinforcement Learning.* (submitted) Conference on Decision and Control (CDC) 2019.

This paper considers policy search in continuous state-action reinforcement learning problems. Typically, one computes search directions using a classic expression for the policy gradient called the Policy Gradient Theorem, which decomposes the gradient of the value function into two factors: the score function and the $Q-$function. This paper presents four results: (i) an alternative policy gradient theorem using weak (measure-valued) derivatives instead of score-function is established; (ii) the stochastic gradient estimates thus derived are shown to be unbiased and to yield algorithms that converge almost surely to stationary points of the non-convex value function of the reinforcement learning problem; (iii) the sample complexity of the algorithm is derived and is shown to be $O(1/\sqrt{k})$; (iv) finally, the expected variance of the gradient estimates obtained using weak derivatives is shown to be lower than those obtained using the popular score-function approach. Experiments on OpenAI gym pendulum environment show superior performance of the proposed algorithm.

APPENDIX E

EFFICIENT POLLING ALGORITHMS USING FRIENDSHIP PARADOX

Below we discuss the main results in the following paper.

- **Bhatt, S.**, Nettasinghe, B., and Krishnamurthy, V. *Efficient Polling Algorithms using Friendship Paradox and Blackwell Dominance.* (submitted) FUSION 2019.

This paper develops efficient polling algorithms that take into account the influence structure of the social network and can track a time-varying fraction of the population having a particular attribute. The proposed algorithms belong to the class of Neighborhood Expectation Polling (NEP) algorithms with two key differences in the sampling mechanism: (i) Friendship Paradox based sampling, (ii) Blackwell Dominance based sampling. NEP algorithm based on Friendship Paradox is provably more efficient, in terms of smaller mean square error, compared to the well known intent polling algorithm. NEP algorithm with Blackwell dominance is developed using a partially observed Markov decision process (POMDP) framework, and is computationally inexpensive to implement.

BIBLIOGRAPHY

[1] Daron Acemoglu, Munther A Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

[2] Lada A. Adamic and Bernardo A. Huberman. Power-law distribution of the world wide web. *Science*, 287(5461):2115–2115, 2000.

[3] Ali N Akansu and Mustafa U Torun. *A Primer for Financial Engineering: Financial Signal Processing and Electronic Trading*. Academic Press, 2015.

[4] Franklin Allen and Douglas Gale. Stock-price manipulation. *The Review of Financial Studies*, 5(3):503–529, 1992.

[5] Juan A Almendral, Luis López, and Miguel AF Sanjuán. Information flow in generalized hierarchical networks. *Physica A: Statistical Mechanics and its Applications*, 324(1):424–429, 2003.

[6] Fredrik Andersson, Helmut Mausser, Dan Rosen, and Stanislav Uryasev. Credit risk optimization with conditional value-at-risk criterion. *Mathematical Programming*, 89(2):273–291, 2001.

[7] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[8] Karl J Åström. *Introduction to stochastic control theory*. Courier Corporation, 2012.

[9] Marco Avellaneda and Sasha Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.

[10] Yossi Aviv and Amit Pazgal. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management*, 10(3):339–359, 2008.

[11] Venkatesh Bala and Sanjeev Goyal. Learning from neighbours. *The review of economic studies*, 65(3):595–621, 1998.

[12] Abhijit V Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, Aug., 1992.

[13] Albert László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[14] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: The topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1):69–77, 2000.

[15] RW Barnard, W Dayawansa, K Pearce, and D Weinberg. Polynomials with nonnegative coefficients. *Proceedings of the American Mathematical Society*, 113(1):77–85, 1991.

[16] Roy F Baumeister and Eli J Finkel. *Advanced social psychology: The state of the science*. OUP USA, 2010.

[17] Dimitir P Bertsekas and Steven Shreve. *Stochastic optimal control: the discrete-time case*. 2004.

[18] Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 1995.

[19] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, Oct., 1992.

[20] David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pages 265–272, 1953.

[21] Colin R Blyth. On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67(338):364–366, 1972.

[22] Javier Borge-Holthoefer, Alejandro Rivero, Iñigo García, Elisa Cauhé, Alfredo Ferrer, Darío Ferrer, David Francos, David Iniguez, María Pilar Pérez, Gonzalo Ruiz, et al. Structural and dynamical patterns on online social networks: The Spanish May 15th movement as a case study. *PloS one*, 6(8):e23883, 2011.

[23] Subir Bose, Gerhard Orosel, Marco Ottaviani, and Lise Vesterlund. Dynamic monopoly pricing and herding. *The RAND Journal of Economics*, 37(4):910–928, 2006.

[24] Subir Bose, Gerhard Orosel, Marco Ottaviani, and Lise Vesterlund. Monopoly pricing in the binary herding model. *Economic Theory*, 37(2):203–241, 2008.

[25] Christina L Boyd. The hierarchical influence of courts of appeals on district courts. *The Journal of Legal Studies*, 44(1):113–141, 2015.

[26] Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, and Matthew Williams. 140 characters to victory?: Using twitter to predict the uk 2015 general election. *Electoral Studies*, 41:230–233, 2016.

[27] Colin F Camerer. Can asset markets be manipulated? a field experiment with racetrack betting. *Journal of Political Economy*, 106(3):457–482, 1998.

[28] Colin F Camerer, George Loewenstein, and Matthew Rabin. *Advances in behavioral economics*. Princeton university press, 2011.

[29] James V Candy. *Bayesian signal processing: classical, modern, and particle filtering methods*, volume 54. John Wiley & Sons, 2016.

[30] Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203, 2015.

[31] Alvaro Cartea and Sebastian Jaimungal. Modelling asset prices for algorithmic and high-frequency trading. *Applied Mathematical Finance*, 20(6):512–547, 2013.

[32] Archishman Chakraborty and Bilge Yılmaz. Manipulation in market order models. *Journal of financial Markets*, 7(2):187–206, 2004.

[33] Christophe Chamley. *Rational herds: Economic models of social learning*. Cambridge University Press, 2004.

[34] Klarissa TT Chang, Wen Chen, and Bernard CY Tan. Advertising effectiveness in social networking sites: Social ties, expertise, and product type. *IEEE Transactions on engineering management*, 59(4):634–643, 2012.

[35] Shimin Chen, Zheng Sun, Song Tang, and Donghui Wu. Government intervention and investment efficiency: Evidence from china. *Journal of Corporate Finance*, 17(2):259–271, 2011.

[36] Yiling Chen and Ian A Kash. Information elicitation for decision making. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 175–182. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[37] Thierry Chonavel. *Statistical signal processing: modelling and estimation*. Springer Science & Business Media, 2002.

[38] Fan Chung and Linyuan Lu. *Complex Graphs and Networks*, volume 107. American Mathematical Society, 2006.

[39] JOEL Cohen, JHB Kempermann, and Gheorghe Zbaganu. *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population*. Springer Science & Business Media, 1998.

[40] Richard A. Cohn, Wilbur G. Lewellen, Ronald C. Lease, and Gary G. Schlarbaum. Individual investor risk aversion and investment portfolio composition. *The Journal of Finance*, 30(2):605–620, May, 1975.

[41] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[42] Anirban Dasgupta, Ravi Kumar, and D Sivakumar. Social sampling. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 235–243. ACM, 2012.

[43] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, pages 1–38, 1977.

[44] Arnoud V den Boer. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18, 2015.

[45] Persi Diaconis and David Freedman. On the consistency of Bayes estimates. *The Annals of Statistics*, pages 1–26, 1986.

[46] Bas Donkers and Arthur Van Soest. Subjective measures of household preferences and financial decisions. *Journal of Economic Psychology*, 20(6):613 – 642, 1999.

[47] Séverine Dubuisson. *Tracking with particle filter for high-dimensional observation and state spaces*. John Wiley & Sons, 2015.

[48] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.

[49] Yiyong Feng, F. Rubio, and D.P. Palomar. Optimal order execution for algorithmic trading: A CVaR approach. In *IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 480–484, June 2012.

[50] Chaim Fershtman and Uzi Segal. Preferences and social influence. *American Economic Journal: Microeconomics*, 10(3):124–42, 2018.

[51] Marianne Frisén. Optimal sequential surveillance for finance, public health, and other areas. *Sequential Analysis*, 28(3):310–337, 2009.

[52] Douglas Gale and Shachar Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346, 2003.

[53] Gourab Ghoshal, Liping Chi, and Albert-László Barabási. Uncovering the role of elementary processes in network evolution. *Scientific Reports*, 3:2920, 2013.

[54] Krista J Gile. Improved inference for respondent-driven sampling data with application to hiv prevalence estimation. *Journal of the American Statistical Association*, 106(493):135–146, 2011.

[55] Ben Golub and Evan Sadler. Learning in social networks. In *The Oxford Handbook of the Economics of Networks*.

[56] Igor Griva, Stephen G Nash, and Ariela Sofer. *Linear and Nonlinear optimization*, volume 108. SIAM, 2009.

[57] Fredrik Gustafsson. *Statistical sensor fusion*. Studentlitteratur, 2010.

[58] Jurgen Habermas and Jürgen Habermas. *The theory of communicative action*, volume 2. Beacon press, 1984.

[59] Jan Hansen, Carsten Schmidt, and Martin Strobel. Manipulation in political stock markets–preconditions and evidence. *Applied Economics Letters*, 11(7):459–463, 2004.

[60] Robin Hanson, Ryan Oprea, and David Porter. Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior & Organization*, 60(4):449–459, 2006.

[61] Simon S Haykin. *Neural networks: a comprehensive foundation*. Tsinghua University Press, 2001.

[62] Nicholas J Higham and Lijing Lin. On pth roots of stochastic matrices. *Linear Algebra and its Applications*, 435(3):448–463, 2011.

[63] Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982.

[64] Daniel Kahneman and Amos Tversky. Choices, values, and frames. In *Handbook of the Fundamentals of Financial Decision Making: Part I*, pages 269–278. World Scientific, 2013.

[65] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

[66] Yashodhan Kanoria and Omer Tamuz. Tractable bayesian social learning on trees. *IEEE Journal on Selected Areas in Communications*, 31(4):756–765, 2013.

[67] Samuel Karlin and Yosef Rinott. Classes of orderings of measures and related correlation inequalities: I. Multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10(4):467–498, 1980.

[68] Elihu Katz, Paul F Lazarsfeld, and Elmo Roper. *Personal influence: The part played by people in the flow of mass communications*. Routledge, 2017.

[69] Steven M Kay. *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.

[70] Steven M Kay. *Fundamentals of statistical signal processing: Practical algorithm development*, volume 3. Pearson Education, 2013.

[71] V. Krishnamurthy and A. Aryan. Quickest detection of market shocks in agent based models of the order book. In *IEEE 51st Annual Conference on Decision and Control (CDC)*, pages 1480–1485, Dec 2012.

[72] Vikram Krishnamurthy. Algorithms for optimal scheduling and management of hidden Markov model sensors. *IEEE Transactions on Signal Processing*, 50(6):1382–1397, 2002.

[73] Vikram Krishnamurthy. Bayesian sequential detection with phase-distributed change time and nonlinear penalty - A POMDP lattice programming approach. *IEEE Transactions on Information Theory*, 57(10):7096–7124, 2011.

[74] Vikram Krishnamurthy. Quickest detection POMDPs with social learning: Interaction of local and global decision makers. *IEEE Transactions on Information Theory*, 58(8):5563–5587, 2012.

[75] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes*. Cambridge University Press, 2016.

[76] Vikram Krishnamurthy and Sujay Bhatt. Sequential Detection of Market Shocks With Risk-Averse CVaR Social Sensors. *IEEE Journal of Selected Topics in Signal Processing*, 10(6):1061–1072, 2016.

[77] Vikram Krishnamurthy and William Hoiles. Online Reputation and Polling Systems: Data Incest, Social Learning, and Revealed Preferences. *IEEE Trans. Comput. Social Systems*, 1(3):164–179, 2014.

[78] Pavlo Krokhmal, Jonas Palmquist, and Stanislav Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of risk*, 4(2):43–68, 2002.

[79] Ziva Kunda. The case for motivated reasoning. *Psychological bulletin*, 108(3):480, 1990.

[80] Harold Kushner. *Weak convergence methods and singularly perturbed stochastic control and filtering problems*. Springer Science & Business Media, 2012.

[81] Jurij Leskovec, Deepayan Chakrabarti, Jon Kleinberg, and Christos Faloutsos. Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 133–145. Springer, 2005.

[82] Churlzu Lim, Hanif D Sherali, and Stan Uryasev. Portfolio optimization by minimizing conditional value-at-risk via nondifferentiable optimization. *Computational Optimization and Applications*, 46(3):391–415, 2010.

[83] Drew A Linzer. Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, 108(501):124–134, 2013.

[84] Luis López, Jose FF Mendes, and Miguel AF Sanjuán. Hierarchical social networks and information flow. *Physica A: Statistical Mechanics and its Applications*, 316(1):695–708, 2002.

[85] Dunia López-Pintado. Diffusion in complex social networks. *Games and Economic Behavior*, 62(2):573–590, 2008.

[86] Dunia López-Pintado. Influence networks. *Games and Economic Behavior*, 75(2):776–787, 2012.

[87] William S Lovejoy. Some monotonicity results for partially observed Markov decision processes. *Oper. Res.*, 35(5):736–743, 1987.

[88] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

[89] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, pages 415–444, 2001.

[90] Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.

[91] Sovan Mitra and Tong Ji. Risk measures in quantitative finance. *International Journal of Business Continuity and Risk Management*, 1(2):125–135, 2010.

[92] Alfred Müller and Dietrich Stoyan. *Comparison Methods for Stochastic Models and Risks*, volume 389. Wiley, 2002.

[93] Pauli Murto and Juuso Välimäki. Learning and information aggregation in an exit game. *The Review of Economic Studies*, 78(4):1426–1461, 2011.

[94] Mohammad Naghshvar and Tara Javidi. Active hypothesis testing: Sequentiality and adaptivity gains. In *46th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2012.

[95] Buddhika Nettasinghe and Vikram Krishnamurthy. What Do Your Friends Think? Efficient Polling Methods for Networks Using Friendship Paradox. *arXiv preprint arXiv:1802.06505*, 2018.

[96] Marcel F Neuts. *Structured stochastic matrices of MG-1 type and their applications*. Dekker, 1989.

[97] Svetlana Obraztsova, Omer Lev, Evangelos Markakis, Zinovi Rabinovich, and Jeffrey S Rosenschein. Distant truth: Bias under vote distortion costs. In *Pro-*

ceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, pages 885–892. International Foundation for Autonomous Agents and Multiagent Systems, 2017.

[98] André Orléan. Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion. *Journal of Economic Behavior & Organization*, 28(2):257–274, 1995.

[99] Marco Ottaviani. *Social learning in markets*. PhD thesis, Massachusetts Institute of Technology, 1996.

[100] Christos H Papadimitriou and John N Tsitsiklis. The complexity of Markov decision processes. *Mathematics of operations research*, 12(3):441–450, 1987.

[101] Yiangos Papanastasiou and Nicos Savva. Dynamic pricing in the presence of social learning and strategic consumers. *Management Science*, 63(4):919–939, 2016.

[102] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200, 2001.

[103] Charles R Plott and Kay-Yut Chen. Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem. 2002.

[104] Charles R Plott, Jorgen Wit, and Winston C Yang. Parimutuel betting markets as information aggregation devices: experimental results. *Economic Theory*, 22(2):311–351, 2003.

[105] H Vincent Poor and Olympia Hadjiliadis. *Quickest Detection*, volume 40. Cambridge University Press, 2009.

[106] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[107] Goran Radanovic and Boi Faltings. Incentives for truthful information elicitation of continuous signals. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[108] Maxim Raginsky. Shannon meets Blackwell and Le Cam: Channels, codes, and statistical experiments. In *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 1220–1224. IEEE, 2011.

[109] Johannes Rauh, Pradeep Kr Banerjee, Eckehard Olbrich, Jürgen Jost, Nils Bertschinger, and David Wolpert. Coarse-graining and the Blackwell order. *Entropy*, 19(10):527, 2017.

[110] Branko Ristic. *Particle filters for random set models*, volume 798. Springer, 2013.

[111] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

[112] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

[113] Stephane Ross, Masoumeh Izadi, Mark Mercer, and David Buckeridge. Sensitivity analysis of POMDP value functions. In *International Conference on Machine Learning and Applications, 2009. ICMLA'09.*, pages 317–323. IEEE, 2009.

[114] David M Rothschild and Justin Wolfers. Forecasting elections: Voter intentions versus expectations. 2011.

[115] Minoru Sakaguchi. *Information theory and decision making*. Statistics Dept., George Washington University, 1964.

[116] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860. ACM, 2010.

[117] Hawraa Salami, Bicheng Ying, and Ali H Sayed. Social learning over weakly connected graphs. *IEEE Transactions on Signal and Information Processing over Networks*, 3(2):222–238, 2017.

[118] Clay Shirky. *Cognitive surplus: Creativity and generosity in a connected age*. Penguin UK, 2010.

[119] Albert N Shiryaev and AB Aries. *Optimal Stopping Rules*, volume 8. Springer Science & Business Media, 2007.

[120] Nate Silver. *The signal and the noise: why so many predictions fail–but some don't*. Penguin, 2012.

[121] Adrian Smith. *Sequential Monte Carlo methods in practice*. Springer Science & Business Media, 2013.

[122] Torsten Söderström. *Discrete-time stochastic systems: estimation and control*. Springer Science & Business Media, 2012.

[123] VG Spokoiny and AN Shiryayev. *Statistical Experiments And Decision, Asymptotic Theory*, volume 8. World Scientific, 2000.

[124] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[125] Donald M Topkis. *Supermodularity and Complementarity*. Princeton University Press, 1998.

[126] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10(1):178–185, 2010.

[127] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[128] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 1998.

[129] Hal R Varian. Differential pricing and efficiency. *First monday*, 1(2), 1996.

[130] Pramod K Varshney. *Distributed detection and data fusion*. Springer Science & Business Media, 2012.

[131] Giampaolo Viglia, Roberta Minazzi, and Dimitrios Buhalis. The influence of e-word-of-mouth on hotel occupancy rate. *International Journal of Contemporary Hospitality Management*, 28(9):2035–2051, 2016.

[132] Dong Wang, Md Tanvir Amin, Shen Li, Tarek Abdelzaher, Lance Kaplan, Siyu Gu, Chenji Pan, Hengchang Liu, Charu C Aggarwal, Raghu Ganti, et al. Using humans as sensors: an estimation-theoretic perspective. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, pages 35–46. IEEE, 2014.

[133] Duncan J Watts and Peter Sheridan Dodds. Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458, 2007.

[134] Ivo Welch. Sequential sales, learning, and cascades. *Journal of Finance*, 47(2):695–732, June, 1992.

[135] Siqi Zhao, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. Human as real-time sensors of social and physical events: A case study of twitter and sports games. *arXiv preprint arXiv:1106.4300*, 2011.

[136] Aviv Zohar and Jeffrey S Rosenschein. Mechanisms for information elicitation. *Artificial Intelligence*, 172(16-17):1917–1939, 2008.