Leveraging Robustness for Information Design in Uncertain Environments

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

You Zu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Krishnamurthy Iyer

,

Acknowledgements

First, I would like to express my deepest gratitude to my advisor, Krishnamurthy Iyer, for his exceptional mentorship over the past five years. He introduced me to the fascinating field of information design and taught me how to think critically as a scientist and pursue my research interests. His kindness, support, and insightful advice have been invaluable throughout this journey. Whether meeting remotely during the pandemic or in person, every discussion with him was a deep learning experience. I am especially grateful for his guidance during the job search process, where he offered thoughtful advice and shared crucial resources. This thesis would not have been possible without his support and mentorship.

I would also like to extend my heartfelt appreciation to my committee members: Haifeng Xu, Ankur Mani, Ying Cui, and Maria Gini. Their guidance and support have been instrumental in shaping my work. Prof. Xu, as my collaborator for all the papers in this thesis, has provided deep insights and encouragement throughout my Ph.D. journey. I am grateful to Prof. Mani, Prof. Cui, and Prof. Gini for their invaluable feedback and advice, starting from the preliminary exam to various stages of this research. Additionally, I am deeply thankful to Prof. Yang Lu for her expertise in statistics and encouragement whenever needed.

I would like to extend my heartfelt thanks to Prof. William L. Cooper, Prof. Jean-Philippe (JP) Richard, Prof. Nick Arnosti, and Prof. Saumya Sinha for their invaluable guidance, encouragement, and generosity in sharing their expertise over the years. I am also deeply grateful to the staff in the department, whose support and assistance were instrumental in ensuring the smooth progression of my Ph.D. journey.

I am deeply thankful to my friends, whose companionship and support made this journey more enjoyable and fulfilling. Special gratitude goes to Jiaqi Liu, my roommate for three wonderful years, and her cat, Simba, who brought joy and comfort to my daily life. I also cherish the time spent with my peers: Xiaotang Yang, Zicheng Wang, Kang Kang, Bingnan Lu, Xuanming Zhang, Chengwenjian Wang, Chen Jiang, Chenyu Wu, Felipe Simon, Jake Roth, and Calvin Roth. To my ping-pong and board game friends—Zhaolin Li, Zhiyu Kang, Chang Li, and Jiuzhou Wang—thank you for the laughter and unforgettable moments that provided a much-needed break from research.

I am deeply thankful to my parents, Chenglin Zu and Xiaocui Xu, for their endless love, support, and encouragement, which carried me through the toughest times. To my cat, Pumpkin: your warmth and companionship brought comfort and joy throughout this journey.

Abstract

Information sharing between platforms and users is becoming increasingly important, with platforms often having information that is not directly visible to users. The platform's information advantage can influence users' decisions, as platforms often have more information about factors that impact outcomes. Consequently, platforms face challenges of sharing information or making recommendations that persuade users to act in ways that result in desirable outcomes. Certainly, the platform cannot make arbitrary recommendations to users without taking into account their incentives as the users may not follow the recommendations. Therefore, platforms seek to send recommendations that are likely to be adopted by users while simultaneously furthering platforms' objectives, such as long-term revenue maximization and welfare outcomes. The field of information design seeks to study such questions by developing systematic approaches to information-sharing strategies tailored to such objectives. This thesis addresses these questions through the work of two papers, each employing a *robustness framework* to mitigate uncertainties in information design. These studies were conducted in collaboration with Prof. Krishnamurthy Iyer and Prof. Haifeng Xu.

To develop our robustness approach, we first study how the platform makes recommendations with limited knowledge of the payoff-relevant state distribution. Specifically, we consider a static persuasion setting with known payoff-relevant state distribution but impose the restriction that the recommended action must be persuasive for all distributions in the neighborhood of the actual state distribution, i.e., we require the persuasion to be *robust*. For this problem, we analyze the *cost of robust persuasion*, i.e., the loss in the platform's expected utility from requiring the action recommendations to be persuasive for all distributions in the neighborhood. We provide upper and lower bounds of the loss under some mild regularity conditions. Using this characterization, we study two information design problems faced by platforms.

The first problem is the repeated persuasion setting between the platform and the users where neither the platform nor the users know the payoff-relevant distribution, and hence the platform has to persuade while learning the distribution. Our first contribution is the notion of *robust persuasiveness* in this setting with the detailed justification supporting the notion. Given this notion, our main result is an algorithm that, with high probability, is *robustly persuasive*. Using our characterization of the *cost of robust persuasion*, we show that our algorithm achieves vanishing average regret. We further prove that no algorithm can achieve better regret (up to logarithmic terms).

The second problem we study is a model of *Markovian persuasion* where the platform shares information about an evolving state and the state transitions are Markovian conditional on the users' actions. In such settings, given the underlying Markovian dynamics, the effectiveness of persuasion is impacted by the users' knowledge of the history, e.g., full knowledge vs. no knowledge of the history. We find sufficient conditions under which the platform's ability to persuade is unaffected by the users' historical information. In general, we consider settings where each user observes the history with an ℓ period lag, where the lag captures the degree of historical information of the users. Using the robustness approach, we propose an algorithm that achieves approximately optimal performance when the lag ℓ is large.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

With the rapid growth of online platforms, information sharing between platforms and users has become increasingly important. Often, platforms have information, e.g., quality, demand, and inventory of products that may not be directly visible to users. For instance, ride-hailing platforms such as Uber and Lyft have more information about the system (overall demand and demand imbalances, etc.), which drivers cannot directly observe. Similarly, online retailers like Amazon and Etsy have more information about products' quality from third-party sellers (after initial exploration or through in-house reviewers). Hiring platforms, such as LinkedIn, often have more comprehensive information about candidates' capabilities, which may be unknown to employers. In these examples, users' decision-making—whether drivers on ride-hailing platforms, buyers on online retailers, or employers on hiring platforms—is significantly influenced by the information shared by the platforms. Given this, the platform often shares relevant information with users, to enable them to make decisions that result in desirable outcomes such as improving overall welfare. For example, Uber provides drivers with information about the estimated wait times at the airport through the driver app. Amazon promotes individual sellers using labels such as "Amazon's choice" and "low price" to convince buyers that the promoted items are of high quality or low price. Therefore, platforms must understand how to make recommendations that result in desirable outcomes. However, to make recommendations that will be adopted by users, platforms need to consider challenges arising from the uncertainty and dynamics of the environment.

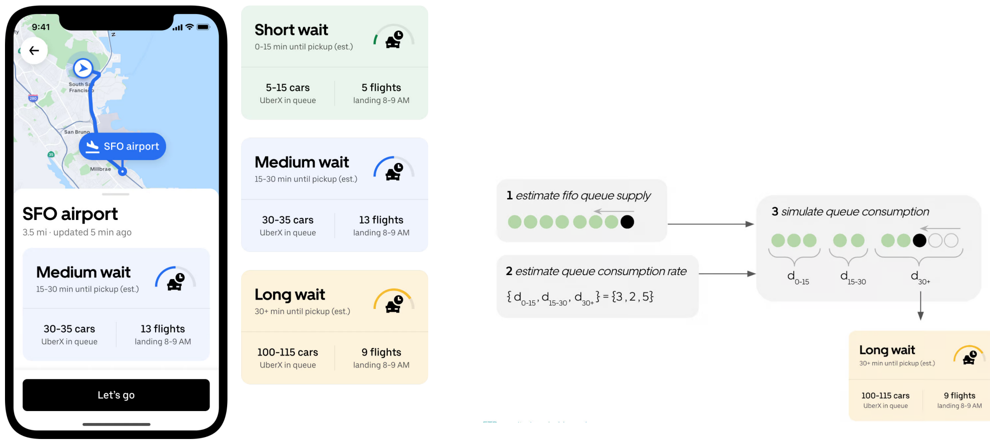
Here, we illustrate the challenges in the following concrete example.

1.1 Ride-sharing platform example

At most airports, drivers in a ride-hailing platform join a "first-in-first-out" (FIFO) queue, and wait to be matched to a ride request. Their wait times before receiving the next trip request depend on the current driver's queue length and the demand. When the demand for trips is high relative to the number of drivers in the queue (referred to as "undersupply"), the queue moves fast, and wait times for drivers are low. Conversely, during periods of low demand relative to queue length, the queue moves slowly and the drivers may face significantly longer wait times. For drivers who are not at the airport, understanding the demand at the airport is important when deciding whether to reposition there. Because drivers typically cannot directly observe demand at the airport before making this decision, Uber shares the information of estimated wait time with drivers. In particular, as in Figure 1.1a, Uber shares the estimated time to request (ETR) through the airport venue marker in the Uber driver app (Cai and Balestra, 2023). The venue marker displays a tile on the primary screen that contains the estimated wait time classified as short ($0 - 15$ minutes), medium ($15 - 30$ minutes), or long (> 30 minutes), the number of the drivers in the queue, and the number of flights arriving in the next hour (together these provide context for the estimated wait time). After observing the information, drivers make decisions about whether or not to join the queue at the airport in periods of undersupply and oversupply. In this context, each driver seeks to spend less waiting time, while the platform itself may have other goals, such as maximizing driver acceptance rate (Uber Technologies, Inc., 2024), capacity utilization rate, trip throughput, welfare[1] (Yan et al., 2020) and minimizing rider's waiting time (Feng et al., 2021), which are not fully aligned with drivers' interests.

If the platform and the drivers know the overall supply and demand distribution, the platform can make reliable recommendations that optimize its own goals while maintaining drivers' satisfaction. The first challenge is that the platform usually

[1]Welfare in the context of a ride-haling platform is defined by the aggregation of the utilities of the rider, driver and platform, and can be viewed as a measure of the total value created for all parties.



(a) Example ETR tiles in the venue marker. (b) ETR model design.

Figure 1.1: Uber's demand and ETR forecasting at airports

does not have the knowledge of the underlying distribution and thus must *learn* to make recommendations over time. For example, to estimated ETR, Uber builds "supply model" and "demand model" to simulate the consumption of the queue up to the drivers' estimated position (Cai and Balestra, 2023), as illustrated in Figure 1.1b. In the "supply model", Uber utilizes the feature of the overall driver dynamics in the queue, e.g., the abandonment rate, the rate of trip radar matching [2], etc. However, when a new driver joins the platform, the platform does not know the driver's abandonment distribution and must *learn* it over time, as having more historical data from the driver. Further, the distribution can vary widely based on factors such as weather conditions, events in the city, and traffic patterns, which pose much uncertainty in the distribution. Given this uncertainty in the environment, Uber needs to understand how to share information or make persuasive recommendations to influence drivers' actions.

The second challenge arises from the system's dynamics. In this example, drivers' actions not only affect their and the platform's immediate rewards but also affect the queue length for drivers over time. As more drivers reposition themselves or queue

[2] Trip radar matching is a feature in the Uber Driver app that provides drivers with nearby requests

up, wait times for drivers increase. Due to these dependencies, the effectiveness of information-sharing is impacted by drivers' knowledge of the history. In practice, drivers' historical knowledge may be limited and outdated. For example, drivers may observe the queue length at the airport during drop-offs or a glimpse of the venue marker at a previous time. Sometimes, a driver's historical information can be outdated due to factors such as infrequent visits to specific locations (e.g., airports), and rapidly changing conditions, such as evolving traffic patterns and rider demand. Given these dynamics, it becomes important for platforms to understand how to make recommendations that align with drivers' incentives while considering long-term objectives.

Motivated by these problems, we tackle two models of information design in uncertain environments in Chapter 3 and Chapter 4. In both models, we use the *robustness approach* to design a signaling mechanism that approximately achieves optimal payoff. In Chapter 2, we describe the static persuasion problem between a *sender* (e.g., a platform) and a *receiver* (e.g., drivers) where the sender makes action recommendations given the observation of a payoff-relevant state. We adopt the assumption common in the literature that prior to observing the realized state, the sender commits to a signaling mechanism that maps each state to a possibly random *action recommendation*. Upon observing the realized state, the sender recommends an action according to the chosen signaling mechanism. The receiver then chooses an action that maximizes her expected utility given the recommendation. Certainly, the sender cannot make arbitrary recommendations regardless of the receiver's interest. A natural requirement for the sender is to make *persuasive* recommendation, i.e., the recommendation that the receiver will find optimal to follow. We first consider the benchmark where the sender and the receiver commonly know the state distribution. From the standard results Kamenica and Gentzkow, 2011; Bergemann and Morris, 2016; Dughmi and Xu, 2021, the sender's persuasion problem can be formulated as a linear program. We consider a robust version of this problem, where we impose the restriction that the signaling mechanism must be persuasive for all distributions in the neighborhood of the actual state distribution, i.e., *robustly persuasive.* We characterize the notion *cost of robust persuasion* as the loss in the sender's expected utility from being robustly persuasive and provide the upper and lower bound of the loss. Using this full characterization of the cost of robust persuasion, we design approximate optimal signaling mechanisms for models in Chapter 3 and Chapter 4.

In Chapter 3, we extend the standard Bayesian persuasion setting and consider the situation where neither the sender nor the receiver knows the distribution of the payoff-relevant state. Instead, the sender learns the distribution over time by observing the state realizations. It captures a more realistic setting where the sender lacks the distributional data initially while making recommendations. Unlike the standard persuasion setting where the persuasive requirement can be easily specified, it is unclear how to specify the persuasive requirements in the absence of distributional knowledge. To tackle this challenge, we pose a natural criterion for persuasiveness that centers *robustness* as a requirement in the face of uncertainty. Specifically, our criterion requires that the sender's recommendations are persuasive under all state distributions in a set of "confidence regions" which contains the true distribution with a given degree of confidence; these confidence regions shrink over time as the sender observes more state realizations. Under this persuasiveness constraint, we leverage the results of robust persuasion in Chapter 2 and propose an efficient algorithm that, with high probability, makes persuasive recommendations and at the same time achieves vanishing average regret. The idea behind the approach is to maintain a set of candidate state distributions based on the observed state realizations in the past and then choose a signaling mechanism that is *robustly* persuasive for all candidate distributions and maximize the sender's utility. Using the characterization of the cost of robust persuasion, we further prove that this regret is optimal (up to logarithmic terms) by constructing a persuasion instance where being robustly persuasive leads to a substantial loss to the sender.

In Chapter 4, we apply the result of robust persuasion to a *Markovian persuasion process*, where the system's next state is fully determined (stochastically) by the current state and receiver's action. In such a setting, given the underlying Markovian dynamics, the effectiveness of persuasion is impacted by the receivers' knowledge of the history. In contrast to Chapter 3 and previous literature (Lingenbrink and Iyer, 2019; Anunrojwong et al., 2022) where the receivers have no information about the history, we focus on the setting where the receivers have limited historical information, e.g., the drivers' information about the driver's queue length from some time ago. Hence, the receivers' beliefs are endogenously determined by the historical realizations of the system and the shared information. In order to make persuasive recommendations, the sender must take into account the existence of such limited historical information. To understand the sender's persuasion problem under the receivers' limited historical information, we define the notion of an *information model*, which specifies how each receiver's belief (prior to receiving a recommendation) is related to the history of the process. In addition to the full-history information model where the receivers observe the entire history and the no-history where the receivers have no historical information, we consider a sequence Φ_ℓ of partial-history information model where each receiver observes the history of the system with an ℓ period lag, for some fixed $\ell \geq 1$. We first formulate the sender's persuasion problem in the full-history and no-history information models as a succinct linear program by observing that the optimal signaling mechanism is history-independent under no-history information model, whereas additionally depends on the previous state-action pair in the full-history information model. However, we show that the analysis of the sender's persuasion problem in the information model Φ_ℓ presents technical intricacies that leave open even the question of the existence of an optimal signaling mechanism. Due to the complexity of solving the persuasion problem optimally under partial-history information models, we take an alternative approach and ask whether simple mechanisms can achieve approximately optimal payoffs while simultaneously persuasive under limited historical information. Using the underlying Markovian dynamics and the robustness approach in Chapter 2, we construct a simple history-independent signaling mechanism that is simultaneously persuasive under the partial-history information model Φ_ℓ for all large ℓ. Our result establishes the effectiveness of a simple history-independent signaling mechanism when the participants in a platform may have limited historical information. Together, our results in Chapter 3 and Chapter 4 contribute to the study of Bayesian persuasion in the dynamic setting and highlight the importance of robustness in designing signaling mechanisms in dynamic settings.

In Chapter 5, we summarize the main contributions of this thesis and highlight areas for future research that can extend the current work to more complex environments. In particular, while our results of robust persuasion in information design provide a foundation for developing effective signaling mechanisms, further exploration is needed to account for scenarios involving dynamically evolving state distribution, multiple senders settings, and partial feedback scenarios. By advancing these avenues, future work can enhance the adaptability of robust signaling mechanisms, allowing platforms to more effectively align user actions with desirable system outcomes in real-world applications.

Chapter 2

Robust Bayesian Persuasion

In this chapter, we will develop the robustness result that underpins our approach to information design in uncertain environments. The work in this chapter appeared as part of Zu et al. (2024).

2.1 Introduction

Consider a static Bayesian persuasion model where a sender shares information about the payoff-relevant state with a receiver. Upon receiving the information, the receiver chooses an action that affects both the sender and receiver's utilities. This model makes a restrictive assumption that the state distribution is fully known to both the sender and the receiver. However, in practical applications, significant uncertainty often surrounds the state distribution, making perfect knowledge unrealistic. For instance, as mentioned in Chapter 1, Amazon promotes individual sellers using labels such as "Amazon's choice" to convince the user that the products are of high quality, but the quality of new products or new sellers is often unknown to both the platform and its users. Given this, how should the platform decide whether or not to recommend the product to its users? To address this question, we depart from the standard Bayesian persuasion setting and consider situations where neither the sender nor the receiver knows the distribution. Instead, the sender has limited knowledge of the state distribution. In this context, "limited knowledge" implies that while the sender may not have access to the exact state distribution, they have the knowledge that the actual state distribution lies within a specific set of possibilities.

To set the stage, in this chapter, we begin by discussing the standard model of Bayesian persuasion with complete knowledge of underlying state distribution and present existing results in (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019; Dughmi and Xu, 2021) that demonstrate that in this setting, the sender's problem can be formulated as a linear program. Then, we study the setting where the sender does not know the exact underlying state distribution but knows that it lies within a certain neighborhood. We address the following question: how can the sender make recommendations that are persuasive for all distributions within that neighborhood, i.e., robustly persuasive recommendations? To characterize the sender's loss from this robustness requirement compared to the complete information setting, we introduce the notion of the *cost of robust persuasion*, defined as the loss in the sender's expected utility from requiring the recommendations to be persuasive for all distributions in the neighborhood. We analyze the cost of *cost of robust persuasion* and provide an upper bound and a matching lower bound of it.

2.2 Related Work

We refer readers to (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019) as well as (Candogan, 2020) for a general overview of the recent developments and (Dughmi, 2017) for a survey from an algorithmic perspective.

Specifically, our work studies the robust persuasion problem where the underlying state distribution is unknown and quantifies the cost of robustness. We briefly discuss some work that have looked at robust persuasion. Kosterina (2018) studies a persuasion setting in the absence of the common prior assumption. In particular, the sender has a known prior, whereas only the set in which the receiver's prior lies is known to the sender. Furthermore, the sender evaluates the expected utility under each signaling mechanism with respect to the worst-case prior to the receiver. Similarly, Hu and Weng (2020) study the problem of the sender persuading a privately informed receiver, where the sender seeks to maximize her expected payoff under the worst-case information of the receiver. Finally, Dworczak and Pavan (2020) study a related setting and propose a lexicographic solution concept where the sender first identifies the signaling mechanisms that maximize her worst-case payoff, and then among them choose the one that maximizes the expected utility under her conjectured prior. In contrast to these works, our model focuses on a setting with common, but unknown, prior, and where the receiver has no private information.

2.3 Static Persuasion Problem with Known State Distribution

In this section, we introduce the standard Bayesian persuasion problem (Kamenica and Gentzkow, 2011) between a single sender and a single receiver where the state distribution is commonly known. Let Ω denote a finite set of states and A denote a finite set of actions. The payoff-relevant state ω is drawn from a commonly known distribution $\mu \in \mathcal{B}_0 \subseteq \Delta(\Omega)$. (Here, for any finite set X, $\Delta(X)$ denotes the set of all probability distributions over X.) Let S denote a finite set of signals. Prior to observing the realized state, the sender commits to a *signaling mechanism* $\sigma : \Omega \to \Delta(S)$ that is a randomized mapping from each state $\omega \in \Omega$ to a probability distribution over signals. Upon observing the realized state, the sender sends a signal to the receiver according to the signaling mechanism σ. The receiver, upon receiving the signal, updates the belief about the realized state and chooses an action $a \in A$ based on the posterior belief. The sender's utility $v(\omega, a)$ and receiver's utility $u(\omega, a)$ depend on both the realized state and the receiver's chosen action. Without loss of generality, we assume that $v(\omega, a) \in [0, 1]$ for all $\omega \in \Omega$ and $a \in A$. Further, to avoid trivialities, we assume $|\Omega| \geq 2$ and $A \geq 2$. We refer to the tuple $\mathcal{I} = (\Omega, A, u, v, \mathcal{B}_0)$ with $u : \Omega \times A \to \mathbb{R}$ and $v : \Omega \times A \to [0, 1]$ as an *instance* of our problem. An important result as highlighted by Bergemann and Morris (2016), an analog of the revelation principle (Myerson, 1979), allows us to restrict our attention to signaling mechanisms that are both *direct* and *persuasive*. A signaling mechanism is *direct* if the signals directly correspond to action recommendations, i.e., $S = A$. For simplicity, we focus on direct signaling mechanisms in the remainder of this chapter. Formally, let $\sigma(\omega, a)$ to denote the probability of recommending action $a \in A$ conditioned on the observing the state realization is $\omega \in \Omega$ and let

$\mathcal{S} = \{\sigma : \sigma(\omega, \cdot) \in \Delta(A) \text{ for each } \omega \in \Omega\}$ denote the set of all direct signaling mechanisms.

A signaling mechanism is *persuasive*, if conditioned on receiving an action recommendation $a \in A$, it is indeed optimal for the receiver to choose action a. Formally, let $a \in A$ be an action with $\sum_{\omega\in\Omega} \mu(\omega)\sigma(\omega, a) > 0$.Upon receiving the recommendation a, the receiver's posterior belief that the realized state is ω is given by Bayes' rule as $\frac{\mu(\omega)\sigma(\omega,a)}{\sum_{\omega'\in\Omega} \mu(\omega')\sigma(\omega',a)}$, and hence $\sum_{\omega\in\Omega} \left(\frac{\mu(\omega)\sigma(\omega,a)}{\sum_{\omega'\in\Omega} \mu(\omega')\sigma(\omega',a)}\right) u(\omega, a')$ denotes her expected utility of choosing action $a' \in A$ conditioned on receiving the recommendation a. For the receiver's expected utility to be maximized by choosing action a, we need $\sum_{\omega\in\Omega} \mu(\omega)\sigma(\omega, a) (u(\omega, a) - u(\omega, a')) \geq 0$ for all $a' \in A$. Since the inequality is trivially satisfied if $\sum_{\omega\in\Omega} \mu(\omega)\sigma(\omega, a) = 0$, the set of persuasive mechanisms $\mathsf{Pers}(\mu)$ is given by

$$\mathsf{Pers}(\mu) \triangleq \left\{\sigma \in \mathcal{S} : \sum_{\omega\in\Omega} \mu(\omega)\sigma(\omega, a) \left(u(\omega, a) - u(\omega, a')\right) \geq 0, \text{ for all } a, a' \in A\right\}. \tag{2.1}$$

First, we note that the set $\mathsf{Pers}(\mu)$ is a convex polytope for all $\mu \in \Delta(\Omega)$. If the distribution is fixed and known, $\mathsf{Pers}(\mu)$ is a set characterized by polynomially-sized linear inequalities (in the number of states and actions). Second, the set $\mathsf{Pers}(\mu)$ is non-empty, since it always contains the "full-information mechanism" which recommends the receiver's optimal action at each state.

Given a persuasive signaling mechanism $\sigma \in \mathsf{Pers}(\mu)$, the the receiver is incentivized to choose the recommended action. Assuming ties are broken in favor of the recommended action, the sender's expected utility is given by

$$V(\mu, \sigma) \triangleq \sum_{\omega\in\Omega} \sum_{a\in A} \mu(\omega)\sigma(\omega, a)v(\omega, a).$$

Since $V(\mu, \sigma)$ is linear in σ, the problem of selecting an optimal persuasive signaling mechanism is given by the following linear program:

$$\mathsf{OPT}_{\mathcal{I}}(\mu) = \sup_{\sigma} V(\mu, \sigma), \text{ subject to } \sigma \in \mathsf{Pers}(\mu). \tag{2.2}$$

2.4 Robust Persuasion with Known State Distribution

In this section, we extend the standard Bayesian persuasion model to the situation where the sender may need to be robustly persuasive, i.e., persuasive for a set of distributions around the actual distribution.

To define a robust persuasion problem, let us fix an instance $\mathcal{I}$. For a given set of distributions $\mathcal{B} \subseteq \mathcal{B}_0$ that contains the true distribution $\mu \in \Delta(\Omega)$, the set of signaling mechanisms that are simultaneously persuasive for all distributions in $\mathcal{B}$ is given by $\mathsf{Pers}(\mathcal{B}) = \cap_{\mu' \in \mathcal{B}} \mathsf{Pers}(\mu')$. We remark that for any non-empty set $\mathcal{B}$, the set $\mathsf{Pers}(\mathcal{B})$ is convex since it is an intersection of convex sets $\mathsf{Pers}(\mu)$, and it is non-empty since it always contains the "full-information signaling mechanism". Hence, the sender's optimal expected utility among all such mechanisms is given by $\sup_{\sigma \in \mathsf{Pers}(\mathcal{B})} V(\mu, \sigma)$.

2.4.1 Cost of Robust Persuasion

In this section, we characterize the loss in the sender's expected utility from requiring the signaling mechanism to be persuasive for all distributions in a neighborhood around the state distribution. To measure this loss, we first define the notion of the *cost of robust persuasion*, a quantity that depends on the neighborhood. Formally, it is defined as

$$\mathsf{Gap}(\mu, \mathcal{B}) \triangleq \sup_{\sigma \in \mathsf{Pers}(\mu)} V(\mu, \sigma) - \sup_{\sigma \in \mathsf{Pers}(\mathcal{B})} V(\mu, \sigma). \tag{2.3}$$

Thus, $\mathsf{Gap}(\mu, \mathcal{B})$ captures the difference in the sender's expected utility (under μ) between using the optimal persuasive signaling mechanism for the distribution μ and using the optimal signaling mechanism that is persuasive for all distributions $\mu' \in \mathcal{B}$. In the following, we will often focus on the case where $\mathcal{B}$ is $\mathsf{B}_1(\mu, \epsilon)$, where $\mathsf{B}_1(\mu, \epsilon) \triangleq \{\mu' \in \Delta(\Omega) : \|\mu' - \mu\|_1 \leq \epsilon\}$ denote the (closed) ℓ_1-ball of radius $\epsilon > 0$ at $\mu \in \Delta(\Omega)$.

For general persuasion instances, one can show that the cost of robust persuasion can be severe: in Appendix B.4.1, we present a persuasion instance and a distribution μ such that for any $\epsilon > 0$, the cost of being robustly persuasive for the set $\mathsf{B}_1(\mu, \epsilon)$ of distributions satisfies $\mathsf{Gap}(\mu, \mathsf{B}_1(\mu, \epsilon)) = \frac{1}{2}$. The instance we present there is pathological, with an action that is optimal for the receiver at a single unique distribution. To obtain meaningful insights on the cost of robust persuasion, we seek to exclude such instances by imposing some regularity condition on the instances.

To state these regularity conditions, we need some notation. For each action $a \in A$, let $\mathcal{P}_a$ denote the set of state distributions for which action a is optimal for a receiver:

$$\mathcal{P}_a \triangleq \left\{\mu \in \Delta(\Omega) : \mathbf{E}_\mu \left[u(\omega, a)\right] \geq \mathbf{E}_\mu \left[u(\omega, a')\right], \text{ for all } a' \in A\right\}.$$

It is without loss of generality to assume that for each $a \in A$, the set $\mathcal{P}_a$ is non-empty. (This is because a receiver can never be persuaded to play an action $a \in A$ for which $\mathcal{P}_a$ is empty, and hence such an action can be dropped from A.)

We consider the following regularity conditions on the persuasion instances:

Assumption 1 (Regularity Conditions)**.** The instance $\mathcal{I}$ satisfies the following conditions:

1. There exists $d > 0$ such that for each $a \in A$, the set $\mathcal{P}_a$ contains an ℓ_1-ball of size d. Let $D > 0$ denote the largest value of d for which the preceding is true, and let $\eta_a \in \mathcal{P}_a$ be such that $\mathsf{B}_1(\eta_a, D) \subseteq \mathcal{P}_a$.
2. There exists a $p_0 > 0$ such that for all $\mu \in \mathcal{B}_0$ we have $\min_\omega \mu(\omega) \geq p_0 > 0$.

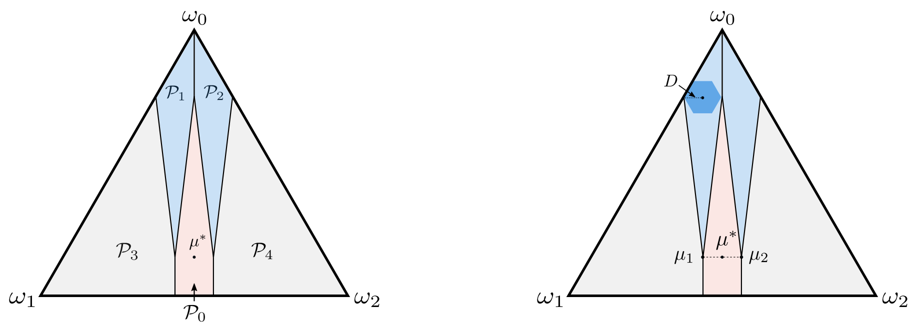
The first condition requires that each such set $\mathcal{P}_a$ has a non-empty relative interior; this excludes the pathological instances like that in Appendix B.4.1, for which there exists an action a with $\mathcal{P}_a$ a singleton. We note that this condition is analogous to the Slater condition in convex optimization, imposing a non-empty interior on the feasibility region to obtain strong duality. The second condition is technical and is made primarily to ensure the potency of the first condition: without it, the sets $\{\mathcal{P}_a\}_{a \in A}$ may satisfy the first condition in $\Delta(\Omega)$, while failing to satisfy it relative to the subset $\Delta(\{\omega : \mu(\omega) > 0\})$ for some $\mu \in \mathcal{B}_0$. Taken together, these regularity conditions serve to avoid pathologies, and henceforth we restrict our attention only to those instances satisfying these regularity conditions.

Under the regularity conditions, our first result shows that the cost of robust persuasion $\mathsf{Gap}(\mu, \mathcal{B})$ is at most linear in the size of the set $\mathcal{B}$.

Proposition 1. For any instance that satisfies the regularity conditions, for all $\mu \in \mathcal{B}_0$ and for all $\epsilon \geq 0$, we have $\mathsf{Gap}(\mu, \mathsf{B}_1(\mu, \epsilon)) \leq \left(\frac{4}{p_0^2 D}\right) \epsilon$.

The proof of the upper bound is obtained through an explicit construction of a signaling mechanism $\widehat{\sigma}$ that is persuasive for all distributions in the set $\mathsf{B}_1(\mu, \epsilon)$, and by showing that the sender's expected payoff under $\widehat{\sigma}$ is close to that under the optimal signaling mechanism in $\mathsf{Pers}(\mu)$. For this construction, we first use the geometry of the instance to split the distribution μ into a convex combination of distributions that either fully reveal the state, or are well-situated in the interior of the sets $\mathcal{P}_a$. (It is here that we make use of the two regularity assumptions.) We then construct the mechanism $\widehat{\sigma}$ to induce, under prior μ, the aforementioned beliefs as posteriors. Finally, we show that for any prior μ' close enough to μ, the posteriors induced by $\widehat{\sigma}$ are close to the posteriors induced under prior μ, implying that these posteriors lie within the sets $\mathcal{P}_a$. This proves the persuasiveness of $\widehat{\sigma}$ for all distributions μ' close to μ. We provide the complete proof in the Appendix A.1.1.

Next, we provide a (worst-case) lower bound on Gap. We accomplish this by carefully constructing a persuasion instance $\mathcal{I}_0$ where being robustly persuasive leads to a substantial loss to the sender. The instance $\mathcal{I}_0$ has three states $\Omega = \{\omega_0, \omega_1, \omega_2\}$ and five actions $A = \{a_0, a_1, a_2, a_3, a_4\}$ for the receiver. At a high level, the receiver's preference can be illustrated in Fig. 2.1a, which depicts the receiver's optimal action for any belief in the simplex. The regions $\mathcal{P}_i$ in the figure corresponds to the set of beliefs that induce action $a_i \in A$ as the receiver's best response. The instance is crafted so that the sets $\mathcal{P}_1$ and $\mathcal{P}_2$ that induce actions a_1 and a_2 respectively are symmetric and extremely narrow with the width controlled by an ℓ_1-ball of radius D contained within, as depicted in Fig. 2.1b. (Since $|\Omega| = 3$, the ℓ_1-ball here is a hexagon.) For completeness, the receiver's utility is listed explicitly in Table 2.1. The sender seeks to persuade the receiver into choosing one of actions a_1 and a_2 (regardless of the state); all other actions are strictly worse for the sender. Formally, we set $v(\omega, a) = 1$ if $a \in \{a_1, a_2\}$ and 0 otherwise, for all ω. The sender's initial knowledge regarding the state distribution is captured by the set $\mathcal{B}_0 = \{\mu \in \Delta(\Omega) : \min_\omega \mu \geq p_0\}$, while the distribution of interest is $\mu^* = (p_0, \frac{1-p_0}{2}, \frac{1-p_0}{2})$, corresponding to the midpoint of the tips of the sets $\mathcal{P}_i$, as shown in Fig. 2.1b. We focus on the setting where the instance parameters D and p_0 satisfy $Dp_0 < 1/64$.



(a) Receiver's preferences (b) Prior μ^*

Figure 2.1: The persuasion instance $\mathcal{I}_0$.

	a_1	a_2	a_3	a_4
ω_0	$2D^2$	$2D^2$	$-2D(1-p_0-2D)$	$-2D(1-p_0-2D)$
ω_1	$(1-2D)(1-D)-p_0$	$(D+1)(2D-1)+p_0$	$2(1-p_0-2D)(1-D)$	$-2(1-p_0-2D)(D+1)$
ω_2	$(D+1)(2D-1)+p_0$	$(1-2D)(1-D)-p_0$	$-2(1-p_0-2D)(D+1)$	$2(1-p_0-2D)(1-D)$

Table 2.1: Receiver's utility in instance $\mathcal{I}_0$, with $u(\omega, a_0)$ normalized to 0 for all $\omega \in \Omega$.

The following proposition shows that in the instance $\mathcal{I}_0$, it is costly to require the signaling mechanism to be robustly persuasive for a set of distributions around μ^*. The result also implies that the bound on $\mathsf{Gap}(\cdot)$ obtained in Proposition 1 is almost tight, except for a factor of $1/p_0$.

Proposition 2. For the instance $\mathcal{I}_0$, we have $\mathsf{OPT}(\mu^*) = 1$. Furthermore, for all $\epsilon \in (0, D)$, we have

$$\mathsf{Gap}(\mu^*, \{\mu^*, \bar{\mu}_1, \bar{\mu}_2\}) \geq \frac{\epsilon}{8Dp_0},$$

where $\bar{\mu}_1 = \mu^* + \frac{\epsilon}{2}(e_1 - e_2)$, $\bar{\mu}_2 = \mu^* + \frac{\epsilon}{2}(e_2 - e_1)$, where the belief e_i puts all its weight on ω_i.

We defer the rigorous algebraic proof of the lower bound to Appendix A.1.2 and present a brief sketch using a geometric argument here. In the instance $\mathcal{I}_0$, the distribution μ^* can be written as a convex combination $\mu^* = (\mu_1 + \mu_2)/2$, where μ_1 and μ_2 are the tips of regions $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively (see Fig. 2.1b). Thus, by the splitting lemma (Aumann et al., 1995a), it follows that the optimal signaling mechanism sends signals that induce posterior beliefs μ_1 and μ_2 leading to receiver's choice of a_1 and a_2 respectively. Since the sender can always persuade the receiver to choose one of her preferred actions, we obtain $\mathsf{OPT}(\mu^*) = 1$. On the other hand, for a signaling mechanism to be robustly persuasive for all distributions ϵ-close to the distribution μ^* for sufficiently small ϵ, the posteriors for the sender's preferred actions a_1, a_2 induced by the signaling mechanism have to be shifted up significantly in the narrow region. Such a large discrepancy ultimately forces the sender to suffer a substantial loss in the expected payoff.

2.4.2 Conclusion

In this section, we studied a static Bayesian persuasion problem where the sender must be robustly persuasive. We denote the notion of robustness and provide the upper and lower bound of the sender's cost of being robust persuasion Gap. Note that this characterization relies on the specifics of the persuasion problem (for instance, the use of the splitting lemma to construct a feasible robust solution) and the regularity conditions to obtain the linear bounds on the cost of robustness, without which the cost could be $\mathcal{O}(1)$, which is a constant independent of the radius of the neighborhood. Whether these regularity conditions could be generalized to other linear settings is an interesting question for further investigation. In Chapter 3, we use this characterization to perform tight regret analysis in a repeated persuasion setting of unknown state distribution setting. In Chapter 4, using the robustness approach, we design a simple signaling mechanism in the Markov persuasion setting.

Chapter 3

Learning to Persuade on the Fly: Robustness Against Ignorance

The work presented in this chapter appears in Zu et al. (2024).

3.1 Introduction

Examples of online platforms recommending content or products to their users abound in online economy. For instance, online retailers like Amazon or Etsy recommend products from third-party sellers to users, styling services like Stitch Fix recommend clothing designs made by custom brands, and online platforms like YouTube or Spotify recommend content or playlist generated by creators. There are two intrinsic challenges in such online recommendations, which we address simultaneously in this chapter. First, the platform making such recommendations often needs to balance the dual objectives of being persuasive (i.e., making obedient recommendations that will be adopted by the users (Bergemann and Morris, 2016)) as well as furthering the platform's goals such as increased sales, fewer returns or more engaged users. Second, the platform often faces a large volume of *new* products/contents/services with a-priori unknown quality/reward distributions and thus has to learn to make good recommendation. We tackle these two challenges by studying learning to persuade on the fly.

3.1.1 Motivating Applications

To motivate the problem we consider, we now describe two concrete examples in the domain of two-sided platforms.

Example 3.1.1 (**Content recommendations by online media platforms**). Consider a media platform like YouTube or TikTok, that recommends content created by independent creators ("channels") to its users. New channels regularly join the platform, and start producing content whose *quality distribution* is unknown to both the platform and its users. Here, by a content's quality, we refer to how engaging, interesting or relevant the users find the content. Despite this lack of knowledge, the platform faces the problem of deciding whether to recommend content from such new channels to its stream of users. In this context, the users seek to consume fresh and high-quality content, while the platform itself may have other goals, such as maximizing user engagement or increasing channel exposure, which are not fully aligned with users' interests. Furthermore, from extensive user-level data, the platform may have good estimates about the utility a user derives from consuming a particular content. A user encountering a new channel may have a prior belief about its quality distribution based on their past experiences in the platform, and from any information provided by the channel itself on their profile. Furthermore, the user may have additional (partial) information from any reviews or ratings left by previous users (or similar summary statistics). For each new content from a channel, the platform observes its quality (perhaps after an initial exploration or through in-house reviewers) and decides whether or not to recommend the content to its users. If the platform and the users know a channel's content quality distribution, the platform can reliably make recommendations that optimize its own goals while maintaining user satisfaction, by consistently mixing high-quality content with some mediocre ones. However, given the lack of such distributional information, the platform must *learn* to make such recommendations over time, as the channel produces more content.

Example 3.1.2 (**Recommendations on hiring platforms**). Consider a hiring platform, where employers receive recommendations about candidates for recruitment (e.g., "recommended matches" in LinkedIn Recruiter). These recommendations are typically tailored to the employer's project requirements. However, within the set of candidates satisfying the requirements, there would be a range of capabilities/fit, whose *distribution* would be unknown to the platform or the employer. Nevertheless, for any particular candidate who might be interested in the position, the platform may be able to assess the candidate's capability based on various candidate features, such as her endorsements, references, etc, using which the platform decides whether or not to recommend the candidate. Similarly, the employer through the course of interviewing different candidates may learn about the capability distribution. While the employer would prefer to be matched with a few high-capability candidates to interview, the platform may have additional incentives from having to cater to the candidates-side of the market, such as increasing the overall number of interviews. Once again, if the distribution of the candidates' capabilities is known to the platform and the employer, the platform could reliably recommend candidates to optimize its goals while simultaneously meeting the employer's preferences. But, without such information, the platform needs to learn to recommend candidates as they apply over time.

In this chapter, we study the problem faced by such a platform learning to make *persuasive* recommendations to a stream of users. While previous work has studied information design in two-sided markets — ranging from recommending products from third-party sellers on e-commerce platforms like Amazon and eBay (Gur et al., 2023; Elliott et al., 2022), recommending drivers by sharing demand trend on ride-sharing services like Uber and Lyft (Yang et al., 2019), to accommodation and rental recommendations in Airbnb (Romanyuk and Smolin, 2019) — the common assumption is that the platform knows the underlying state distribution. Our work in this chapter contributes to this literature by relaxing this strong assumption.

3.1.2 Modeling Contributions

Formally, we study a repeated persuasion setting between a sender and a stream of receivers, where at each time t, the sender shares some information correlated to some payoff-relevant *state* with the corresponding receiver. The state at each time t is drawn independently and identically from an unknown distribution, and subsequent to receiving information about it, the newly-arriving myopic receiver chooses an action from a finite set, generates payoffs, and then leaves the system forever. The sender seeks to persuade this stream of receivers into choosing actions that are aligned with her preference by selectively sharing information about the state at each round.

To tackle the practical challenge of making recommendations in the absence of distributional data, we depart from the standard Bayesian persuasion setting and consider situations where neither the sender nor the receiver knows the distribution of the payoff relevant state. Instead, the sender learns this distribution over time by observing the state realizations. We adopt the assumption common in the literature on Bayesian persuasion that the sender commits to a signaling mechanism that, at each time step, maps the realized state to a possibly random *action recommendation.* Such a commitment assumption is well-justified for settings of interest to this work since online platforms typically design and implement the information sharing policy as software in advance, rendering frequent changes unlikely. This advance design serves as a commitment device organically.

As in Chapter 2, we require that the sender makes persuasive recommendations that the receiver will find optimal to follow. This incentive compatibility requirement can be easily justified by an application of revelation principle. In the case where the sender and the receivers know the state distribution, the persuasiveness requirement implies that, subsequent to each recommendation, the recommended action maximizes the receiver's expected utility under the conditional state distribution (given the recommendation). However, in the absence of such distributional knowledge, it is not immediately clear how to impose persuasiveness.

Our main modeling contribution addresses this issue by proposing a natural criteria for persuasiveness when neither the sender nor the receivers know the state distribution. The starting point of our approach is the observation that any persuasiveness criteria directly corresponds to a model of receivers' response on receiving a recommendation (just as in the case of known state distribution). Thus, by considering reasonable behavioral models for the receiver, we develop in Section 3.3.2 a persuasiveness criterion that centers *robustness* as a requirement in the face of uncertainty. Specifically, our criterion requires that the sender's recommendations are persuasive under all state distributions in a set of "confidence regions" which contain the true distribution with a given degree of confidence; these confidence regions shrink over time as the sender observes more state realizations. This is in line with the approach in statistics that uses confidence regions to address the uncertainty in parameter estimates. Furthermore, this robustness requirement naturally leads to conservative recommendations, thereby making it likely that the recommendations will be accepted. We refer to this notion as β-robustly persuasiveness where $1-\beta$ denotes the confidence level.

3.1.3 Algorithmic Contribution and Regret Characterization

A sender who simply recommends the receiver's best action at the realized state will certainly be persuasive with complete confidence ($\beta = 0$), but may end up with a significant loss in her utility when compared to her utility had she known the state distribution. However, since the sender observes the state realizations over time, she has the opportunity to make more profitable recommendations with greater confidence in their persuasiveness as she obtains more information. Thus, the sender's goal is to carefully manage this tradeoff between the confidence in persuasiveness and her utility, and achieve low regret against the optimal signaling mechanism with the knowledge of the state distribution.

The primary theoretical contribution of this chapter is an efficient algorithm that, with high probability, makes persuasive recommendations and at the same time achieves vanishing average regret. The algorithm we propose proceeds by maintaining at each time a set of candidate state distributions, based on the observed state realizations in the past. The algorithm then chooses a signaling mechanism that is simultaneously persuasive for each of the candidate distributions and maximizes the sender's utility. Due to this aspect of the algorithm, we name it the *Robustness against Ignorance* ($\mathfrak{Rai}$) algorithm.

By a careful choice of the candidate set of distributions at each time period, we show in Theorem 1 that the $\mathfrak{Rai}$ algorithm satisfies the β-robustly persuasiveness criterion for $\beta = o(T)$, where T is the horizon length. Furthermore, exploiting the structure of the problem, we show in Proposition 3 that the $\mathfrak{Rai}$ algorithm involves solving a polynomially-sized (in number of states and actions) linear program at each period. Taken together, these results establish our algorithm's persuasiveness and its computational efficiency.

To characterize the regret of the $\mathfrak{Rai}$ algorithm, we use the cost of robust persuasion in 2.4.1. Recall that $\mathsf{Gap}(\mu, \mathcal{B})$ captures the loss in the sender's expected utility (under distribution μ) from using a signaling mechanism that is persuasive for all distributions in the set $\mathcal{B}$, as opposed to using one that is persuasive only for the distribution μ. In Proposition 1, we establish that, under some regularity conditions, the sender's cost of robust persuasion $\mathsf{Gap}(\mu, \mathcal{B})$ is at most linear in the radius of the set $\mathcal{B}$. Further, we provide a matching lower bound in Proposition 2 by carefully crafting a persuasion instance and using its geometry to prove a linear cost of robust persuasion; this instance thus serves as a lower-bound example for robust persuasion.

Using this characterization of the cost of robust persuasion, we perform a tight regret analysis of persuasion under unknown state distribution in Section 3.5. Our positive result, Theorem 2, establishes that for any persuasion setting satisfying the aforementioned regularity conditions, the $\mathfrak{Rai}$ algorithm achieves $\mathcal{O}(\sqrt{T \log T})$ regret with high probability. Furthermore, in Theorem 3, we provide a matching lower-bound (up to $\log T$ terms) for the regret of any algorithm that makes persuasive recommendations. In addition to the characterization of Gap and the custom persuasion instance from Propositions 1 and 2, the proofs of these theorems rely on concentration results for sums of independent random vectors in Banach spaces.

Our results contribute to the work on online learning that seeks to evaluate the value of knowing the underlying distributional parameters in settings with repeated interactions (Kleinberg and Leighton, 2003). In particular, our results fully characterize the sender's *value of knowing the state distribution* for repeated persuasion. Our well-motivated approach to relax the strong assumption of complete distributional knowledge in the standard persuasion setting is also aligned with the prior-independent mechanism design literature (Dhangwatnotai et al., 2015; Chawla et al., 2013).

3.2 Related Work

Our work contributes to the burgeoning literature on Bayesian persuasion and information design in economics, operations research and computer science.

Online learning & mechanism design. Our work in this chapter subscribes to the recent line of work that studies the interplay of learning and mechanism design in incomplete-information settings, in the absence of common knowledge on the prior. We briefly discuss the ones closely related to our work.

Castiglioni et al. (2020) focus on persuasion setting with a commonly known prior distribution of the state but unknown receiver types chosen *adversarially* from a finite set. They show that effective learning, in this case, is computationally intractable but does admit $\mathcal{O}(\sqrt{T})$ regret learning algorithm, after relaxing the computability constraint. Our model complements theirs by focusing on known receiver types but unknown state distributions in a *stochastic* setup. Moreover, we achieve a similar (and tight) regret bound through a computationally *efficient* algorithm. Also relevant to us is the recent line of work on Bayesian exploration (Kremer et al., 2014; Mansour et al., 2015, 2016) which is also motivated by online recommendation systems. In contrast to our setting, these models assume the prior is commonly known but the realized state is unobservable and thus needs to be learned during the repeated interactions.

Dispensing with the common prior itself, Camara et al. (2020) study an adversarial online learning model where both a mechanism designer and the agent learn about the states over time. The agent is long-lived and is assumed to minimize her counterfactual (internal) regret in response to the mechanism designer's policy, which is assumed to be non-responsive to the agent's actions. The authors use a reinforcement learning approach to mechanism design and characterize the policy regret of the mechanism designer, taking into account the agents' responses, relative to the best-in-hindsight fixed mechanism. Similar to our work, the regret bounds require the characterization of a "cost of robustness" of the underlying design problem. While related, the receivers in our model are short-lived and myopic. Furthermore, our model is stochastic rather than adversarial, and thus a prior exists in our model. More broadly, our model is similar in spirit to the prior-independent mechanism design literature (Dhangwatnotai et al., 2015; Chawla et al., 2013), though our setup is different. Moreover, our algorithm is measured by the regret whereas approximation ratios are often adopted for prior-independent mechanism design.

Recent works by Hahn et al. (2019, 2020) study information design in online optimization problems such as the secretary problem (Hahn et al., 2019) and the prophet inequalities (Hahn et al., 2020), and propose constant-approximation persuasive schemes. These online optimization problems often take the adversarial approach, which is different from our stochastic setup and learning-focused tasks. Therefore, our results are not comparable.

Safe online learning: Our work in this chapter also relates to safe online learning. The work by Moradipari et al. (2021) is the most relevant to our work. They study a safe online learning problem where the linear reward and a single linear constraint depend on different unknown parameters. The learner has access to both the reward and the side information about the safety set. In this setting, they propose an algorithm based on linear Thompson Sampling and achieve the regret $\mathcal{O}(\sqrt{T \log^3 T})$. The key difference is that their analysis relies on the assumption that a known safe action is an interior point of the safety set for all possible values of the unknown parameter. Under our regularity conditions, it is true that for every distribution there exists a signaling mechanism for which all the persuasiveness constraints hold strictly (that is, the order of the quantifiers from above is interchanged). However, it is unclear if this weaker assumption would be sufficient for their setting.

Amani et al. (2019) study a linear stochastic multi-armed bandit problem where the linear reward function and a *single* linear safety constraint depend on an unknown parameter. Their main algorithm and its analysis depend on knowing (a lower bound on) the safety gap, i.e., the slack in the safety constraint for the optimal solution under the true parameter. When the safety gap is known and positive (i.e., the constraint is inactive), they prove a regret of $\mathcal{O}(\log T\sqrt{T})$. On the other hand, if the safety gap is known to be zero, they only achieve a regret of $\tilde{O}(T^{2/3})$. They provide a separate algorithm for the case of an unknown safety gap and state a regret bound of $\tilde{O}(T^{2/3})$. In our setting, there are *multiple* persuasiveness constraints, and many of these would be active for the true distribution in nontrivial settings. Thus even if their work can be extended to multiple constraints, it may only guarantee $\tilde{O}(T^{2/3})$ regret bound.

Usmanova et al. (2019) seek to minimize a smooth convex function over a set of uncertain linear constraints where both the coefficients and constant parameters are unknown. Although our problem is a specific case of theirs, our model does not meet their central assumption of being able to evaluate the constraints at any point within a small neighborhood of the feasible set.

Recent works by (Pacchiano et al., 2021; Khezeli and Bitar, 2020; Moradipari et al., 2020, 2021) study a similar safe learning problem in different contexts. Pacchiano et al. (2021) require that at each time, the chosen action has an expected cost below a certain threshold. Khezeli and Bitar (2020); Moradipari et al. (2020) study safe learning where in addition to maximizing the expected reward, one requires the reward to be above a threshold with high probability. In these settings, the objective and the constraint are aligned. Our setup is different because the sender's and the receivers' preferences, corresponding respectively to the objective and constraints, need not be aligned with each other. Most importantly, all these work impose a single constraint at each round, whereas our persuasiveness condition requires multiple constraints at each round.

Online linear/convex optimization: Since the persuasion problem can be posed as a linear program, our work also relates to the online convex optimization problem. Mostly, the focus here is on adversarial setting where the loss function (objective) is adversarially chosen and revealed at the end of each time period. Some papers (Cao et al., 2019; Mahdavi et al., 2013) focus on the stochastic setting, but either study an unconstrained problem (Cao et al., 2019) or study a batch algorithm rather than an online algorithm (Mahdavi et al., 2013). Focusing on the constraints, and using the terminology of (Kim and Lee, 2023), these work typically consider either a *long-term constraint* formulation (Yu et al., 2017; Mahdavi et al., 2011; Neely and Yu, 2017; Yi et al., 2021; Kim and Lee, 2023; Cao and Liu, 2018), or consider a *cumulative constraint* formulation (Yuan and Lamperski, 2018; Yi et al., 2022; Guo et al., 2022). The *long-term constraint* formulation requires feasibility on average in the long run. Such constraints are reasonable in applications where the constraints are on aggregate quantities, such as budgets in online advertising (Liakopoulos et al., 2019), covering constraints in sensor networks, capacity constraints in online routing (Agrawal and Devanur, 2014), etc. However, this type of constraint is not reasonable in our setting as it would permit the sender to make poor recommendations in some rounds as long as it can be compensated by good recommendations in other rounds. In contrast, the *cumulative* constraint formulation focuses on bounding the sum of the positive-parts of the constraints (which require some quantity to be non-positive). This formulation is equivalent to our formulation if the cumulative constraint can be made zero. However, most previous work allow for some constraint violation and seek to bound the order of the violations. In the presence of such violations, our formulation is stronger.

Finally, by characterizing the persuasion problem as a Stackelberg game between the sender's choice of a signaling mechanism and the receiver's subsequent choice of an action, our work is related to the broader work on the characterization of regret in repeated Stackelberg settings (Balcan et al., 2015; Dong et al., 2018; Chen et al., 2020).

3.3 Model

Consider a persuasion setting with a single long-run *sender* persuading a stream of homogeneous *receivers* who arrive sequentially over a time horizon of length T. At each time $t \in [T] = \{0, \cdots, T-1\}$, a state $\omega_t \in \Omega$ is drawn independently and identically from a state distribution $\mu^* \in \Delta(\Omega)$. (Here, for any finite set X, $\Delta(X)$ denotes the set of all probability distributions over X.) We focus on the setting where Ω is a known finite set, however, the distribution μ^* is *unknown* to both the sender and the receivers. To capture the sender's initial knowledge (before time $t = 0$) about the distribution μ^*, we assume that the sender knows that μ^* lies in the set $\mathcal{B}_0 \subseteq \Delta(\Omega)$.

At each time $t \in [T]$, the sender observes the realized state ω_t, and shares with the arriving receiver an *action recommendation* $a_t \in A$ (chosen according to a *signaling algorithm*, as described below), where A is a finite set of actions available to the receivers. The receiver then chooses an action $\hat{a}_t \in A$ (not necessarily equal to a_t). This results in the receiver obtaining a utility $u(\omega_t, \hat{a}_t)$ and the sender obtaining a utility $v(\omega_t, \hat{a}_t)$. The setup of the problem follows the same assumption and notation in Chapter 2, where we define the tuple $\mathcal{I} = (\Omega, A, u, v, \mathcal{B}_0)$ and assume that $v(\omega, a) \in [0, 1]$ for all $\omega \in \Omega$ and $a \in A$. For completeness, we recall that $u : \Omega \times A \to \mathbb{R}$ and $v : \Omega \times A \to [0, 1]$ with $|\Omega| \geq 2$ and $|A| \geq 2$ to avoid trivial cases.

Before we proceed, we make a few remarks on the persuasion instance. First, the preceding description does not specify a model of the receivers' actions $\hat{a}_t$. As we discuss below in Section 3.3.2, this issue is intertwined with the *persuasiveness constraints* that we impose on the sender's signaling algorithm, and hence, we postpone the discussion until then. Second, while we have assumed that the sender shares information in the form of action recommendations, under the persuasiveness constraints we consider it can be shown that this is without loss of generality. Third, while our definition of an instance assumes that the receivers are homogeneous, it can be extended to allow for heterogeneity of receivers' utility; our results continue to apply in the setting where the receivers' types are observable to the sender. Finally, we assume that the sender knows the receivers' utility. This is justified in the context of our applications of interest, namely online platforms, where given the scale, the platform may have good estimates about user utility from extensive user-level data.

Informally, given a persuasion instance $\mathcal{I}$, the sender's goal is to systematically make action recommendations such that her long-run total utility is maximized. We now describe the formal algorithmic aspects of the sender's goal.

As each time t, the sender chooses an action recommendation a_t based on past state realizations, the past action recommendations as well as the past actions chosen by the receivers. To separate the historical information from that about the present, we define the *history* h_t at the beginning of time t as follows: $h_t = \cup_{\tau<t}\{(\tau, \omega_\tau, a_\tau, \hat{a}_\tau)\}$ (with $h_0 = \emptyset$), and note that the sender observes (h_t, ω_t) prior to making the recommendation a_t at time t. We also note that, since the receivers do not know the state distribution μ^*, neither the past actions recommended by the sender nor the past actions chosen by the receivers carry any information about μ^* beyond that contained in the state realizations. Thus, the part of the history that is relevant to the sender consists of only the state realizations until time t.

A *signaling algorithm* $\mathfrak{a} \equiv \mathfrak{a}(\mathcal{I})$ for the sender specifies, at each time $t \in [T]$ and after any history h_t and state ω_t, a probability distribution $\sigma^{\mathfrak{a}}(h_t, \omega_t, \cdot) \in \Delta(A)$ over the set of actions. (We sometimes drop the superscript $\mathfrak{a}$ when it is clear from the context.) Specifically, once the state ω_t is realized, the sender draws the action recommendation a_t independently according to the distribution $\sigma(h_t, \omega_t, \cdot) \in \Delta(A)$.

Thus, the probability that the sender recommends an action $a \in A$ is given by $\sigma(h_t, \omega_t, a)$. Implicitly, the notion of a signaling algorithm reflects the assumption that the sender *commits* to a mechanism for sending recommendations.

Given an instance $\mathcal{I}$ and a signaling algorithm $\mathfrak{a}$, the sender's total (realized) utility is given by

$$V_{\mathcal{I}}(\mathfrak{a}, T) \triangleq \sum_{t \in [T]} v(\omega_t, \hat{a}_t).$$

Thus, to evaluate the performance of a signaling algorithm, we need a model of the receivers' response subsequent to receiving the action recommendations. Rather than directly specifying such a response model, we instead model conditions on the signaling algorithm $\mathfrak{a}$ which result in *obedient* responses from the receivers, i.e., which lead each receiver to choose the action recommended: $\hat{a}_t = a_t$.

Any such condition on the signaling algorithm $\mathfrak{a}$ implies a model of receivers' response, and the converse can be assumed without loss of generality by invoking incentive compatibility and the revelation principle. Henceforth, we refer to such a condition as a *persuasiveness criterion*.

To motivate these persuasiveness criteria on the signaling algorithms, we first discuss the setting where the sender and the receivers commonly know the state distributions. This setting will also serve as a benchmark to compare the performance of any signaling algorithm satisfying certain persuasiveness requirements.

3.3.1 Benchmark: Known State Distribution

Consider the setting where the sender and the receivers commonly know the state distribution $\mu^* = \mu \in \Delta(\Omega)$. In this setting, each receiver responds by choosing the action that maximizes her expected utility under the posterior belief about the state given the action recommendation. In particular, the sender's problem decouples across time periods, and the sender's problem at each time is a static Bayesian persuasion model, which can be formulated as a linear program. Recall that in Section 2.3, the problem of selecting an optimal persuasive signaling mechanism is given by the following linear program:

$$\mathsf{OPT}_{\mathcal{I}}(\mu) \triangleq \max_{\sigma} V(\mu, \sigma), \text{ subject to } \sigma \in \mathsf{Pers}(\mu). \tag{3.1}$$

Finally, letting σ^* denote an optimal signaling mechanism to the preceding optimization problem, the algorithm $\mathfrak{a}$ that sets $\sigma^{\mathfrak{a}}(h_t, \omega_t, a) = \sigma^*(\omega_t, a)$ after any history h_t optimizes the sender's total expected utility when the state distribution is known, with total expected utility given by $T \cdot \mathsf{OPT}_{\mathcal{I}}(\mu)$.

3.3.2 Persuasiveness Criterion: Unknown Distribution

We now return to the setting with unknown state distribution, and discuss refined persuasiveness conditions on the signaling algorithm under which the receivers' response can be reasonably assumed to equal the recommendation. In particular, we propose and motivate a condition on the signaling algorithm, namely the *robust persuasiveness* criterion as described in Definition 2, and provide detailed justification supporting the notion.

We begin with the simplest criterion inspired from the known distribution setting. As the sender observes the past state realizations, the empirical distribution γ_t, with $\gamma_t(\omega) \triangleq \frac{1}{t} \sum_{\tau < t} \mathbf{I}\{\omega_\tau = \omega\}$, provides an estimate for the unknown distribution μ^*. A natural first idea, which we call the *naive criterion*, simply requires the algorithm to act as if this estimate is exact:

Definition 1. *A signaling algorithm* $\mathfrak{a}$ *satisfies the* naive criterion *if each* $\sigma^{\mathfrak{a}}[h_t]$ *is persuasive under the empirical distribution at time* t*, i.e.,* $\sigma^{\mathfrak{a}}[h_t] \in \mathsf{Pers}(\gamma_t)$ *for all* $t \in [T]$.

The naive criterion can be motivated through a particular behavioral model of the receivers involving social learning. Specifically, consider a platform setting where each receiver (i.e., a user) arrives with an uninformative Dirichlet prior over the state distribution μ^*, and observes all the past state realizations. The latter holds if we assume there is social learning among the receivers, where each receiver leaves a feedback that is read by all subsequent receivers. Then, at each time t the corresponding receiver's belief about the state would be exactly the empirical distribution γ_t, and thus the receiver would optimally accept the recommendation made by the platform if it uses a signaling algorithm satisfying the naive criterion.

However, from a practical perspective, the preceding model makes very restrictive assumptions. First, in a platform setting, the users' prior belief over μ^*, if such a prior exists at all, is unlikely to be known to the platform, and need not be the same across different users (let alone be the uninformative Dirichlet prior). Second, even with social learning, the users typically would not observe all the past state realizations (or even just the empirical distribution); this is because not all users leave reviews in a platform, and a user would typically read only a subset of available reviews. Thus, under a realistic model of social learning, the receivers' belief about the state would be in general different from the empirical distribution.

In addition to relying on restrictive behavioral assumptions, there are other deficiencies with the naive criterion that render it ill-suited as a criterion for ensuring persuasiveness. First, the naive criterion is especially weak in the initial stages of persuasion due to the lack of sufficient data; at these initial stages, the constraint based on the empirical distribution may not constrain the sender's recommendations. For instance, if the empirical distribution at the beginning happens to be skewed and concentrates on very few states, then the naive criterion imposes no restriction on the action recommendations at any previously unseen state since it has zero empirical probability. Second, an algorithm satisfying the naive criterion may still make *inconsistent* recommendations across time. That is, for such an algorithm, there may not exist a single belief μ for which the recommendations as a whole are persuasive, i.e., $\sigma^{\mathfrak{a}}[h_t] \in \mathsf{Pers}(\mu)$ for all t. Any such belief μ, if it exists, provides a justification for the signaling algorithm, and larger the set of such beliefs the stronger the justification. For instance, the "full-information" signaling algorithm $\mathfrak{Full}$, which always recommends the receivers' best action $a_t \in \operatorname{argmax}_{a \in A} u(\omega_t, a)$ after any history h_t, has the strongest justification since all beliefs $\mu \in \Delta(\Omega)$ satisfy $\sigma^{\mathfrak{Full}}[h_t] \in \mathsf{Pers}(\mu)$. On the other hand, one can easily construct examples where an algorithm satisfying naive criterion fails to have even a single belief justifying it, due to inconsistencies in recommendations across different periods.

Summarizing, the primary reason for the weaknesses of the naive criterion is its reliance on the point estimate γ_t in the place of receivers' inherently uncertain beliefs about the state. Even for basic inferential tasks, such point estimates are seldom sufficient. Without explicitly incorporating this uncertainty into its conditions, an algorithm would provide no *confidence* that the receivers will accept and act according to the recommendations. To remedy these weaknesses, we propose the following criterion that embraces the notion of robustness in its conditions.

Definition 2. *Given $\beta \geq 0$, a signaling algorithm $\mathfrak{a}$ is* β-robustly persuasive*, if there exists (history-dependent) sets $\mathcal{C}_t \subseteq \mathcal{B}_0$ for all time t, such that*

1. ***Robustness:*** *The signaling mechanism $\sigma^{\mathfrak{a}}[h_t]$ is persuasive for* all *beliefs in the set $\mathcal{C}_t$: for each $t \in [T]$, we have*

$$\sigma^{\mathfrak{a}}[h_t] \in \mathsf{Pers}(\mathcal{C}_t) \triangleq \cap_{\mu \in \mathcal{C}_t} \mathsf{Pers}(\mu).$$

2. ***Coverage:*** *The sets $\mathcal{C}_t$ all contain the true state distribution μ^* with high probability:*

$$\mathbf{P}_{\mu^*} \left(\cap_{t \in [T]} \mathcal{C}_t \ni \mu^* \right) \geq 1 - \beta.$$

(Here, $\mathbf{P}_{\mu^*}$ represents the probability with respect to the (unknown) distribution μ^* and any independent randomization in the algorithm.)

The first condition in the criterion enforces robustness, requiring that the signaling mechanism at time t, $\sigma^{\mathfrak{a}}[h_t]$, is persuasive with respect to *all* beliefs in the set $\mathcal{C}_t$. These sets implicitly capture the uncertainties regarding the receivers' beliefs, and by depending on the history, reflect any learning occuring over time. (We note that the set $\mathsf{Pers}(\mathcal{C}_t)$ is indeed non-empty, as it contains the "full-information" mechanism.) The second condition in the criterion requires these sets to have good coverage properties, i.e., these sets contain the state distribution μ^* with high probability.

To further motivate the criterion, we delve a bit into the perspective of social learning in a platform setting mentioned earlier. Here, while it is a strong assumption to require the receivers to know the exact empirical distribution, it is fair to assume that the receivers observe (summary statistics about) a sizeable proportion of past state realization. In particular, many common empirical principles, such as the "90-9-1 rule"(Antelmi et al., 2019; Van Mierlo et al., 2014), posit that a constant fraction of the users leave feedback in the platform. In this context, a receiver who starts with some sufficiently diffuse prior over μ^*, and who learns from past (incomplete) feedback, will have a belief about the state that is close enough to the empirical distribution. Thus, a signaling algorithm that makes recommendations that are persuasive for *all* beliefs close to the empirical distribution would ensure that such a receiver would find it optimal to follow the recommended action. Our proposed criterion, by using a robustness approach, abstracts away from the details of such an explicit model, and captures the receivers' response through the uncertainty sets $\mathcal{C}_t$.

Observe that as long as the sets $\mathcal{C}_t$ contain the empirical distribution γ_t, the preceding criterion is stronger than the naive criterion. More importantly, in addition to capturing more realistic models of social learning, the coverage and the robustness conditions together also overcome the other inadequacies of the naive criterion that we discussed above. To see this, note that, at the initial stages t when the data is insufficient, good coverage requires the set $\mathcal{C}_t$ to be large, and thus the action recommendations are severely constrained (even at the states that have not been realized), unlike the case with the naive criterion. Similarly, the robustness ensures that any belief $\mu \in \cap_{t\in[T]}\mathcal{C}_t$ provides a justification for the signaling algorithm, thus precluding any inconsistencies across time. In particular, with probability at least $1-\beta$, the true state distribution μ^* justifies all the recommendations made by a β-robustly persuasive signaling algorithm: $\mathbf{P}_{\mu^*}\left(\sigma^{\mathfrak{a}}[h_t] \in \mathsf{Pers}(\mu^*) \text{ for all } t \in [T]\right) \geq 1-\beta$.

The parameter β in the criterion plays the same role as that played by significance level in inference. In particular, low values of β correspond to high level of confidence in the uncertainty sets $\mathcal{C}_t$. Finally, it is easy to see that β-robustly persuasive algorithms exist for any $\beta \geq 0$; in fact, choosing the sets $\mathcal{C}_t = \mathcal{B}_0$ for all $t \in [T]$, it follows that the algorithm $\mathfrak{Full}$ is 0-robustly persuasive.

Given the preceding discussion, we hereafter assume that for any signaling algorithm $\mathfrak{a}$ that is β-robustly persuasive for some (small) $\beta \geq 0$, the receivers' response $\hat{a}_t$ equals the action recommendation a_t at each time t. Thus, for any such algorithm $\mathfrak{a}$, the sender's total utility reduces to $V_{\mathcal{I}}(\mathfrak{a}, T) = \sum_{t\in[T]} v(\omega_t, a_t)$.

3.3.3 Sender's Learning Problem

Finally, we describe the evaluation metric for the performance of any algorithm satisfying the preceding persuasiveness criterion by comparing the sender's utility $V_{\mathcal{I}}(\mathfrak{a}, T)$ against the known-distribution benchmark given by $T \cdot \mathsf{OPT}(\mu^*)$. Specifically, we measure the sender's *regret* from using a β-robustly persuasive algorithm $\mathfrak{a}$ by

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{a}, T, \mu^*) \triangleq T \cdot \mathsf{OPT}_{\mathcal{I}}(\mu^*) - V_{\mathcal{I}}(\mathfrak{a}, T) = T \cdot \mathsf{OPT}_{\mathcal{I}}(\mu^*) - \sum_{t \in [T]} v(\omega_t, a_t). \quad (3.2)$$

We are now ready to formalize the sender's learning problem. Begin by noticing that one must require the signaling algorithm $\mathfrak{a}$ to be β-robustly persuasive for some small β in order for the second equality above to hold, i.e., for the receivers' responses to match the recommendations. At the same time, 0-robustly persuasiveness is an excessive requirement, with no hope of resulting in a sub-linear regret. (In Appendix B.4.2, we present an example instance where any 0-robustly persuasive algorithm necessarily obtains a linear regret.) Thus, the central problem is to design, for any given instance $\mathcal{I}$, an algorithm $\mathfrak{a}$ that is β-robustly persuasive for small (vanishing) β and simultaneously achieves sublinear regret with high probability.

3.4 The Robustness Against Ignorance ($\mathfrak{Rai}$) Algorithm

Having described the learning problem faced by the sender, in this section, we present a signaling algorithm that we call the Robustness Against Ignorance ($\mathfrak{Rai}$) algorithm. Here, we show that the $\mathfrak{Rai}$ algorithm is β-robustly persuasive with $\beta = o(1)$, relegating the regret analysis to Section 3.5.

Before describing our proposed algorithm, we briefly motivate our design approach. Observe that if the state distribution μ^* is known, then the sender's problem is given by the linear program (2.2), and thus the optimal signaling mechanism can be efficiently computed. Thus, a natural learning approach is to solve at each time t the estimated version of the LP (2.2), where the unknown distribution μ^* is replaced by the empirical distribution γ_t, and use the corresponding optimal signaling mechanism for that time period. However, this alone is not sufficient to obtain an algorithm that is β-robustly persuasive, which requires the signaling mechanisms to be persuasive for all distributions in some small neighborhood of μ^*. To elaborate, simply solving the estimated LP may yield solutions that are only ϵ-feasible for distributions close to the empirical distribution, i.e., some of the persuasiveness constraints for such nearby distributions may get violated. In fact, optimizing the estimated LP may result in a mechanism that is not persuasive for *any* other distribution close to the empirical distribution. Thus, an immediate challenge is in determining how to use the empirical distribution estimate to find well-performing signaling mechanisms that are persuasive (with high probability) for *all* distributions in a small neighborhood around the unknown state distribution. Part of this challenge is to carefully choose the corresponding neighborhoods without significantly sacrificing the performance of the mechanism.

The algorithm we propose is adaptive. An alternative is to adopt an "explore-then-commit" design (Lattimore and Szepesvári, 2020), where the algorithm uses the state realizations in the first t periods (for some carefully chosen t) to estimate the unknown distribution and subsequently commits to a single signaling mechanism for the remaining time periods. However, it is unlikely that such a algorithmic design can achieve strong regret guarantees in our setting, since it is known that such an approach yields the sub-optimal $\mathcal{O}(T^{2/3})$ regret in simple multi-armed bandit problems (Lattimore and Szepesvári, 2020). This observation illustrates the need for adaptivity to obtain order-wise optimal regret.

To meet these challenges, our algorithm $\mathfrak{Rai}$ proceeds by adaptively maintaining, at each time $t \geq 0$, a set $\mathcal{B}_t$ of candidates for the (unknown) distribution μ^*. This set is a (closed) ℓ_1-ball of radius ϵ_t at the empirical distribution γ_t. It then selects a signaling mechanism that maximizes the sender expected utility w.r.t. the empirical estimate γ_t among mechanisms that are persuasive for all distributions $\mu \in \mathcal{B}_t$. Finally, it makes an action recommendation a_t using this signaling mechanism, given the state realization ω_t. The $\mathfrak{Rai}$ algorithm is formally described in Algorithm 1. Here, we use the notation $\mathsf{Pers}(\mathcal{B})$ to denote the set of signaling mechanisms that are simultaneously persuasive under all distributions μ in the set $\mathcal{B} \subseteq \Delta(\Omega)$: $\mathsf{Pers}(\mathcal{B}) = \cap_{\mu \in \mathcal{B}} \mathsf{Pers}(\mu)$. We remark that for any non-empty set $\mathcal{B} \subseteq \Delta(\Omega)$, the set $\mathsf{Pers}(\mathcal{B})$ is convex since it is an intersection of convex sets $\mathsf{Pers}(\mu)$, and is non-empty since it contains the full-information signaling mechanism. Furthermore, we let $\mathsf{B}_1(\mu, \epsilon) \triangleq$

ALGORITHM 1: The Robustness Against Ignorance ($\mathfrak{Rai}$) algorithm

Input: Instance $\mathcal{I}$, Time horizon T
Parameters: $\gamma_0 \in \mathcal{B}_0$, $\{\epsilon_t > 0 : t \in [T]\}$
Output: $a_t \in A$ for each $t \in [T]$
for $t = 0$ *to* $T-1$ **do**
  Choose any $\sigma[h_t] \in \arg\max_\sigma \{V(\gamma_t, \sigma) : \sigma \in \mathsf{Pers}(\mathcal{B}_t)\}$;
  Recommend $a_t = a \in A$ with probability $\sigma(\omega_t, a; h_t)$;
  Update $\gamma_{t+1}(\omega) \leftarrow \frac{1}{t+1} \sum_{\tau=0}^{t} \mathbf{I}\{\omega_\tau = \omega\}$ for each $\omega \in \Omega$;
  Set $\mathcal{B}_{t+1} \leftarrow \mathsf{B}_1(\gamma_{t+1}, \epsilon_{t+1})$;
end

$\{\mu' \in \Delta(\Omega) : \|\mu' - \mu\|_1 \leq \epsilon\}$ denote the (closed) ℓ_1-ball of radius $\epsilon > 0$ at $\mu \in \Delta(\Omega)$.

From the intuitive description, it follows that the sets $\mathcal{B}_t = \mathsf{B}_1(\gamma_t, \epsilon_t)$ naturally play the role of the covering sets $\mathcal{C}_t$ in the definition of β-robustly persuasiveness. Specifically, the parameters $\{\epsilon_t : t \in [T]\}$ control the degree of persuasiveness of the algorithm: larger values of ϵ_t imply that the algorithm is β-robustly persuasive for smaller values of β. (In particular, if all ϵ_t are larger than 2, the algorithm reduces to the full-information algorithm $\mathfrak{Full}$, and is 0-robustly persuasive.) Unsurprisingly, larger values of ϵ_t also lead to larger regret, and hence the sender must choose ϵ_t to optimally trade-off the persuasiveness of the algorithm against its regret.

Our first main result characterizes $\mathfrak{Rai}$'s persuasiveness for a particular choice of parameter values which we show in Section 3.5 to be regret-optimal.

Theorem 1. *For each $t \in [T]$, let $\epsilon_t = \min\{\sqrt{\frac{|\Omega|}{t}}\left(1 + \sqrt{\Phi \log T}\right), 2\}$ with $\Phi > 0$. Then, the $\mathfrak{Rai}$ algorithm is β-robustly persuasive with*

$$\beta = \sup_{\mu^* \in \mathcal{B}_0} \mathbf{P}_{\mu^*}\left(\cap_{t \in [T]} \mathcal{B}_t \not\ni \mu^*\right) \leq T^{1 - \frac{3\Phi\sqrt{\Omega}}{56}}.$$

In particular, for $\Phi > 20$, we have $\beta \leq T^{-0.5}$.

The proof of the persuasiveness of $\mathfrak{Rai}$ follows by showing that the empirical distribution γ_t concentrates around the unknown state distribution μ^* with high probability. Since, after any history h_t, the signaling mechanism $\sigma[h_t]$ chosen by the algorithm is persuasive for all distributions in an ℓ_1-ball around γ_t, we deduce that it is persuasive under μ^* as well. To show the concentration result, we use a concentration inequality for independent random vectors in a Banach space (Foucart and Rauhut, 2013); the full proof is provided in Appendix B.1.1.

We observe that to get strong persuasiveness guarantees, the choice of ϵ_t in the preceding theorem requires the knowledge of the time horizon T. However, applying the standard doubling tricks (Besson and Kaufmann, 2018), one can convert our algorithm to an *anytime* version that has the same regret upper bound guarantee, at the cost of a weakened persuasiveness guarantee, where the persuasiveness β is weakened to a constant arbitrarily close to 0.

Next, note that the $\mathfrak{Rai}$ algorithm requires finding at each time t a signaling mechanism that is persuasive for *all* distributions in a neighborhood around the empirical distribution. The following result shows that this is a simple computational task requiring a polynomial running time. Thus, the result establishes the $\mathfrak{Rai}$ algorithm's computational efficiency.

Proposition 3. The $\mathfrak{Rai}$ algorithm requires solving at each time a linear program with size polynomial in $|\Omega|$ and $|A|$.

Proof. To see the efficiency of the $\mathfrak{Rai}$ algorithm, note that at each time t the algorithm has to solve the optimization problem $\max_\sigma\{V(\gamma_t, \sigma) : \sigma \in \mathsf{Pers}(\mathcal{B}_t)\}$. Since $\mathcal{B}_t = \mathsf{B}_1(\gamma_t, \epsilon_t)$ is an ℓ_1-ball of radius ϵ_t, it is a convex polyhedron with at most $|\Omega| \cdot (|\Omega| - 1)$ vertices. (These vertices are all of the form $\gamma_t + \frac{\epsilon_t}{2}(e_\omega - e_{\omega'})$, where e_ω is the belief that puts all its weight on ω.) By the linearity of the obedience constraints and the convexity of $\mathcal{B}_t$, it follows that $\mathsf{Pers}(\mathcal{B}_t)$ is obtained by imposing the obedience constraints at priors corresponding to each of these vertices. Since there are $\mathcal{O}(|\Omega| + |A|^2)$ obedience constraints for each distribution, we obtain that the optimization problem is a polynomially-sized linear program, and hence can be solved efficiently. □

Having addressed the persuasiveness and the computational efficiency of the $\mathfrak{Rai}$ algorithm, we devote the rest of the chapter to analyzing its regret. To do this, we first take a digression to define (and bound) the cost of robust persuasion in static persuasion problems. Armed with this result, we then characterize the algorithm's regret in Section 3.5.

3.5 Regret Analysis

We now return to the regret analysis of the online persuasion setting. The regret bounds we establish in this section make critical use of the characterization of the cost of robust persuasion from the preceding section.

Our main result establishes a upper bound on the regret of the $\mathfrak{Rai}$ algorithm in instances satisfying the regularity conditions. While p_0 appears in our regret bound, it is not required by the $\mathfrak{Rai}$ algorithm for its operation.

Theorem 2. *Suppose the instance $\mathcal{I}$ satisfies the regularity condition. For $t \in [T]$, let $\epsilon_t = \min\{\sqrt{\frac{|\Omega|}{t}}(1 + \sqrt{\Phi \log T}), 2\}$ with $\Phi > 0$. Then, for all $\mu^* \in \mathcal{B}_0$, with probability at least $1 - T^{1-\frac{3\Phi\sqrt{\Omega}}{56}} - T^{-8\Phi|\Omega|}$, the $\mathfrak{Rai}$ algorithm satisfies*

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) \leq 2\left(\frac{20}{p_0^2 D} + 1\right)\left(1 + \sqrt{|\Omega| T}(1 + 2\sqrt{\Phi \log T})\right).$$

In particular, the regret is of order $\mathcal{O}\left(\frac{\sqrt{\Omega}}{p_0^2 D}\sqrt{T \log T}\right)$ with high probability.

The central step in the proof is the following decomposition of the regret, established in Lemma 6 in Appendix B.2.1:

$$\begin{aligned}\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) \leq & \sum_{t \in [T]} \mathsf{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \|\mu^* - \gamma_t\|_1)) + \sum_{t \in [T]} \mathsf{Gap}(\gamma_t, \mathsf{B}_1(\gamma_t, \epsilon_t)) \\ & + \sum_{t \in [T]} \|\mu^* - \gamma_t\|_1 + \sum_{t \in [T]} \left(\mathbf{E}_{\mu^*}[v(\omega_t, a_t) | h_t] - v(\omega_t, a_t)\right).\end{aligned}$$

Observe that on the event $\{\mu^* \in \cap_{t \in [T]} \mathcal{B}_t\}$, we have $\|\mu^* - \gamma_t\|_1 \leq \epsilon_t$. Thus, on this event, the first two terms on the right-hand side of the preceding inequality capture the cost of persuading robustly for all distributions in an ℓ_1-ball of radius ϵ_t around the distribution μ^* and its estimate γ_t. Moreover, the third term represents the estimation error between μ^* and γ_t. Together with Proposition 1, we thus obtain that the first three terms are of order $\sum_{t \in [T]} \epsilon_t = O(\sqrt{T \log T})$. Finally, the last term, which captures the randomness in the sender's payoff, is also of the same order due to a simple application of the Azuma-Hoeffding inequality. The details are provided in Appendix B.2.1.

3.5.1 Lower Bound

In this section, we show that our regret upper bound in Theorem 2 is essentially tight with respect to the parameters D, T (up to a lower order $\sqrt{\log T}$ factor). We also show that the inverse polynomial dependence on p_0, the smallest probability of states, is necessary though the exact order of the dependence on p_0 is left as an interesting open question.

Theorem 3. *For the instance $\mathcal{I}_0$ and distribution $\mu^* \in \mathcal{B}_0$ considered in Proposition 2, there exists a $T_0 > 0$ such that for any $T \geq T_0$ and any β_T-robustly persuasive algorithm $\mathfrak{a}$ the following holds with probability at least $\frac{1}{3} - 2\beta_T$:*

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{a}, T, \mu^*) = T \cdot \mathsf{OPT}(\mu^*) - \sum_{t \in [T]} v(\omega_t, a_t) \geq \frac{\sqrt{T}}{32 D p_0}.$$

We provide a sketch here. First the regret can be split into two terms:

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{a}, T, \mu^*) = T \cdot \mathsf{OPT}(\mu^*) - \sum_{t \in [T]} V(\mu^*, \sigma^{\mathfrak{a}}[h_t]) + \sum_{t \in [T]} V(\mu^*, \sigma^{\mathfrak{a}}[h_t]) - \sum_{t \in [T]} v(\omega_t, a_t)$$

Let $\mathcal{E}_T(\mu)$ be the event under which the signaling mechanism $\sigma^{\mathfrak{a}}[h_t]$ chosen by the algorithm $\mathfrak{a}$ after any history $h_t \in \mathcal{E}_T(\mu)$ is persuasive for the distribution μ. Hence on the event $\mathcal{E}_T(\mu^*) \cap \mathcal{E}_T(\bar{\mu}_1) \cap \mathcal{E}_T(\bar{\mu}_2)$, the signaling mechanism $\sigma^{\mathfrak{a}}[h_t]$ is persuasive for all three distributions $\mu^*, \bar{\mu}_1 = \mu^* + \frac{\epsilon}{2}(e_1 - e_2)$ and $\bar{\mu}_2 = \mu^* + \frac{\epsilon}{2}(e_2 - e_1)$. From Proposition 2, we have that on this event, the first term, which is the sender's expected loss, is no less than $T \cdot \mathsf{Gap}(\mu^*, \{\mu^*, \bar{\mu}_1, \bar{\mu}_2\})$. We lower bound the second term using the Azuma-Hoeffding inequality. The remaining step is to show that the probability of the event $\mathcal{E}_T(\mu^*) \cap \mathcal{E}_T(\bar{\mu}_1) \cap \mathcal{E}_T(\bar{\mu}_2)$ does not vanish as T goes to infinity, which follows from robust persuasiveness of the algorithm $\mathfrak{a}$ and careful choice of ϵ. The details are provided in Appendix B.2.2.

3.6 Conclusion

We studied a repeated Bayesian persuasion problem where the distribution of payoff-relevant states is unknown to the sender. The sender learns this distribution from observing state realizations while making recommendations to the receiver. We propose the $\mathfrak{Rai}$ algorithm which persuades robustly and achieves $\mathcal{O}(\sqrt{T \log T})$ regret against the optimal signaling mechanism under the knowledge of the state distribution. To match this upper-bound, we construct a persuasion instance for which no persuasive algorithm achieves regret better than $\Omega(\sqrt{T})$. Taken together, our work precisely characterizes the value of knowing the state distribution in repeated persuasion.

While in our analysis we have assumed that the receiver's utility is fixed across time periods, our model and the analysis can be easily extended to accommodate heterogeneous receivers, as long as the sender observes the receiver's type prior to making the recommendation, and the cost of robustness Gap can be uniformly bounded across different receiver types. More interesting is the setting where the sender must persuade a receiver with an unknown type. In such a setting, assuming the sender cannot elicit the receiver's type prior to making the recommendation, the sender makes a menu of action recommendations (one for each receiver type). It can be shown the complete information problem in this setting corresponds to public persuasion of a group of receivers with no externality, which is known to be a computationally hard linear program with exponentially many constraints (Dughmi and Xu, 2017). Consequently, our algorithm ceases to be computationally efficient. Nevertheless, our results imply that the algorithm continues to maintain the $O(\sqrt{T \log T})$ regret bound.

Our main technical contribution is using the characterization of the cost of robust persuasion in Chapter 2 to perform a tight regret analysis for the online learning problem. Though our characterization of the cost of robust persuasion relies on the specifics of the persuasion setting, the regret analysis is more agnostic to the setting. Given this, we believe our approach can be extended to other online linear programming settings as long as one can obtain a characterization of the corresponding cost of robustness.

Chapter 4

Markov Persuasion Processes with Endogenous Agent Beliefs

In this chapter[1], we explore another information design problem in a dynamic setting motivated by online platforms. Unlike Chapter 3, where the state is independently drawn at each round, we shift our attention to a setting where the state is determined stochastically by the current state and the receiver's action. In such a setting, we analyze the sender's persuasion problem, accounting for the possibility that the receivers may have limited (partial) historical information. We apply a similar robustness approach to construct a simple signaling mechanism with good performance guarantees.

4.1 Introduction

Many platform services and markets involve *freelance* service providers (drivers in ride-hailing markets, hosts in accommodation services, etc.) who make voluntary decisions on when and where to provide their services, at what quality, and at which price. Often, the participants of these platforms lack all the necessary information about the system (overall demand, demand imbalances, etc.) to act optimally. Given that the platform is typically better informed, many of them provide recommendations to the participants on their actions in the system. For example, ride-hailing

[1]A preliminary version of this work appeared as an extended abstract Iyer et al. (2023)

platforms such as Uber and Lyft share real-time demand information with drivers to enable them to make repositioning decisions.

To study such settings, we consider a model of Markovian persuasion (Wu et al., 2022; Gan et al., 2022; Ely, 2017; Farhadi and Teneketzis, 2022; Lehrer and Shaiderman, 2022), where a single long-lived sender seeks to persuade a stream of short-lived receivers by sharing information about a payoff-relevant state. The state transitions are assumed to be Markovian, where the system's next state is fully determined (stochastically) by the current state and the receiver's action. The state of the system is observable to the sender but not to the receivers. In line with the literature (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019; Dughmi and Xu, 2021), we assume that the sender *commits* to a signaling mechanism, which recommends an action based on the current state and the history of the process. The receivers are myopic, and choose an action that maximizes their expected payoff under their posterior beliefs given the recommendation. The goal of the sender is to maximize the long-run average reward.

In such settings, given the underlying Markovian dynamics, the effectiveness of persuasion is impacted by the receivers' knowledge of the history. Past analyses of Markovian persuasion settings either assume the receivers have exogenous beliefs (Gan et al., 2022; Wu et al., 2022), or assume that the receivers have no information about the history (Lingenbrink and Iyer, 2019; Anunrojwong et al., 2022). (Few works assume that the receivers have information about past signals (Ely, 2017; Farhadi and Teneketzis, 2022; Renault et al., 2017; Ashkenazi-Golan et al., 2022) or past states (Lehrer and Shaiderman, 2022), but restrict themselves to the setting where the state evolves independently of the receivers' actions.) However, from a practical perspective, both these assumptions are restrictive. In particular, the participants in a platform typically have beliefs that are influenced by their past experiences therein. Furthermore, these participants are likely to have some limited information of the history. For instance, in a ride-hailing setting, a driver, in addition to knowing the typical demand patterns at different locations, may also have some stale historical information about demand at a particular location from having dropped off a rider there earlier. Sometimes, drivers' historical information can be outdated, owing to various factors such as infrequent visits to the airport and the rapid evolution of traffic patterns and rider demand. In order to ensure that the driver heeds a recommendation to move to that location, a platform must take into account the existence of such limited historical information.

In this markovian-persuasion, we seek to understand the sender's persuasion problem when receivers may have limited information about the history. To do this, we define the notion of an *information model*, which specifies how each receiver's belief (prior to receiving a recommendation) is related to the history of the process. In addition to the full-history information model Φ_{full} (where the receivers observe the entire history) and the no-history information model Φ_{no} (where receivers have no historical information), we consider a sequence Φ_ℓ of partial-history information models where each receiver observes the history of the system with an ℓ periods lag, for some fixed $\ell \geq 1$. These partial-history information models provide lower-bounds on the sender's payoff in more complex information models, and thus serve as a standard for comparison.

Our main contributions are as follows:

1. **Establishing benchmarks.** We begin with the analysis of the two benchmark information models, i.e., no-history and full-history information. We prove that, under the no-history information model, the optimal signaling mechanism is history-independent, whereas in the full-history information model, the optimal mechanism depends on the current state as well as the previous state-action pair. Consequently, these characterizations allow us to formulate the sender's persuasion problem as a succinct linear program under both information models.

2. **Ordering and solving partial-history information models.** We then analyze the sequence of partial-history information models Φ_ℓ. We show that the sender's optimal payoff under any such model is less than that under the no-history information model Φ_{no}, but greater than that under full-history information model Φ_{full}. Moreover, we show that the sender's optimal payoff increases as the lag ℓ increases. We then identify sufficient conditions on the model primitives that ensure that the sender's optimal payoff in the two benchmark information models is equal and hence partial information about the history on the receivers' part has no adverse impact on the sender's payoff. Nevertheless, we show that the analysis of the persuasion problem in the

information models Φ_ℓ presents technical intricacies that leave open even the question of the existence of an optimal signaling mechanism. Due to this, we study the sender's problem restricting attention to signaling mechanisms that only depend on a fixed length of past history. Here, we show that the sender's problem can be written as a *bilinear* program, whose size grows exponentially in the lag ℓ. This suggests that solving to optimality the sender's problem in the partial-history information models can be computationally challenging as well.

3. **Simple and approximately optimal persuasion.** Due to the complexity of solving the persuasion problem optimally under partial-history information models, we take an alternative approach and ask whether simple *history-independent* mechanisms can achieve approximately optimal payoffs while simultaneously being persuasive under limited historical information. Using the underlying Markovian dynamics and a robust persuasion approach in Chapter 2, we answer the preceding question positively. In particular, we construct a history-independent signaling mechanism whose payoff is close to the optimal payoff under the no-history information model, and which is simultaneously persuasive in information models Φ_ℓ for all large enough ℓ. To obtain this construction, we prove an extension of the splitting lemma (Aumann et al., 1995b; Kamenica and Gentzkow, 2011) to Markovian settings.

Our results contribute to the literature on information design and persuasion in dynamic settings, with endogenous beliefs of the receivers. From a theoretical perspective, our results establish the effectiveness of simple history-independent signaling mechanisms in such settings. Furthermore, our results highlight the importance of robustness in designing signaling mechanisms; when participants in a platform may have limited historical information, a simple but robust signaling mechanism can achieve good performance while being persuasive.

4.2 Related Work

Our work is related to a branch of literature on Bayesian persuasion (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2016, 2019; Dughmi and Xu, 2021) in dynamic settings. Specifically, our work relates to the following streams of literature.

Markov persuasion. Several papers (Gan et al., 2022; Wu et al., 2022; Bacchiocchi et al., 2024) have studied the Markov persuasion problem where the system's next state is updated on the current state and the receivers' performed actions. Gan et al. (2022) study an infinite-horizon setting where the sender privately observes the payoff-relevant parameter and advises actions the receiver to take over time. The external parameter is drawn from a known prior distribution to both the sender and the receiver. In particular, they consider two types of receivers: myopic (short-lived) and farsighted (optimizes cumulative rewards over time). In both settings, they assume that the signal at each round does not depend on historical realizations and the receiver cannot observe history. However, in our setting, the receivers may observe the historical realizations and their beliefs are endogenous determined by the sender's signaling strategy.

Wu et al. (2022) focus on a finite-horizon setting where the sender persuades a stream of myopic receives in an uncertain Markovian environment. At each round, the sender observes an outcome that is drawn from a prior distribution that is only known to the receiver. The sender aims to maximize the long-term rewards without the knowledge of the sender's utility functions, prior distributions, and the Markov transition kernels. The authors use a reinforcement learning approach to design an online learning algorithm that achieves $\mathcal{O}(\sqrt{T})$ regret. Similar to the work (Gan et al., 2022), the receiver's prior belief is exogenous and the sender is restricted to history-independent signaling strategies. A recent paper by Bacchiocchi et al. (2024) extends the work to the setting where the sender has no knowledge about transitions, prior distributions over outcomes, sender's stochastic rewards, and receivers' ones. However, they allow that the signaling mechanism can violate the persuasiveness constraint and guarantee the sublinear violation.

Another relevant area of study is dynamic Bayesian persuasion (Ely, 2017; Renault et al., 2017; Farhadi and Teneketzis, 2022), where the state evolution follows a Markov process independent of the receiver's actions. Ely (2017) study a dynamic bank runs game with binary states and actions. The bank sends public and private signals to the agents over time to prevent a run on deposits. They design the optimal policy by selecting minimal correlated signals. Renault et al. (2017) also consider a setting with two states and two actions, showing that a myopic strategy is optimal for the sender in two-state scenario. They further show that with more than two states, the optimal strategy need not to be myopic and provide a sufficient condition on the Markov process for the optimal strategy being myopic. Ashkenazi-Golan et al. (2022) extends to a finite number of actions and characterizes the sender's optimal signaling strategy in the limit, as the length of each round goes to zero. Recent paper by Farhadi and Teneketzis (2022); Lehrer and Shaiderman (2022) study the optimal signaling strategy in various dynamic settings. Farhadi and Teneketzis (2022) study the optimal information disclosure in two-state Markov chain jump games where the sender aims to delay the receiver's detection of the jump from a good state to a bad state. Lehrer and Shaiderman (2022) study the Markov persuasion problem with stochastic revelation, assuming that at each round, after the receiver takes an action, the true state and her payoff, are revealed to her with a positive probability. In contrast to these works, our model assumes that the receiver's actions affect the state evolution.

Other papers also explore dynamic persuasion in various application contexts. Bernasconi et al. (2022) study the sequential persuasion problem between a sender and a farsighted receiver within a tree structure. They assume that in each round, the sender privately observes the state realization, which is drawn from a distribution unknown to both the sender and the receiver. They demonstrate that without knowledge of the state distribution, no algorithm can consistently be persuasive at each round with high probability.Additional related works include (Li and Norman, 2021; Wu, 2023; Board and Lu, 2018; Orlov et al., 2020; Bizzotto et al., 2021; Alizamir et al., 2020).

As examples of Markovian persuasion where the receivers have no information about the history, Lingenbrink and Iyer (2019) study the information-sharing problem in a single-server queue offering services at a fixed price. The service provider observes the queue and shares the information with the delay-sensitive Poisson arriving customers. The authors formulate the service providers' decision problem of maximizing the revenue as an infinite linear program. A similar approach is taken by Anunrojwong et al. (2022) to study information design to manage congestion in queues.

Robust persuasion. As our proposed signaling mechanism relies on the robust persuasion framework, our work is also related to robust persuasion. Chapter 3 study a repeated Bayesian persuasion problem where neither the sender nor the receiver knows the payoff-relevant state distribution. They propose a robust signaling mechanism that recommends persuasive recommendations at all rounds with high probability, achieving $\mathcal{O}(\sqrt{T \log T})$ regret. For our robustness results, we extend their approach to settings where the receivers' beliefs are endogenous. Kosterina (2022) also assumes an unknown prior distribution, where only the set containing the true prior is known to the sender. The authors characterize the support of the signal realization and show that the optimal signal forms a hyperbola.

Other relevant works (Ui, 2022; Hu and Weng, 2021; Dworczak and Pavan, 2022) study the robust persuasion problem in the context of the receiver's external private information. They leverage the robustness approach facing the uncertainty of the receivers' private information. While these studies focus on static persuasion models with robustness to exogenous receiver beliefs, our work addresses dynamic robust persuasion with endogenous receivers' beliefs in sequential setups. Finally, there are also studies of robust persuasion with respect to receiver payoffs (e.g., (Babichenko et al., 2022)), though these are less relevant to the present work.

4.3 Model

Informally, we study a dynamic persuasion setting between a long-lived sender and a stream of short-lived receivers where the underlying payoff-relevant state evolves as a Markov persuasion process. At each time t, a new receiver arrives to whom the sender, after observing the current state, recommends an action. The receiver then chooses an action, possibly different from the sender's recommendation, after which the state updates according to a Markov transition kernel which is common knowledge among the sender and the receivers. Each receiver seeks to maximize the expected utility with respect to her (posterior) beliefs, given the sender's recommendation and her (partial) information about the history of the process. The sender's problem, our object of investigation, is to decide how to recommend actions that maximize her long-run average payoff. We now describe this model formally.

We consider a sequential setting where at each time $t \in \mathbb{Z}$, the payoff-relevant state is given by $\bar{\omega}_t \in \Omega$. Here Ω is a finite set of states. We denote the *signal* shared by the sender as $\bar{s}_t \in S$, and the action chosen by the receiver as $\bar{a}_t \in A$, where again S is a finite set of signals and A is a finite set of actions. (We describe how the sender shares the signals and how the receivers choose their actions in detail below.) The state evolution is Markovian given the receiver's action: $\mathbf{P}(\bar{\omega}_t = \omega|\bar{h}_t) = p(\omega|\bar{\omega}_{t-1}, \bar{a}_{t-1})$ for each $\omega \in \Omega$, where $\bar{h}_t$ denotes the history at time t, i.e., the infinite sequence of state, action and signals up to (but not including) time t. Here, $p\colon \Omega \times \Omega \times A \to [0,1]$ is a stationary Markovian transition kernel, with $p(\omega'|\omega, a)$ denoting the probability of the state transitioning from $\bar{\omega}_{t-1} = \omega$ to $\bar{\omega}_t = \omega'$ after the receiver takes action $\bar{a}_{t-1} = a$. At the end of each time t, the corresponding receiver obtains a payoff given by $u(\bar{\omega}_t, \bar{a}_t) \in \mathbb{R}$, whereas the sender obtains a reward given by $v(\bar{\omega}_t, \bar{a}_t) \in [0,1]$. We note that the transition kernel, the receivers' payoffs and the sender's reward do not directly depend on the signals.

4.3.1 Signaling Mechanisms

We assume that at each time t the sender observes the history $\bar{h}_t$ and the current state $\bar{\omega}_t$. On the other hand, the receiver at time t does not observe the current state, but, as we discuss later, may have some information about the history. To convey payoff-relevant information about the state at each time t, the sender shares a *private* signal $\bar{s}_t$ to the corresponding receiver. In particular, the sender commits to sharing these signals using a *signaling mechanism*, which in general, maps the history $\bar{h}_t$ and the state $\bar{\omega}_t$ at any time t to a signal $\bar{s}_t$. However, we circumscribe the class of signaling mechanisms in the following ways. First, we restrict our attention to signaling mechanisms that depend only on a finite part of the history at each time. While this assumption is primarily motivated by practical concerns, it also allows us to avoid some technical issues in defining the sender's long-run average payoff if the signaling mechanism depends on the infinite history. Second, given our focus on stationary environments, we study signaling mechanisms that do not directly depend on the calendar time. Third, we assume that the signal at each time t depends only on the historical state-action pairs, and not on the past signals. This assumption ensures that we do not implicitly induce dependence on the infinite history via past signals. Finally, we focus on *direct signaling mechanisms* (Bergemann and Morris, 2019) where the sender shares signals that are action recommendations, i.e., $S = A$.

Given our assumption that signals are private, it follows by the revelation principle (Ely, 2017) that considering direct signaling mechanisms is without loss of generality. Further, it is sufficient to restrict our attention to direct signaling mechanisms that are *persuasive*, i.e., ones where the action recommendations are optimally adopted by the receivers. In such settings, the information in past signals is already contained in the past actions, and hence the assumption that the signals only depend on past state-action pairs is not restrictive. Thus, the main restrictive assumption we make is that the signals only have finite history dependence.

Before formalizing the preceding discussion, we introduce some notation to simplify some cumbersome expressions. We let $\mathcal{X} = \Omega \times A$ denote the set of state-action pairs, and we denote a generic element of $\mathcal{X}$ by $x = (\omega, a)$. Thus, $\bar{x}_t = (\bar{\omega}_t, \bar{a}_t) \in \mathcal{X}$ denotes the state-action pair at time t, and $p(\omega'|x)$ with $x = (\omega, a)$ stands for $p(\omega'|\omega, a)$. Next, for any $k \geq 1$ and at any time t, a *slice* of history $\bar{h}_t^k$ of length k describes the sequence of states-action pairs in the past k time periods: $\bar{h}_t^k = (\bar{x}_{t-k}, \ldots, \bar{x}_{t-1}) \in \mathcal{X}^k$. We denote a generic element of $\mathcal{X}^k$ by $h^k = (x_{-k}, \ldots, x_{-1})$. Finally, we let $\mathcal{X}^0$ denote the singleton set consisting of the unique (empty) slice of history of length zero.

A signaling mechanism is a mapping $\sigma \colon \mathcal{X}^k \times \Omega \to \Delta(A)$ (for some $k \geq 0$) that specifies for each $h^k \in \mathcal{X}^k$ and $\omega \in \Omega$, the probability $\sigma(a|h^k, \omega)$ with which the sender shares the signal $\bar{s}_t = a \in A$ if the (slice of) history is $\bar{h}_t^k = h^k$ and the current state is $\bar{\omega}_t = \omega$. We let Σ_k denote the set of all signaling mechanisms that depend only on history slices of length k, and let $\Sigma = \cup_{k \geq 0} \Sigma_k$. The set Σ_0 contains the signaling mechanisms that do not depend on the history.

4.3.2 Beliefs and Persuasiveness

Next, we describe the notion of persuasiveness as applied to signaling mechanisms. To do this, we need to model the receivers' beliefs about the history of the process, which in general depends endogenously on how much information they have about the past. We capture this endogenous level of historical information through the concept of an *information model* (see Section 4.3.4). However, to develop our concepts, we will initially consider the receiver's prior beliefs as exogenously specified.

Suppose the sender commits to a signaling mechanism $\sigma \in \Sigma_k$ for some $k \geq 0$.

Fix a time t, and let the corresponding receiver's belief over the history $\bar{h}_t$ and the current state $\bar{\omega}_t$ (prior to receiving any signal) be denoted by ϕ_t. Then, upon receiving an action recommendation $\bar{s}_t = a$, the receiver's posterior belief that $\bar{\omega}_t = \omega$ can be found using Bayes' rule as

$$F(\omega|a;\phi_t,\sigma) = \frac{\sum_{h^k} \phi_t(h^k,\omega)\sigma(a|h^k,\omega)}{\sum_{\omega'} \sum_{h^k} \phi_t(h^k,\omega')\sigma(a|h^k,\omega')}.$$

Here, $\phi_t(h^k,\omega)$ denotes the receiver's marginal belief that the history slice of length k is $\bar{h}_t^k = h^k \in \mathcal{X}^k$ and the state is $\bar{\omega}_t = \omega$. The receiver then chooses an action that maximizes their expected utility under their posterior belief $F(\cdot|a;\phi_t,\sigma)$. We say the signaling mechanism σ is persuasive w.r.t. the belief ϕ_t, if the recommended action $\bar{s}_t = a$ is optimal for the receiver, i.e., the following inequality holds:

$$\sum_{\omega} F(\omega|a;\phi_t,\sigma)\ \partial u(\omega,a,a') \geq 0, \text{ for all } a,a' \in A,$$

where $\partial u(\omega,a,a') \triangleq u(\omega,a) - u(\omega,a')$ denotes the incremental payoff for the receiver for choosing an action $a \in A$ over action $a' \in A$ at state $\omega \in \Omega$. The inequality states that the receiver's expected utility with the action a is higher than that with a' when action a is recommended.

More generally, let $\Phi = \{\phi_t : t \in \mathbb{Z}\}$ denote the sequence of receivers' beliefs at each time $t \in \mathbb{Z}$. For any such sequence Φ, the set $\mathsf{Pers}(\Phi)$ of persuasive signaling mechanisms contains all signaling mechanisms σ that are persuasive w.r.t. ϕ_t for each $t \in \mathbb{Z}$. We note that the set $\mathsf{Pers}(\Phi)$ is non-empty, since the mechanism that recommends the receivers' preferred action at each state is persuasive for sequence Φ.

4.3.3 Invariant Distribution

As a step towards describing the models of endogenous historical information held by the receivers, we next analyze the induced dynamics under a signaling mechanism to characterize its invariant distribution. Suppose the sender chooses a signaling mechanism $\sigma \in \Sigma_k$ for some $k \geq 0$. Assuming that all the receivers follow the sender's recommendations, the induced process dynamics can be described as a

Markov chain with state space $\mathcal{X}^k \times \Omega$; the state of this Markov chain at time t is given by $(\bar{h}_t^k, \bar{\omega}_t)$ (or simply $\bar{\omega}_t$ for $k = 0$). For $k \geq 1$, an invariant distribution $\pi \in \Delta(\mathcal{X}^k \times \Omega)$ of this chain satisfies the following balance equations:

$$\sum_{x_{-k} \in \mathcal{X}} \pi\left((x_{-k}, h^{k-1}), \omega\right) \sigma(a|(x_{-k}, h^{k-1}), \omega) p(\omega'|\omega, a) = \pi\left((h^{k-1}, \omega, a), \omega'\right). \quad (4.1)$$

for each $h^{k-1} = (x_{-(k-1)}, \dots, x_{-1}) \in \mathcal{X}^{k-1}$, $(\omega, a) \in \mathcal{X}$ and $\omega' \in \Omega$. Here, the left-hand side expression gives the probability that the state $(\bar{h}_{t+1}^k, \bar{\omega}_{t+1})$ of the chain equals $\left((h^{k-1}, \omega, a), \omega'\right)$ after a Markovian transition if the state $(\bar{h}_t^k, \bar{\omega}_t)$ is distributed as π and the receiver at time t follows the sender's recommendation. The equation states that for an invariant distribution, this probability must be the same as that under π itself. For $k = 0$, the balance equation for an invariant distribution $\pi \in \Delta(\Omega)$ given by

$$\sum_{(\omega_{-1}, a_{-1}) \in \mathcal{X}} \pi(\omega_{-1}) \sigma(a_{-1}|\omega_{-1}) p(\omega|\omega_{-1}, a_{-1}) = \pi(\omega), \quad \text{for all } \omega \in \Omega.$$

Since the state of the induced Markov chain includes the receivers' actions, in general there might be multiple invariant distributions π corresponding to a signaling mechanism. (As a trivial example, consider a setting with $\Omega = \{0\}$, $A = \{0, 1\}$ and a receiver who is indifferent between the two actions. Let $\sigma \in \Sigma_1$ be a signaling mechanism that recommends $\bar{s}_t = 0$ if $(\bar{\omega}_{t-1}, \bar{a}_{t-1}, \bar{\omega}_t) = (0, 0, 0)$ and recommends $\bar{s}_t = 1$ if $(\bar{\omega}_{t-1}, \bar{a}_{t-1}, \bar{\omega}_t) = (0, 1, 0)$. Then, assuming all the receivers follow the recommendations, any distribution over $\mathcal{X} \times \Omega$ is an invariant distribution under σ.) Hereafter, in cases where there are multiple invariant distributions, we focus on the one under which the sender's long-run average reward is maximized (with ties broken arbitrarily). We denote such a distribution by $\mathsf{Inv}(\sigma)$. Note that this assumption is aligned with the notion of sender-preferred equilibrium common in the persuasion literature (Kamenica and Gentzkow, 2011).

Below, we abuse the notation slightly by letting $\pi = \mathsf{Inv}(\sigma)$ also denote the distribution of the stationary Markov process induced under a signaling mechanism $\sigma \in \Sigma_k$, i.e., the distribution of the entire history $\bar{h}_t$ at each time t, assuming all the receivers follow their recommendations. Furthermore, for any $\ell \geq 1$, we let $\pi(h^\ell)$

denote the (marginal) distribution of a slice of history $\bar{h}_t^\ell$.

4.3.4 Modeling Receivers' Endogenous Information

We now formally describe the notion of an *information model*, which captures the receivers' endogenous information about the historical evolution of the process. We consider two benchmark settings, one where each receiver fully observes the history, and the other where the receivers have no information about the history. In addition, we consider a sequence of settings where the receivers have partial information about the history.

In general, when receivers have information about the history, their belief sequence $\Phi = \{\phi_t : t \in \mathbb{Z}\}$ itself will depend on the process. The nature of this dependence is determined by the amount of information the receivers have about the past.

1. **Full-history information model:** To motivate the notion, fix a signaling mechanism $\sigma \in \Sigma$ and consider first the setting where at each time t, the corresponding receiver has complete knowledge of the history $\bar{h}_t$. Then, the receiver's belief ϕ_t over $(\bar{h}_t, \bar{\omega}_t)$ must put all its weight on the realized value of $\bar{h}_t$. In other words, we have for all $t \in \mathbb{Z}$,

$$\mathbf{P}^\sigma\big(\phi_t = e_h \otimes p(\cdot|x_{-1}) \,\big|\, \bar{h}_t = h, \bar{\omega}_t = \omega\big) = 1, \quad \text{for all } h \in \mathcal{X}^\infty \text{ and } \omega \in \Omega,$$

   where e_h is the distribution that puts all its weight on $h = (\dots, x_{-2}, x_{-1}) \in \mathcal{X}^\infty$, and $e_h \otimes p(\cdot|x_{-1})$ encodes the fact that the receivers' belief about $\bar{\omega}_t$ comes from the resulting Markovian transition $p(\cdot|x_{-1})$. (Here, $\mathbf{P}^\sigma$ denotes the probability measure induced by the signaling mechanism σ together with the underlying Markovian dynamics, assuming that the receivers adopt the sender's recommendations.) When the preceding condition holds, we denote the resulting belief sequence $\{\phi_t : t \in \mathbb{Z}\}$ by Φ_{full} and call it the *full-history information model*.

2. **No-history information model:** At the other extreme, consider the case where the receivers have no information about the history of the process. Then, at any time t, the receiver's belief ϕ_t must be independent of the realized

history. A natural choice, motivated by the requirement of consistency, is to let each belief ϕ_t equal the invariant distribution $\mathsf{Inv}(\sigma)$. (In certain cases, this modeling assumption can be established formally. For instance, if time periods denote the Poisson arrival times of receivers to a stochastic system, then the receivers observe the system distributed as the time-average (Wolff, 1982), which equals the expectation with respect to the invariant distribution when the latter is unique.) Specifically, we have for each $t \in \mathbb{Z}$,

$$\mathbf{P}^\sigma\big(\phi_t = \mathsf{Inv}(\sigma) \otimes P \,\big|\, \bar{h}_t = h, \bar{\omega}_t = \omega\big) = 1, \text{for all } h \in \mathcal{X}^\infty \text{ and } \omega \in \Omega.$$

Here, $\mathsf{Inv}(\sigma) \otimes P$ encodes the distribution of $(\bar{h}_t, \bar{\omega}_t)$ where the history $\bar{h}_t$ is distributed as $\mathsf{Inv}(\sigma)$, and the state $\bar{\omega}_t$ is obtained from a subsequent transition from the Markov kernel P. For the setting where the preceding condition holds, we denote the belief sequence $\{\phi_t : t \in \mathbb{Z}\}$ by Φ_{no} and call it the *no-history information model*.

3. **Partial-history information models:** Between the two extremes described above lie a multitude of information models where receivers possess partial information about the process history. In such partial-history models, the belief sequence Φ would have a complex dependence on the history $\bar{h}_t$. Because a comprehensive analysis of all such models is challenging, we focus on a particular sequence of information models to capture realistic scenarios where the receivers may have some "delayed" information about the process. Such delayed information about the process could plausibly arise from the receivers having interacted with the process in the past; however, we do not formalize such repeated interactions in our model.

   Specifically, for a fixed $\ell \geq 0$, consider the setting where the receivers observe the process with an ℓ-period lag. In other words, at each time t, the receiver observes the history $\bar{h}_{t-\ell}$, i.e., all the state-action pairs before time $t-\ell$. Then, we have for each $t \in \mathbb{Z}$,

$$\mathbf{P}^\sigma\Big(\phi_t = e_{h_{-\ell}} \otimes P_\sigma^\ell \otimes P \,\Big|\, \bar{h}_t = h, \bar{\omega}_t = \omega\Big) = 1, \text{for all } h \in \mathcal{X}^\infty \text{ and } \omega \in \Omega.$$

   Here, $e_{h_{-\ell}}$ is the distribution that puts all its weight on the realization $\bar{h}_{t-\ell} =$

$h_{-\ell}$, P^{ℓ}_{σ} encodes the subsequent ℓ transitions of the process, i.e., the distribution of $(\bar{x}_{t-\ell}, \dots, \bar{x}_{t-1})$ under the signaling mechanism σ, and finally, the kernel P captures the subsequent distribution of the $\bar{\omega}_t$. When the preceding holds, we denote the resulting belief sequence $\{\phi_t : t \in \mathbb{Z}\}$ as Φ_ℓ, and call it the *partial-history information model with lag* ℓ. We note that Φ_0 is same as the full-history information model Φ_{full}.

An advantage of studying the sequence $\{\Phi_\ell\}_{\ell \geq 0}$ of information models is that they serve as a standard of comparison for other more complex information models. In particular, one can show that the sender's payoff under the information model Φ_ℓ acts as a lower-bound on her payoff in settings where the receivers only have limited, but arbitrary, information about states and action ℓ periods and further back. Thus, while we do not capture all possible partial-history information models, our choice provides a lower bound of many other information models and gives insight into the problem's fundamental difficulty, and on the influence of the lag on the sender's payoff.

4.3.5 Sender's Persuasion Problem

Finally, we are ready to formally describe the sender's persuasion problem. We focus on settings where the sender seeks to maximize the long-run average reward over the infinite horizon. Given the Markovian state evolution, this is equivalent to the sender choosing a signaling mechanism to maximize the expected rewards under the resulting invariant distribution. Formally, we denote the sender's problem under the information model Φ as

$$
\begin{aligned}
\mathsf{MPP}(\Phi) \triangleq \max_{\sigma,\pi} \quad & \mathbf{E}^{\pi}[v(\omega, a)] \\
\text{subject to,} \quad & \sigma \in \mathsf{Pers}(\Phi) \cap \Sigma, \quad \pi = \mathsf{Inv}(\sigma),
\end{aligned} \tag{4.2}
$$

and let $\mathsf{OPT}(\Phi)$ denote its optimal value. Furthermore, for $k \geq 0$, we analogously define $\mathsf{MPP}(\Phi, \Sigma_k)$ (and $\mathsf{OPT}(\Phi, \Sigma_k)$) as the sender's problem (and its optimal value) when the signaling mechanism is restricted to lie in the set Σ_k. In the preceding optimization problem, unlike a static persuasion problem, the expectation in the objective is taken with respect to the invariant distribution π which is in turn determined by the signaling mechanism σ.

Hereafter, we make the following standard *unichain* assumption (Puterman, 2014; Tsitsiklis, 2007), which is common in the analysis of average-reward Markov decision processes. To state formally, a stationary Markovian policy is a decision rule that chooses a possibly randomized action based solely on the current state. Such a policy naturally induces a Markov chain over the state space. The unichain condition requires the induced Markov chain to have a single ergodic class.

Assumption 2 (Unichain)**.** Under any stationary Markovian policy, the resulting Markov chain has a single ergodic class, i.e., it is aperiodic and irreducible.

This assumption ensures that the invariant distribution under any signaling mechanism $\sigma \in \Sigma_0$, assuming the receivers adopt the recommendations, is unique, and thus the long-run averages are independent of the initial conditions.

4.4 Benchmarking Markovian Persuasion with Historical Information

With the goal towards studying the sender's persuasion problem in general information models, we first analyze the sender's problem (4.2) under the benchmark full-history and no-history information models. As we show later, the sender's optimal payoffs in the two benchmark models provide bounds on the sender's optimal payoff in partial-history information models. Moreover, the results here set the stage for our subsequent analysis of the partial-history information models.

4.4.1 Analysis of the Benchmark Information Models

Our analysis of the benchmark information models begins with the following lemma, which establishes that in each case, there exists an optimal signaling mechanism that is fairly simple, and does not heavily depend on the history. In particular, the optimal mechanism under the no-history information model Φ_{no} is history-independent, whereas it additionally depends on the previous state-action pair under the full-history information model Φ_{full}.

Lemma 1. In the no-history information model Φ_{no}, there exists an optimal signaling mechanism σ that is history-independent, i.e., $\sigma \in \Sigma_0$. Similarly, under the full-history information model Φ_{full}, there exists an optimal signaling mechanism $\sigma \in \Sigma_1$, which depends only on the current state and the previous state-action pair.

The proof uses the underlying Markovian dynamics of the process, and is provided in Appendix C.1 (as are the proofs of all results in this section). Most importantly, the lemma allows us to show that the sender's problem (4.2) in the benchmark settings can be formulated as a polynomially-sized linear program. This LP formulation plays a key role in particular in Theorem 4, where we characterize sufficient conditions under which the sender's optimal payoffs in the two benchmark information models are equal.

To obtain the LP formulations, we recall that for any signaling mechanism $\sigma \in \Sigma_1$, assuming that the receivers follow the recommendations, the induced process dynamics can be described as a Markov chain with states $(\bar{x}_{t-1}, \bar{\omega}_t) \in \mathcal{X} \times \Omega$. Similarly, for any $\sigma \in \Sigma_0$, if all receivers follow the recommendations, the states $\bar{\omega}_t$ form a Markov chain. We now introduce some notation to unify the presentation. First, define $\mathcal{X}_{\mathsf{full}} \triangleq \mathcal{X}$ and $\mathcal{X}_{\mathsf{no}} \triangleq \mathcal{X}^0$. For each $(\omega, a) \in \mathcal{X}$ and $\omega' \in \Omega$, let $p_{\mathsf{no}}(x, \omega'|\omega, a) \triangleq p(\omega'|\omega, a)$ for all $x \in \mathcal{X}_{\mathsf{no}}$ and $p_{\mathsf{full}}(x, \omega'|\omega, a) \triangleq p(\omega'|\omega, a)\mathbf{I}\{(\omega, a) = x\}$ for all $x \in \mathcal{X}_{\mathsf{full}}$. For $i \in \{\mathsf{full}, \mathsf{no}\}$, consider the following linear program $\mathsf{LP}(i)$, with variables $z(x, \omega, a)$ for each $x \in \mathcal{X}_i$ and $(\omega, a) \in \mathcal{X}$:

$$
\begin{aligned}
\mathsf{LP}(i) \triangleq \max_{z \geq 0} \quad & \sum_{x \in \mathcal{X}_i} \sum_{\omega \in \Omega} \sum_{a \in A} z(x, \omega, a) v(\omega, a) \\
& \sum_{\omega \in \Omega} z(x, \omega, a) \partial u(\omega, a, a') \geq 0, \quad \text{for all } a, a' \in A \text{ and } x \in \mathcal{X}_i \\
& \sum_{\hat{x} \in \mathcal{X}_i} \sum_{\hat{\omega} \in \Omega} \sum_{\hat{a} \in A} z(\hat{x}, \hat{\omega}, \hat{a}) p_i(x, \omega | \hat{\omega}, \hat{a}) = \sum_{a \in A} z(x, \omega, a), \text{ for all } x \in \mathcal{X}_i, \, \omega \in \Omega \\
& \sum_{x \in \mathcal{X}_i} \sum_{\omega \in \Omega} \sum_{a \in A} z(x, \omega, a) = 1.
\end{aligned} \tag{4.3}
$$

The following proposition, whose proof is provided in Appendix C.1.1, relates each linear program $\mathsf{LP}(i)$ to the corresponding problem $\mathsf{MPP}(\Phi_i)$.

Proposition 4. For $i \in \{\mathsf{full}, \mathsf{no}\}$, the sender's problem $\mathsf{MPP}(\Phi_i)$ can be equivalently formulated as the corresponding linear program $\mathsf{LP}(i)$. In particular, let z_i^*

be any optimal solution of $\mathsf{LP}(i)$. Then, the signaling mechanism σ_i is optimal for $\mathsf{MPP}(\Phi_i)$, where for each $x \in \mathcal{X}_i$, $\omega \in \Omega$ and $a \in A$,

$$\sigma_i(a|x,\omega) = \frac{z_i^*(x,\omega,a)}{\sum_{a'} z_i^*(x,\omega,a')}$$

if the denominator is positive; otherwise, σ_i recommends any action that is optimal for the receivers at state ω.

To interpret the linear program $\mathsf{LP}(i)$, we focus on the full-history information model Φ_{full}, and consider a signaling mechanism $\sigma \in \Sigma_1 \cap \mathsf{Pers}(\Phi_{\mathsf{full}})$. Then, writing the balance equations (4.1) for the invariant distribution $\pi \in \mathsf{Inv}(\sigma)$, we obtain

$$\sum_{x\in\mathcal{X}} \pi(x,\omega)\sigma(a|x,\omega)p(\omega'|\omega,a) = \pi\left(\omega,a,\omega'\right), \quad \text{for all } (\omega,a) \in \mathcal{X} \text{ and } \omega' \in \Omega.$$

Since this equation is bilinear in π and σ, we introduce the variables $z(x,\omega,a) \triangleq \pi(x,\omega)\sigma(a|x,\omega)$, and note that z constitutes the joint distribution of two consecutive state-action pairs $(\bar{x}_{t-1},\bar{x}_t)$ under π. The first equality in $\mathsf{LP}(\mathsf{full})$ follows from the balance equations, and the second equality follows because z is a distribution. The fact that σ is persuasive under Φ_{full} yields the inequality. Thus, the variables $z(x,\omega,a)$ defined above are feasible for $\mathsf{LP}(\mathsf{full})$, with the objective value equal to the sender's payoff under σ. The proposition follows upon showing the converse, i.e., for any z feasible for $\mathsf{LP}(\mathsf{full})$, there exists a signaling mechanism $\sigma \in \Sigma_1 \in \mathsf{Pers}(\Phi_{\mathsf{full}})$ with sender's payoff equal to the objective value under z (and a similar statement for the Φ_{no} case).

Note that the linear program $\mathsf{LP}(i)$ has $\mathcal{O}(|\mathcal{X}_i| \times |\mathcal{X}|)$ variables and $\mathcal{O}(|A|^2 \cdot |\mathcal{X}_i| + |\mathcal{X}_i| \cdot |\Omega|)$ constraints. Thus, for the no-history information model Φ_{no}, the linear program $\mathsf{LP}(\mathsf{no})$ has $\mathcal{O}(|\Omega| \cdot |A|)$ variables and $\mathcal{O}(|A|^2 + |\Omega|)$ constraints, whereas for the full-history information model Φ_{full}, the linear program $\mathsf{LP}(\mathsf{full})$ has $\mathcal{O}(|\Omega|^2 \cdot |A|^2)$ variables and $\mathcal{O}(|A|^3 \cdot |\Omega| + |\Omega|^2 \cdot |A|)$ constraints. Together with the preceding result, we thus conclude that, in addition to having an LP formulation, the sender's problem in the benchmark models can be solved efficiently. As we discuss later in Section 4.5.1, this is in stark contrast to the case under partial-history information models.

4.4.2 Ordering and Bounding Partial-history Information Models

Our next result justifies our choice of the two benchmarks, by showing that there is a natural nested order relating the different information models, with the two benchmark models occupying the extremes.

Lemma 2. For $\ell \geq 0$, we have $\mathsf{Pers}(\Phi_{\mathsf{full}}) \subseteq \mathsf{Pers}(\Phi_\ell) \subseteq \mathsf{Pers}(\Phi_{\ell+1}) \subseteq \mathsf{Pers}(\Phi_{\mathsf{no}})$ and consequently, $\mathsf{OPT}(\Phi_{\mathsf{full}}) \leq \mathsf{OPT}(\Phi_\ell) \leq \mathsf{OPT}(\Phi_{\ell+1}) \leq \mathsf{OPT}(\Phi_{\mathsf{no}})$.

Intuitively, the result follows from the fact that with less information available to the receivers, the sender's ability to persuade them improves. Formally, this result is established by showing, e.g., that any signaling mechanism that is persuasive under the model Φ_{full} remains persuasive under the model Φ_{no}, because the sender can always share additional historical information if needed. Thus, the result implies a trade-off: by choosing the optimal signaling mechanism for the model Φ_{full}, the sender can simultaneously be persuasive for all the partial-history information models Φ_ℓ, but at the cost of lower payoffs. We illustrate the magnitude of this trade-off in the following example.

Example 4.4.1. Consider a setting with $\Omega = \{0, 1\}$ and $A = \{0, 1\}$. The receivers' utility is given by $u(\omega, a) = \mathbf{I}\{\omega = a\}$, i.e., the receiver desires to match the action with the state. The sender strictly prefers the receiver choosing action $a = 1$ over action $a = 0$ in all states, i.e., $v(\omega, a) = \mathbf{I}\{a = 1\}$ for all ω. The transition probabilities are such that when taking action $a = 0$, the state remains the same with probability 0.8 and switches with probability 0.2, whereas when taking action $a = 1$, the state switches with probability 0.8 and stays the same with probability 0.2. By solving the LP formulations in the preceding section, we find that the sender's optimal payoff in the no-history information model equals $\mathsf{OPT}(\Phi_{\mathsf{no}}) = 1$, i.e., when the receivers have no historical information, the sender can persuade the receivers to always choose her preferred action $a = 1$. On the other hand, when the receivers can observe the complete history, the sender obtains a strictly lower payoff, namely $\mathsf{OPT}(\Phi_{\mathsf{full}}) = 0.52$.

Thus, the example shows that, in general, the sender's optimal payoff significantly depends on the level of historical information the receivers possess. A natural question then is whether there are conditions under which historical information does not affect the sender's ability to persuade the receivers. The following proposition characterizes one such sufficient condition.

To state the result, we need some definitions. For $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$, with $\pi = \mathsf{Inv}(\sigma)$, let $\tau_a^\sigma \triangleq \sum_{\omega \in \Omega} \pi(\omega)\sigma(a|\omega)$ denote the probability with which action a is recommended by σ in steady state. Let $\mathcal{X}_+^\sigma \triangleq \{(\omega, a) \in \mathcal{X} : \pi(\omega)\sigma(a|\omega) > 0\}$ denote the set of all state-action pairs that occur with positive probability under σ, and let $A_+^\sigma \triangleq \{a \in A : \tau_a^\sigma > 0\}$ denote the set of all actions that σ recommends with positive probability. Let μ_a^σ denote the posterior belief of a receiver after being recommended $a \in A_+^\sigma$. Finally, let $\mathcal{B}_{\mathsf{no}}^\sigma \triangleq \{\mu_a^\sigma : a \in A_+^\sigma\}$ denote the set of all posterior beliefs so induced, and let $\mathsf{Conv}(\mathcal{B}_{\mathsf{no}}^\sigma)$ denote its convex hull. (When it is clear from the context, we drop the superscript σ from the notation to avoid clutter.)

Theorem 4. *For a fixed $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$, the following statements are equivalent:*

1. *There exists a signaling mechanism $\widehat{\sigma} \in \mathsf{Pers}(\Phi_{\mathsf{full}}) \cap \Sigma_1$ that results in the same invariant distribution over the state-action pairs as under σ, and that induces the same posterior belief μ_a^σ after any action recommendation $a \in A_+^\sigma$;*

2. *There exists weights $\{\lambda(a|x) \geq 0 : a \in A_+, x \in \mathcal{X}_+\}$ satisfying*

$$\sum_{a \in A_+} \lambda(a|x)\mu_a(\omega) = p(\omega|x), \quad \text{for all } x \in \mathcal{X}_+ \text{ and } \omega \in \Omega,$$

$$\sum_{x = (\hat{\omega}, \hat{a}) \in \mathcal{X}_+} \lambda(a|x)\tau_{\hat{a}}\mu_{\hat{a}}(\hat{\omega}) = \tau_a, \quad \text{for all } a \in A_+.$$

In particular, the preceding statements hold if (i) the set of posterior beliefs $\mathcal{B}_{\mathsf{no}}^\sigma$ is linearly independent, and (ii) the transition kernels lie in the convex hull of these beliefs, i.e., $p(\cdot|x) \in \mathsf{Conv}(\mathcal{B}_{\mathsf{no}}^\sigma)$ for all $x \in \mathcal{X}_+$.

The theorem, together with Lemma 2, implies that when the conditions in the theorem statement hold, the sender can achieve the same optimal payoffs no matter the level of historical information possessed by the receivers. In particular, we have $\mathsf{OPT}(\Phi_\ell) = \mathsf{OPT}(\Phi_{\mathsf{no}})$ for all $\ell \geq 0$. The proof of the theorem uses the weights $\lambda(a|x)$ to explicitly construct a signaling mechanism $\widehat{\sigma} \in \Sigma_1$ which is persuasive for the model Φ_{full}, and induces the same set $\mathcal{B}_{\mathsf{no}}$ of posterior beliefs for the receivers with the same distribution, resulting in the same payoff for the sender. Finally, observe that the beliefs $\mu_a \in \mathcal{B}_{\mathsf{no}}$ can be easily computed from the optimal solution of the linear program $\mathsf{LP}(\mathsf{no})$; thus, the sufficient conditions in the theorem statement are straightforward to verify.

4.5 Optimal Persuasion in Partial-history Information Models via Robustness

We now turn to the study of optimal persuasion in the general partial-history information model. We first discuss the technical intricacies in finding an optimal signaling mechanism for $\mathsf{MPP}(\Phi_\ell)$, and the associated computational challenges for the problem $\mathsf{MPP}(\Phi_\ell, \Sigma_k)$, which we formulate as a bilinear optimization program. Given these challenges, we design an approximately optimal signaling mechanism for $\mathsf{MPP}(\Phi_\ell)$ for large enough ℓ, that is "simple" in the sense that it is history-independent and computationally efficient. Our key idea is to leverage the fast mixing property of underlying Markov chains, whereby after sufficiently many transitions, the state distribution will be close to, though not exactly the same as, the invariant distribution. To guarantee persuasiveness for this distribution, it suffices for our design to simply guarantee robust persuasiveness for every belief that is close to the invariant distribution. We show that such robust persuasiveness can be employed to yield a simple and approximately optimal persuasion signaling mechanism for reasonably large ℓ.

4.5.1 Intricacies of Persuasion in Partial-history Information Models

Consider the sender's persuasion problem $\mathsf{MPP}(\Phi_\ell)$ in the partial-history information model Φ_ℓ for general $\ell \geq 1$. In these models, the receivers neither have complete information about the history, nor do they completely lack history information. As we show next, this intermediate level of historical information makes the sender's persuasion problem challenging and technically intricate. In fact, even determining the degree of history dependence of the optimal signaling mechanism is difficult. This intricacy presents itself even in the simplest partial-history information model, namely Φ_1, as we explain next.

Recall that in the model Φ_1, the receivers observe the history with one-period lag, i.e., at time t, the corresponding receiver observes the history $\bar{h}_{t-1}$ at time $t-1$. Thus, this receiver knows the realization of $\bar{x}_{t-2} = (\bar{\omega}_{t-2}, \bar{a}_{t-2})$, but does not know $\bar{x}_{t-1} = (\bar{\omega}_{t-1}, \bar{a}_{t-1})$ and $\bar{\omega}_t$. An initial guess then is to consider signaling mechanisms in the set Σ_2, i.e., ones that make recommendations based on $(\bar{x}_{t-2}, \bar{x}_{t-1}, \bar{\omega}_t)$. (This comports well with the full-history information model $\Phi_{\mathsf{full}} = \Phi_0$, where the optimal signaling mechanism lies in the set Σ_1.) With such a choice of σ, after seeing $\bar{h}_{t-2} = h_{-2} = (\dots, x_{-3}, x_{-2})$, the receiver's belief about $(\bar{x}_{t-1}, \bar{\omega}_t)$ is given by

$$\mathbf{P}^{\sigma}\left(\bar{x}_{t-1} = x_{-1}, \bar{\omega}_t = \omega \;\middle|\; \bar{h}_{t-2} = h_{-2}\right) = p(\omega_{-1}|x_{-2})\sigma(a_{-1}|x_{-3}, x_{-2}, \omega_{-1})p(\omega|x_{-1}).$$

Thus, the receiver's belief depends not just on x_{-2}, but also on the realization x_{-3} of $\bar{x}_{t-3}$. Requiring the signaling mechanism σ to be persuasive for different beliefs of the receiver corresponding to different realization of x_{-3}, without depending on x_{-3} explicitly, is unlikely to yield optimality. Consequently, one is tempted to consider signaling mechanisms $\sigma \in \Sigma_3$, i.e., ones that make recommendation based on $(\bar{x}_{t-3}, \bar{x}_{t-2}, \bar{x}_{t-1}, \bar{\omega}_t)$. However, a similar argument as above would imply that for such signaling mechanisms, the receiver's belief would depend on the realization of $(\bar{x}_{t-4}, \dots \bar{x}_{t-2})$. In general, for any signaling mechanism $\sigma \in \Sigma_k$ with $k \geq 1$, the receiver at time t has a different belief for different realizations of $(\bar{x}_{t-(k+1)}, \dots, \bar{x}_{t-2})$, but the signaling mechanism σ does not base its recommendation on the realization of $\bar{x}_{t-(k+1)}$. Due to this mismatch of dependencies, it is unclear what the right dependence of the optimal signaling mechanism is on the history, or for that matter, even whether there exists an optimal signaling mechanism within the class Σ of signaling mechanisms.

Given the ambiguity regarding the degree of history dependence, one may instead consider optimizing the sender's payoff within a restricted subset Σ_k of signaling mechanisms, for some fixed k. However, even this restricted problem turns out to be computationally challenging since, unlike the case for Φ_{full} and Φ_{no}, it does not reduce to a linear program. In particular, the following result formulates the sender's problem $\mathsf{MPP}(\Phi_1, \Sigma_1)$ as a bilinear program.

Proposition 5. The sender's problem $\mathsf{MPP}(\Phi_1, \Sigma_k)$ can be formulated as the following bilinear program:

$$
\begin{aligned}
&\max_{z \geq 0} \sum_{h \in \mathcal{X}^{k+1}} \sum_{(\omega,a)\in\mathcal{X}} z(h,\omega,a)v(\omega,a) \\
&\quad \sum_{x\in\mathcal{X}} \sum_{\omega\in\Omega} z(h,x,\omega,a)\partial u(\omega,a,a') \geq 0, \qquad \forall h \in \mathcal{X}^k,\ a,a' \in A \\
&\quad \sum_{x'\in\mathcal{X}} z(x',h,x)p(\omega|x) - \sum_{a\in A} z(h,x,\omega,a) = 0, \quad \forall h \in \mathcal{X}^k,\ x\in\mathcal{X},\ \omega\in\Omega \\
&\quad \sum_{a'\in A} \big(z(x,h,\omega,a)\cdot z(x',h,\omega,a') - z(x',h,\omega,a)\cdot z(x,h,\omega,a')\big) = 0, \\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall h \in \mathcal{X}^k,\ x',x,(\omega,a)\in\mathcal{X} \\
&\quad \sum_{h \in \mathcal{X}^{k+1}} \sum_{(\omega,a)\in\mathcal{X}} z(h,\omega,a) = 1.
\end{aligned}
\tag{4.4}
$$

To elaborate, for any $\sigma \in \Sigma_k$ and under the information model Φ_1, given our preceding discussion, the relevant state from the perspective of a receiver at time t is $(\bar{h}_t^{k+1}, \bar{\omega}_t)$, of which the receiver has observed $\bar{h}_{t-1}^k$. Letting π denote the steady state distribution of $(\bar{h}_t^{k+1}, \bar{\omega}_t)$ under σ, and after defining $z(x,h,\omega,a) :- \pi(x,h,\omega)\sigma(a|h,\omega)$ for $x \in \mathcal{X}$, $h \in \mathcal{X}^k$ and $(\omega,a) \in \mathcal{X}$, the first equality in the bilinear program follows from the balance equations for the steady-state. The final equality follows from the fact that z denotes the joint distribution of $(\bar{h}_t^{k+1}, \bar{\omega}_t, \bar{a}_t)$ in steady state. The inequality follows from the persuasiveness constraint on σ, analogous to the same constraint in $\mathsf{LP}(\mathsf{full})$ and $\mathsf{LP}(\mathsf{no})$. However, it is in the second equality that we have the bilinearity of the optimization problem, which arises due to the restriction that σ lies in the set Σ_k, and hence cannot depend on $\bar{x}_{t-k-1}$. In particular, the non-linear equality constraint in (4.4) explicitly encodes the requirement that $\sigma(a|h,\omega) = z(x,h,\omega,a)/\pi(x,h,\omega)$ does not depend on $x \in \mathcal{X}$.

A similar argument holds for any partial-history information model Φ_ℓ, where for any signaling mechanism $\sigma \in \Sigma_k$ for $k \geq 0$, the belief of a receiver at time t depends on the realization of $(\bar{x}_{t-\ell-k}, \ldots, \bar{x}_{t-\ell-1})$, but the signaling mechanism σ does not base its recommendation on $(\bar{x}_{t-\ell-k}, \ldots, \bar{x}_{t-k-1})$. Again, the sender's problem $\mathsf{MPP}(\Phi_\ell, \Sigma_k)$ can be shown to be a bilinear optimization problem, whose size is exponential in $\ell + k$. We skip the details for the sake of brevity.

k	0	1	2	3	4	5
$\mathsf{OPT}(\Phi_1, \Sigma_k)$	0.576	0.772	0.799	0.808	0.811	0.821

Figure 4.1: (Example 4.4.1 contd.) Sender's optimal payoff in $\mathsf{MPP}(\Phi_1, \Sigma_k)$ for different values of k. The optimal values are obtained by numerically solving bilinear optimization programs analogous to (4.4) for different values of k. Here, $\mathsf{OPT}(\Phi_{\mathsf{full}}) = 0.52$ and $\mathsf{OPT}(\Phi_{\mathsf{no}}) = 1$.

The preceding discussion hints at a trade-off faced by the sender in the model Φ_ℓ for some $\ell \geq 1$. On one hand, the sender can adopt the optimal signaling mechanism for the full-history information model Φ_{full}, which is simple in that it only uses the previous state-action pair (and the current state) to recommend an action, and is persuasive for the model Φ_ℓ, as shown in Proposition 2. However, this simplicity may come at the cost of substantially lower payoffs, especially if ℓ is large. On the other hand, the sender may choose a large k and solve a non-linear program akin to (4.4) to find the best signaling mechanism within the class k, which likely will yield higher payoffs, at the cost of substantial computational complexity. (See e.g., Fig 4.1.) In the following section, we provide an approach to overcome this trade-off, as long as one is satisfied with approximate optimality.

4.5.2 Approximately Optimal Persuasion via Robustness

In this section, we ask and answer the following questions: in partial-history information models, can "simple" signaling mechanisms guarantee persuasiveness without sacrificing the sender's payoff too much? And if so, can we find such a mechanism in a computationally efficient manner? To answer these questions positively, we take the robustness approach from Chapter 2. Our starting point is the observation that, for a signaling mechanism σ, if the underlying Markov chain mixes rapidly, the belief of the receiver who has stale historical information must be close to the invariant distribution $\pi = \mathsf{Inv}(\sigma)$. Thus, if σ is simultaneously persuasive for all distributions close to π, it must be persuasive under the information models Φ_ℓ for all large enough ℓ. Using this insight, we explicitly construct a *robustly persuasive* history-independent signaling mechanism with good payoff guarantees.

To begin, recall that for any history-independent signaling mechanism $\sigma \in \Sigma_0$, assuming the receivers follow the recommendation, $\bar{x}_t = (\bar{\omega}_t, \bar{a}_t) \in \mathcal{X}$ forms a Markov chain. Let $\pi = \mathsf{Inv}(\sigma)$; we abuse the notation slightly by letting π also denote the marginal over $\bar{\omega}_t$, i.e., $\pi(\omega) = \sum_{a \in A} \pi(\omega, a)$ for $\omega \in \Omega$. For $\epsilon \geq 0$, let $\mathsf{B}_1(\pi, \epsilon)$ denote the set of all distributions $\mu \in \Delta(\Omega)$ that are ϵ-close to π in ℓ_1-norm: $\mathsf{B}_1(\pi, \epsilon) \triangleq \{\mu \in \Delta(\Omega) : \|\mu - \pi\|_1 \leq \epsilon\}$.

An ϵ-*robustly persuasive* signaling mechanism $\sigma \in \Sigma_0$ is one whose recommendations would be optimally adopted by any receiver whose prior belief about $\bar{\omega}_t$ lies in the set $\mathsf{B}_1(\pi, \epsilon)$:

$$\sum_{\omega} \mu(\omega)\sigma(a|\omega)\partial u(\omega, a, a') \geq 0, \text{ for all } a, a' \in A \text{ and all } \mu \in \mathsf{B}_1(\pi, \epsilon),$$

where $\pi = \mathsf{Inv}(\sigma)$. We denote the set of ϵ-robustly persuasive signaling mechanisms by $\mathsf{RP}(\epsilon)$. The value of ϵ captures the degree of robustness of a mechanism $\sigma \in \mathsf{RP}(\epsilon)$, with smaller values corresponding to lower robustness. Observe that for all $\epsilon \geq 0$, we have $\mathsf{RP}(\epsilon) \subseteq \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$, with equality for $\epsilon = 0$. Furthermore, the set $\mathsf{RP}(\epsilon)$ is non-empty for all $\epsilon \geq 0$, as it contains the signaling mechanism that recommends a receiver-optimal action at each state.

Our next result describes the relation between $\mathsf{RP}(\epsilon)$ and the set $\mathsf{Pers}(\Phi_\ell)$ for large ℓ. For $\sigma \in \Sigma_0$ and $\ell \geq 1$, let $Q_\sigma^\ell(x, \omega) \triangleq \mathbf{P}^\sigma(\bar{\omega}_\ell = \omega | \bar{x}_{-1} = x)$ denote the distribution of $\bar{\omega}_\ell$ under σ, given $\bar{x}_{-1} = x \in \mathcal{X}$. Define $d_\ell(\sigma)$ as the maximum ℓ_1-distance between $Q_\sigma^\ell(x)$ and $\pi = \mathsf{Inv}(\sigma)$ over $x \in \mathcal{X}$:

$$d_\ell(\sigma) \triangleq \sup_{x \in \mathcal{X}} \left\| Q_\sigma^\ell(x) - \pi \right\|_1 = \sup_{x \in \mathcal{X}} \sum_{\omega} \left| Q_\sigma^\ell(x, \omega) - \pi(\omega) \right|.$$

Finally, let $\gamma_\star(\sigma)$ denote the absolute spectral gap (Levin and Peres, 2017) of the Markov chain $\{\omega_t\}$ under σ and $\pi_{\min}(\sigma) = \min_\omega \pi(\omega) > 0$. We have the following result.

Lemma 3. Suppose the signaling mechanism $\sigma \in \Sigma_0$ is ϵ-robustly persuasive for $\epsilon > 0$. If $\ell \geq 0$ satisfies $d_\ell(\sigma) \leq \epsilon$, then $\sigma \in \mathsf{Pers}(\Phi_\ell)$. In particular, $\sigma \in \mathsf{Pers}(\Phi_\ell)$ for all $\ell \geq \frac{1}{\gamma_\star(\sigma)} \log\left(\frac{2}{\epsilon \pi_{\min}(\sigma)}\right)$.

The proof of the bound in the lemma statement uses the unichain assumption (Assumption 2) to bound the mixing time of the underlying Markov chain. The result implies that in order to find a signaling mechanism in $\mathsf{Pers}(\Phi_\ell)$, it suffices to find a history-independent signaling mechanism in the set $\mathsf{RP}(\epsilon)$ for small enough ϵ. We highlight that the required value of ϵ decays exponentially in ℓ, and hence the robustness requirements are not too stringent.

Given this preceding result, we seek to identify a robustly persuasive mechanism with good guarantees on the sender's payoff. We prove such a result next. To state the result, we need a definition. Define the sets $\mathcal{P}_a \subseteq \Delta(\Omega)$ as follows:

$$\mathcal{P}_a \triangleq \left\{ \mu \in \Delta(\Omega) : a \in \operatorname*{argmax}_{a'} \mathbf{E}_\mu[u(\omega, a')] \right\}.$$

In other words, $\mathcal{P}_a$ is the set of beliefs for which the receiver finds it optimal to choose action a. Similar to Chapter 3, we make the following regularity assumption on the receivers' utility function.

Assumption 3 (Regularity)**.** There exists a positive constant $D > 0$ and beliefs $\eta_a \in \mathcal{P}_a$ for $a \in A$ such that $\mathsf{B}_1(\eta_a, D) \subseteq \mathcal{P}_a$ for each $a \in A$, where $\mathsf{B}_1(\eta, D)$ is a radius-D ℓ_1-ball centered at η.

The regularity assumption ensures that each action for the receiver is optimal for a set of beliefs with non-zero (Lebesgue) measure. This ensures the exclusion of pathological instances, where there is an action that is optimal for the receiver under a unique belief. Furthermore, Chapter 2 establishes that the regularity assumption ensures that, in static problems, the cost of requiring robustness scales linearly in the degree of robustness.

Next, let $a_\omega \in A$ be a best response for a receiver at state $\omega \in \Omega$, i.e., $a_\omega \in \arg\max_{a \in A} u(\omega, a)$ for each $\omega \in A$. Let $P_f(\omega, \omega') \triangleq p(\omega' | \omega, a_\omega)$ denote the transition probability from state ω to state ω' on choosing the action a_ω, and let P_f denote the transition matrix of the underlying process. Note that the unichain assumption implies that P_f is ergodic. Let $\nu_f \in \Delta(\Omega)$ denote the steady state distribution under the transition kernel P_f. Furthermore, let $\tau \triangleq \max_\omega 1/\nu_f(\omega)$ denote the maximum expected *first return time* across all states. Finally, let s_f be the smallest positive singular value of the matrix $I - P_f$.

With these definitions in place, we are now ready to present the main result of this section.

Theorem 5. *For* $\epsilon < \frac{s_f w_{\min} D}{2(s_f + 2(1+\tau)\sqrt{|\Omega|})}$, *there exists a signaling mechanism* $\widehat{\sigma} \in \mathsf{RP}(\epsilon)$ *with the sender's payoff bounded below by*

$$\left(1 - \frac{2\epsilon}{w_{\min} D}\left(1 + \frac{2(1+\tau)\sqrt{|\Omega|}}{s_f}\right)\right) \cdot \mathsf{OPT}(\Phi_{\mathsf{no}}),$$

where $w_{\min}$ *is the smallest positive probability of recommending an action under the optimal mechanism under* Φ_{no}.

The preceding result, together with Lemma 3, implies that for the partial-history model Φ_ℓ with large enough ℓ, the sender need not solve a non-linear program. Instead, the sender can use a simple history-independent signaling mechanism to obtain approximately optimal payoffs. The proof involves an explicit construction of such a signaling mechanism $\widehat{\sigma} \in \mathsf{RP}(\epsilon)$. From a computational perspective, constructing such a mechanism requires solving $\mathsf{MPP}(\Phi_{\mathsf{no}})$ (equivalently the linear program $\mathsf{LP}(\mathsf{no})$), and solving a separate linear program (C.8) with $\mathcal{O}(|\Omega|)$ variables and constraints (see Lemma 11 in Appendix C.2 for details). Thus, not only the proposed mechanism obtains approximately optimal payoffs, but it also can be computed efficiently.

To construct the mechanism $\widehat{\sigma} \in \mathsf{RP}(\epsilon)$ with good guarantees on the sender's payoff, we use a similar approach as in Chapter 3, where we first identify a set of beliefs that we seek to induce as the receivers' posterior beliefs under the constructed mechanism. These beliefs are chosen to lie strictly in the interior of the sets $\mathcal{P}_a$, to ensure that the actions remain optimal for all close-by beliefs. However, unlike the static setting in Chapter 2, the endogeneity of the receivers' prior belief in our setting raises the question of whether there exists a mechanism that induces these beliefs as posteriors. To exhibit such a mechanism, we prove the following analog of the splitting lemma (Aumann et al., 1995b; Kamenica and Gentzkow, 2011) for the Markovian persuasion setting, providing conditions on a set of beliefs under which a signaling mechanism exists that induces those beliefs as posteriors.

Lemma 4. For a finite set S, let $\{\mu_s : s \in S\}$ be a set of beliefs, and for each $s \in S$, let $a_s \in A$ be such that $\mu_s \in \mathcal{P}_{a_s}$. Suppose there exists a set of weights $\{w_s \geq 0 : s \in S\}$ satisfying $\sum_{s\in S} w_s = 1$ and $\sum_{s\in S}\sum_{\omega\in\Omega} w_s \mu_s(\omega) p(\cdot|\omega, a_s) = \sum_{s\in S} w_s \mu_s$. Then, there exists a history-independent signaling mechanism σ, sending signals in the set

S, such that, under the no-history information model, the signaling mechanism σ is persuasive (where a signal s is interpreted as an action recommendation of a_s), with a receiver's posterior belief upon receiving signal s given by μ_s.

Conversely, for any history-independent signaling mechanism $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$, there exists weights $\{w_a \geq 0 : a \in A\}$ and beliefs $\{\mu_a \in \mathcal{P}_a : a \in A\}$ satisfying $\sum_{a \in A} w_a = 1$ and $\sum_{a \in A} \sum_{\omega \in \Omega} w_a \mu_a(\omega) p(\cdot|\omega, a) = \sum_{a \in A} w_a \mu_a$, such that, under the no-history information model, a receiver's posterior belief upon receiving an action recommendation $a \in A$ is given by μ_a.

With this Markov splitting lemma in hand, we construct our robustly persuasive mechanism by proving the existence of weights satisfying the preceding condition. We provide the complete proof in Appendix C.2.

4.6 Conclusion

We consider a Markovian persuasion setting between a single long-lived sender and a stream of receivers, where the sender commits to a signaling mechanism to maximize the long-run average reward. To capture settings where the receiver may have limited historical information, we analyze a set of endogenous information models. We observe that the sender's persuasion problem can be posed as simple linear programs under the full-history and the no-history information models. However, when the receiver has partial information about the history, the sender's problem presents technical intricacies, and is computationally challenging due to its non-linear nature. To overcome this difficulty, we adopt a robust persuasion approach to construct a simple history-independent signaling mechanism with strong guarantees on the payoff, that nevertheless is persuasive for all models with sufficiently limited historical information. Furthermore, the robust mechanism can be computed efficiently by solving simple linear programs. From a theoretical perspective, our work in this chapter quantifies the trade-off between higher sender's payoffs and being persuasive under a larger class of information models.

We have focused on the setting where the sender seeks to maximize the long-run average payoff. An alternative objective is to maximize the cumulative discounted reward. However, note that in endogenous information models, the receivers' belief is related to the invariant distribution of the process, which equals long-run averages in stationary models. Thus, the persuasion problem with discounted rewards is similar to a constrained Markov decision process where the objective involves discounting and the constraint requires averaging. Even in the classical context of constrained MDPs, problems with distinct discount factors in the objective and the constraints are challenging (note that averaging can be interpreted as the limit where the discount factor converges to one). For instance, Feinberg and Shwartz (1994, 1995) show that in such settings the optimal policy need not be stationary. An additional complexity that arises with discounting rewards is the dependence on the initial conditions. Given these challenges, a systematic analysis of Markov persuasion process with endogenous beliefs and discounted rewards is an interesting direction for further theoretical research.

Chapter 5

Conclusion and Discussion

This thesis leverages robustness to study two models of information design in uncertain environments. Starting with a static Bayesian persuasion problem involving a single sender and receiver, we introduced the notion of robust persuasiveness and characterized the cost of robust persuasion. Building on this, we extended our study to dynamic settings of a single sender and a stream of receivers. Chapter 3 considers the problem where the sender must learn the unknown state distribution while sending robustly persuasive recommendations. Using the robustness approach, we propose the $\mathfrak{Rai}$ algorithm which persuades robustly and achieves $\mathcal{O}(\sqrt{T \log T})$ regret against the optimal signaling mechanism under the knowledge of the state distribution. To account for state evolution, Chapter 4 studied the Markov persuasion problem, where the receiver's action also determines the system's next state (stochastically). Here, we developed a robust and simple signaling mechanism with strong guarantees on the payoff when the receivers have sufficiently limited historical information. Together, these contributions demonstrate how robustness provides a powerful tool for studying persuasion problems in uncertain environments.

Our results assume that the underlying state distribution remains fixed over time. However, in real-world settings, such distributions as product quality or demand/supply distribution often evolve dynamically. An important extension would explore how the sender can adapt the recommendations to such changing environments while maintaining robustness. This extension may require incorporating techniques from online learning, especially robust algorithms for non-stationary environments. For example, methods like regret minimization (Gajane et al., 2018; Cheung et al., 1906) could enable the sender to make recommendations that are persuasive and resilient to shifts in the underlying state distribution.

Another promising research direction is to extend our model to settings with multiple simultaneous senders. In Chapters 3 and 4, we focused on the single sender setting. However, many real-world platforms naturally involve multiple senders persuading the same receivers. In such settings, senders may have conflicting objectives or even engage in direct competition, introducing strategic interactions that fundamentally differ from those in the single-sender case. For example, a seller listed on both Amazon and eBay, with each platform aiming to convince the buyer of the product's quality. However, their goals may not be aligned: Amazon aims to increase sales volume through customer reviews and product labels like "Amazon's Choice", while eBay may focus on attracting more bids and increasing competition between buyers. The user, having access to information from both platforms, may be influenced differently depending on the recommendations from each platform on the same product. Similarly, Uber and Lyft both have real-time demand information and strategically use it to persuade drivers to reposition to specific pickup areas. While both platforms aim to optimize their own operations, each platform is focused on persuading drivers to pick up locations that benefit their own goals, whether it's maximizing efficiency or revenue, leading to competing recommendations for the same set of drivers. This raises a central question about how each sender should design their signaling mechanisms in the presence of others. Should senders coordinate their signals, exploit differences in their objectives, or adapt dynamically to the strategies of their rivals? Moreover, incorporating uncertainty about the environment adds further complexity. How can each sender share persuasive recommendations while accounting for state uncertainty and the potential actions of completing senders?

Finally, future work is to investigate robust persuasion with partial receiver feedback, where the sender has limited or noisy information about the receiver's actions or outcomes following a recommendation. For instance, platforms may only receive aggregate-level data or delayed feedback from the receivers, making it challenging to accurately compute the sender and receiver's utilities. This partial feedback environment introduces complexities in designing persuasive signaling mechanisms. Future work could focus on designing robustly persuasive signaling mechanisms that can account for the uncertainty in feedback while guaranteeing good performance.

References

S. Agrawal and N. R. Devanur. Bandits with concave rewards and convex knapsacks. In *Proceeding of the fifteenth ACM conference on Economics and computation*, pages 989–1006, 2014.

S. Alizamir, F. de Véricourt, and S. Wang. Warning against recurring risks: An information design approach. *Management Science*, 66(10):4612–4629, 2020.

S. Amani, M. Alizadeh, and C. Thrampoulidis. Linear stochastic bandits under safety constraints. *Advances in Neural Information Processing Systems*, 32, 2019.

A. Antelmi, D. Malandrino, and V. Scarano. Characterizing the behavioral evolution of twitter users and the truth behind the 90-9-1 rule. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 1035–1038, 2019.

J. Anunrojwong, K. Iyer, and D. Lingenbrink. Persuading risk-conscious agents: A geometric approach. *Available at SSRN 3386273*, 2020.

J. Anunrojwong, K. Iyer, and V. Manshadi. Information design for congested social services: Optimal need-based persuasion. *Management Science*, 2022.

G. Ashkenazi-Golan, P. Hernández, Z. Neeman, and E. Solan. Markovian persuasion with two states. *arXiv preprint arXiv:2209.06536*, 2022.

R. J. Aumann, M. Maschler, and R. E. Stearns. *Repeated games with incomplete information*. MIT press, 1995a.

R. J. Aumann, M. Maschler, and R. E. Stearns. *Repeated games with incomplete information*. MIT press, 1995b.

Y. Babichenko, I. Talgam-Cohen, H. Xu, and K. Zabarnyi. Regret-minimizing bayesian persuasion. *Games and Economic Behavior*, 136:226–248, 2022.

F. Bacchiocchi, F. E. Stradi, M. Castiglioni, A. Marchesi, and N. Gatti. Markov persuasion processes: Learning to persuade from scratch. *arXiv preprint arXiv:2402.03077*, 2024.

M.-F. Balcan, A. Blum, N. Haghtalab, and A. D. Procaccia. Commitment without regrets: Online learning in stackelberg security games. In *Proceedings of the sixteenth ACM conference on economics and computation*, pages 61–78, 2015.

D. Bergemann and S. Morris. Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics*, 11(2):487–522, 2016.

D. Bergemann and S. Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

M. Bernasconi, M. Castiglioni, A. Marchesi, N. Gatti, F. Trovò, et al. Sequential information design: Learning to persuade in the dark. In *Thirty-sixth Conference on Neural Information Processing Systems*, pages 1–25, 2022.

L. Besson and E. Kaufmann. What doubling tricks can and can't do for multi-armed bandits. *arXiv preprint arXiv:1803.06971*, 2018.

J. Bizzotto, J. Rüdiger, and A. Vigier. Dynamic persuasion with outside information. *American Economic Journal: Microeconomics*, 13(1):179–94, 2021.

S. Board and J. Lu. Competitive information disclosure in search markets. *Journal of Political Economy*, 126(5):1965–2010, 2018.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

M. Cai and M. Balestra. Demand and ETR Forecasting at Airports. `https://www.uber.com/blog/demand-and-etr-forecasting-at-airports`, 2023.

M. K. Camara, J. D. Hartline, and A. Johnsen. Mechanisms for a no-regret agent: Beyond the common prior. In *2020 IEEE 61st Annual Symposium on Foundations* *of Computer Science (FOCS)*, volume 1, pages 259–270, Los Alamitos, CA, USA, Nov 2020. IEEE Computer Society. doi: 10.1109/FOCS46700.2020.00033.

O. Candogan. Information design in operations. In *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*, pages 176–201. INFORMS, 2020.

X. Cao and K. R. Liu. Online convex optimization with time-varying constraints and bandit feedback. *IEEE Transactions on automatic control*, 64(7):2665–2680, 2018.

X. Cao, J. Zhang, and H. V. Poor. On the time-varying distributions of online stochastic optimization. In *2019 American Control Conference (ACC)*, pages 1494–1500. IEEE, 2019.

M. Castiglioni, A. Celli, A. Marchesi, and N. Gatti. Online bayesian persuasion. *Advances in Neural Information Processing Systems*, 33, 2020.

S. Chawla, J. D. Hartline, D. Malec, and B. Sivan. Prior-independent mechanisms for scheduling. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 51–60, 2013.

Y. Chen, Y. Liu, and C. Podimata. Learning strategy-aware linear classifiers. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 15265–15276. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper/2020/file/ae87a54e183c075c494c4d397d126a66-Paper.pdf`.

W. C. Cheung, D. Simchi-Levi, and R. Zhu. Drifting reinforcement learning: The blessing of (more) optimism in face of endogenous & exogenous dynamics. *Arxiv*, 1906.

P. Dhangwatnotai, T. Roughgarden, and Q. Yan. Revenue maximization with a single sample. *Games and Economic Behavior*, 91:318–333, 2015.

J. Dong, A. Roth, Z. Schutzman, B. Waggoner, and Z. S. Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70, 2018.

S. Dughmi. Algorithmic information structure design: a survey. *ACM SIGecom Exchanges*, 15(2):2–24, 2017.

S. Dughmi and H. Xu. Algorithmic persuasion with no externalities. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 351–368, 2017.

S. Dughmi and H. Xu. Algorithmic bayesian persuasion. *SIAM Journal on Computing*, 50(3):STOC16–68–STOC16–97, 2021.

P. Dworczak and A. Pavan. Preparing for the worst but hoping for the best: Robust (Bayesian) persuasion. *CEPR Discussion Paper No. DP15017*, 2020.

P. Dworczak and A. Pavan. Preparing for the worst but hoping for the best: Robust (bayesian) persuasion. *Econometrica*, 90(5):2017–2051, 2022.

M. Elliott, A. Galeotti, A. Koh, and W. Li. Matching and information design in marketplaces. *Available at SSRN*, 2022.

J. C. Ely. Beeps. *American Economic Review*, 107(1):31–53, 2017.

F. Farhadi and D. Teneketzis. Dynamic information design: A simple problem on optimal sequential information disclosure. *Dynamic Games and Applications*, 12 (2):443–484, 2022.

E. A. Feinberg and A. Shwartz. Markov decision models with weighted discounted criteria. *Mathematics of Operations Research*, 19(1):152–168, 1994.

E. A. Feinberg and A. Shwartz. Constrained markov decision models with weighted discounted rewards. *Mathematics of Operations Research*, 20(2):302–320, 1995.

G. Feng, G. Kong, and Z. Wang. We are on the way: Analysis of on-demand ride-hailing systems. *Manufacturing & Service Operations Management*, 23(5): 1237–1256, 2021.

S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer, 2013. doi: 10.1007/978-0-8176-4948-7.

P. Gajane, R. Ortner, and P. Auer. A sliding-window algorithm for markov decision processes with arbitrarily changing rewards and transitions. *arXiv preprint arXiv:1805.10066*, 2018.

J. Gan, R. Majumdar, G. Radanovic, and A. Singla. Bayesian persuasion in sequential decision-making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36(5), pages 5025–5033, 2022.

H. Guo, X. Liu, H. Wei, and L. Ying. Online convex optimization with hard constraints: Towards the best of two worlds and beyond. In *Advances in Neural Information Processing Systems*, 2022.

Y. Gur, G. Macnamara, I. Morgenstern, and D. Saban. Information disclosure and promotion policy design for platforms. *Management Science*, 2023.

N. Hahn, M. Hoefer, and R. Smorodinsky. The secretary recommendation problem. *arXiv preprint arXiv:1907.04252*, 2019.

N. Hahn, M. Hoefer, and R. Smorodinsky. Prophet inequalities for bayesian persuasion. In *Proc. 29th Int. Joint Conf. Artif. Intell.(IJCAI)*, pages 175–181, 2020.

J. Hu and X. Weng. Robust persuasion of a privately informed receiver. *Economic Theory*, pages 1–45, 2020.

J. Hu and X. Weng. Robust persuasion of a privately informed receiver. *Economic Theory*, 72(3):909–953, 2021.

K. Iyer, H. Xu, and Y. Zu. Markov persuasion processes with endogenous agent beliefs. *arXiv preprint arXiv:2307.03181*, 2023.

E. Kamenica and M. Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

K. Khezeli and E. Bitar. Safe linear stochastic bandits. *ArXiv*, abs/1911.09501, 2020.

Y. Kim and D. Lee. Online convex optimization with stochastic constraints: Zero constraint violation and bandit feedback. *arXiv preprint arXiv:2301.11267*, 2023.

R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, page 594, USA, 2003. IEEE Computer Society. ISBN 0769520405.

S. Kosterina. Persuasion with unknown beliefs. *Work. Pap., Princeton Univ., Princeton, NJ*, 2018.

S. Kosterina. Persuasion with unknown beliefs. *Theoretical Economics*, 17(3):1075–1107, 2022.

I. Kremer, Y. Mansour, and M. Perry. Implementing the "wisdom of the crowd". *Journal of Political Economy*, 122(5):988–1012, 2014.

T. Lattimore and C. Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020. ISBN 9781108486828. URL `https://books.google.com/books?id=bbjpDwAAQBAJ`.

E. Lehrer and D. Shaiderman. Markovian persuasion with stochastic revelations. *arXiv preprint arXiv:2204.08659*, 2022.

D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

F. Li and P. Norman. Sequential persuasion. *Theoretical Economics*, 16(2):639–675, 2021.

N. Liakopoulos, A. Destounis, G. Paschos, T. Spyropoulos, and P. Mertikopoulos. Cautious regret minimization: Online optimization with long-term budget constraints. In *International Conference on Machine Learning*, pages 3944–3952. PMLR, 2019.

D. Lingenbrink and K. Iyer. Optimal signaling mechanisms in unobservable queues. *Operations research*, 67(5):1397–1416, 2019.

M. Mahdavi, R. Jin, and T. Yang. Trading regret for efficiency: Online convex optimization with long term constraints. *CoRR*, abs/1111.6082, 2011. URL `http://arxiv.org/abs/1111.6082`.

M. Mahdavi, T. Yang, and R. Jin. Stochastic convex optimization with multiple objectives. *Advances in neural information processing systems*, 26, 2013.

Y. Mansour, A. Slivkins, and V. Syrgkanis. Bayesian incentive-compatible bandit exploration. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 565–582, 2015.

Y. Mansour, A. Slivkins, V. Syrgkanis, and Z. S. Wu. Bayesian exploration: Incentivizing exploration in bayesian games. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 661–661, 2016.

A. Moradipari, C. Thrampoulidis, and M. Alizadeh. Stage-wise conservative linear bandits. *Advances in neural information processing systems*, 33:11191–11201, 2020.

A. Moradipari, S. Amani, M. Alizadeh, and C. Thrampoulidis. Safe linear thompson sampling with side information. *IEEE Transactions on Signal Processing*, 69: 3755–3767, 2021.

R. B. Myerson. Incentive compatibility and the bargaining problem. *Econometrica: journal of the Econometric Society*, pages 61–73, 1979.

M. J. Neely and H. Yu. Online convex optimization with time-varying constraints. *arXiv preprint arXiv:1702.04783*, 2017.

D. Orlov, A. Skrzypacz, and P. Zryumov. Persuading the principal to wait. *Journal of Political Economy*, 128(7):2542–2578, 2020.

A. Pacchiano, M. Ghavamzadeh, P. Bartlett, and H. Jiang. Stochastic bandits with linear constraints. In *International Conference on Artificial Intelligence and Statistics*, pages 2827–2835. PMLR, 2021.

M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

J. Renault, E. Solan, and N. Vieille. Optimal dynamic information provision. *Games and Economic Behavior*, 104:329–349, 2017.

G. Romanyuk and A. Smolin. Cream skimming and information design in matching markets. *American Economic Journal: Microeconomics*, 11(2):250–76, 2019.

J. N. Tsitsiklis. Np-hardness of checking the unichain condition in average cost mdps. *Operations research letters*, 35(3):319–323, 2007.

Uber Technologies, Inc. Understanding acceptance and cancellation rates, 2024. URL `https://www.uber.com/blog/understanding-acceptance-and-cancellation-rates/?uclick_id=c27bca9c-a290-4ca9-9b22-8d4b9bd04bc1`. Accessed: 2024-11-02.

T. Ui. Optimal and robust disclosure of public information. *arXiv preprint arXiv:2203.16809*, 2022.

I. Usmanova, A. Krause, and M. Kamgarpour. Safe convex learning under uncertain constraints. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2106–2114. PMLR, 2019.

T. Van Mierlo et al. The 1% rule in four digital health social networks: an observational study. *Journal of medical Internet research*, 16(2):e2966, 2014.

R. W. Wolff. Poisson arrivals see time averages. *Operations Research*, 30(2):223–231, 1982. ISSN 0030364X, 15265463. URL `http://www.jstor.org/stable/170165`.

J. Wu, Z. Zhang, Z. Feng, Z. Wang, Z. Yang, M. I. Jordan, and H. Xu. Sequential information design: Markov persuasion process and its efficient reinforcement learning. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 471–472, 2022.

W. Wu. Sequential bayesian persuasion. *Journal of Economic Theory*, 214:105763, 2023.

C. Yan, H. Zhu, N. Korolko, and D. Woodard. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)*, 67(8):705–724, 2020.

P. Yang, K. Iyer, and P. Frazier. Information design in spatial resource competition. *arXiv preprint arXiv:1909.12723*, 2019.

X. Yi, X. Li, T. Yang, L. Xie, T. Chai, and K. Johansson. Regret and cumulative constraint violation analysis for online convex optimization with long term constraints. In *International Conference on Machine Learning*, pages 11998–12008. PMLR, 2021.

X. Yi, X. Li, T. Yang, L. Xie, T. Chai, and H. Karl. Regret and cumulative constraint violation analysis for distributed online constrained convex optimization. *IEEE Transactions on Automatic Control*, 2022.

H. Yu, M. Neely, and X. Wei. Online convex optimization with stochastic constraints. *Advances in Neural Information Processing Systems*, 30, 2017.

J. Yuan and A. G. Lamperski. Online convex optimization for cumulative constraints. *CoRR*, abs/1802.06472, 2018. URL `http://arxiv.org/abs/1802.06472`.

Y. Zu, K. Iyer, and H. Xu. Learning to persuade on the fly: Robustness against ignorance. *Operations Research*, 2024.

Appendix A

Appendix to Chapter 2

A.1 Proofs from Section 2.4.1

This section provides the proof of the propositions in Section 2.4.1. Throughout, we use the same notation as in the main text.

A.1.1 Proof of Proposition 1

Proof of Proposition 1. Observe that for $\epsilon > \frac{p_0^2 D}{4}$, we have $\frac{4\epsilon}{p_0^2 D} > 1$, and hence the specified bound is trivial. Hence, hereafter, we assume $\epsilon \leq \frac{p_0^2 D}{4}$.

To begin, let $\sigma \in \arg\max_{\sigma' \in \mathsf{Pers}(\mu)} V(\mu, \sigma')$ denote the optimal signaling mechanism under the distribution μ. Let $A_+ = \{a \in A : \sum_{\omega \in \Omega} \sigma(\omega, a) > 0\}$ denote the set of all actions that are recommended with positive probability under σ. For each $a \in A_+$, let μ_a denote the receiver's posterior belief (under signaling mechanism σ) upon receiving the action recommendation a. Note that since σ is persuasive under μ, we must have $\mu_a \in \mathcal{P}_a$. By the splitting lemma (Aumann et al., 1995a), it then follows that μ can be written as a convex combination $\sum_{a \in A_+} w_a \mu_a$ of $\{\mu_a : a \in A_+\}$, where $w_a \in [0, 1]$ is given by $w_a = \sum_{\omega \in \Omega} \mu(\omega) \sigma(\omega, a)$.

We next explicitly construct a signaling mechanism $\widehat{\sigma}$. To simplify the proof argument, the signaling mechanism $\widehat{\sigma}$ we construct is not a *straightforward* mechanism, in the sense that it reveals more than just action recommendations for signals in S. Using revelation principle, one can construct an equivalent straightforward mechanism $\bar{\sigma}$ by *coalescing* (Anunrojwong et al., 2020) signals with the same best response for the signal. We omit the details of this reduction. We start with some definitions that are needed to construct the signaling mechanism $\widehat{\sigma}$.

Let $\eta_a \in \mathcal{P}_a$ be such that $\mathsf{B}_1(\eta_a, D) \subseteq \mathcal{P}_a$. For $\delta = \frac{2\epsilon}{p_0 D} \in [0,1]$, define $\xi_a = (1-\delta)\mu_a + \delta\eta_a \in \mathcal{P}_a$ for each $a \in A_+$ and let $\xi = \sum_{a\in A_+} w_a \xi_a$. Furthermore, since $\mu_a \in \mathcal{P}_a$ and $\mathsf{B}_1(\eta_a, D) \subseteq \mathcal{P}_a$, the convexity of the set $\mathcal{P}_a$ implies that $\mathsf{B}_1(\xi_a, \delta D) \subseteq \mathcal{P}_a$.

Since $\mu \in \mathcal{B}_0 \subseteq \mathsf{relint}(\Delta(\Omega))$, we have $\frac{1}{1-\rho}(\mu - \rho\xi) \in \Delta(\Omega)$ for all small enough $\rho > 0$. Let $\bar{\rho} \triangleq \sup\left\{\rho \in [0,1] : \frac{1}{1-\rho}(\mu - \rho\xi) \in \Delta(\Omega)\right\}$ be the largest such value in $[0,1]$, and define χ as

$$\chi \triangleq \begin{cases} \frac{1}{1-\bar{\rho}}\left(\mu - \bar{\rho}\xi\right), & \text{if } \bar{\rho} < 1; \\ \mu, & \text{if } \bar{\rho} = 1. \end{cases}$$

Then, we obtain $\mu = \bar{\rho}\xi + (1-\bar{\rho})\chi$. Furthermore, if $\bar{\rho} < 1$, we have

$$\bar{\rho} = \frac{\|\chi - \mu\|_1}{\|\chi - \mu\|_1 + \|\mu - \xi\|_1} \geq \frac{p_0}{p_0 + \delta},$$

where the inequality follows from $\|\mu - \xi\|_1 \leq \sum_{a\in A_+} w_a \|\mu_a - \xi_a\|_1 = \delta \sum_{a\in A_+} w_a \|\eta_a - \xi_a\|_1 \leq 2\delta$ and from the fact that χ lies in the boundary of $\Delta(\Omega)$, which implies $\|\chi - \mu\|_1 \geq 2\min_\omega \mu(\omega) \geq 2p_0$.

With the preceding definitions in place, we are now ready to construct the mechanism $\widehat{\sigma}$. Let a_ω be a best response for the receiver at state $\omega \in \Omega$, and let $S = \{(\omega, a_\omega) \in \Omega \times A : \chi(\omega) > 0\}$. Consider the signaling mechanism $\widehat{\sigma}$, with the set of signals $A_+ \cup S$, defined as follows: for each $\omega \in \Omega$, let

$$\widehat{\sigma}(\omega, s) \triangleq \begin{cases} \bar{\rho}\frac{w_a \xi_a(\omega)}{\mu(\omega)}, & \text{for } s = a \in A_+; \\ (1-\bar{\rho})\frac{\chi(\omega)}{\mu(\omega)}, & \text{for } s = (\omega, a_\omega) \in S; \\ 0, & \text{otherwise.} \end{cases} \tag{A.1}$$

We now show that the signaling mechanism $\widehat{\sigma}$ is persuasive for all distributions in $\mathsf{B}_1(\mu, \epsilon)$, in the sense that for all signals $s \in A_+$ it is optimal for the receiver to play s, and for all signals $s = (\omega, a_\omega) \in S$, it is optimal for the receiver to play a_ω. To see this, for any $\gamma \in \mathsf{B}_1(\mu, \epsilon)$, let $\gamma(\cdot|s)$ denote the receiver's posterior under signaling mechanism $\widehat{\sigma}$ upon receiving the signal $s \in A_+ \cup S$. For $s = (\omega, a_\omega) \in S$, we have $\gamma(\cdot|s) = e_\omega$, where e_ω is the belief that puts all its weight on $\omega \in \Omega$. Thus, upon receiving the signal $s = (\omega, a_\omega)$ it is optimal for the receiver with the distribution γ to take action a_ω. Thus, it only remains to show that signals $s = a \in A_+$ are persuasive.

For $a \in A_+$, we have for $\omega \in \Omega$,

$$\begin{aligned}
\mu(\omega|a) &= \frac{\mu(\omega)\widehat{\sigma}(\omega, a)}{\sum_{\omega' \in \Omega} \mu(\omega')\widehat{\sigma}(\omega', a)} = \xi_a(\omega) \\
\gamma(\omega|a) &= \frac{\gamma(\omega)\widehat{\sigma}(\omega, a)}{\sum_{\omega' \in \Omega} \gamma(\omega')\widehat{\sigma}(\omega', a)} = \frac{\gamma(\omega)}{\mu(\omega)} \cdot \frac{\xi_a(\omega)}{\sum_{\omega' \in \Omega} \frac{\gamma(\omega')\xi_a(\omega')}{\mu(\omega')}}.
\end{aligned}$$

Then, using triangle inequality and some algebra, we obtain

$$\begin{aligned}
\|\gamma(\cdot|a) - \mu(\cdot|a)\|_1 &= \sum_{\omega \in \Omega} |\gamma(\omega|a) - \xi_a(\omega)| \\
&\leq \sum_{\omega \in \Omega} \left| \gamma(\omega|a) - \frac{\gamma(\omega)}{\mu(\omega)} \cdot \xi_a(\omega) \right| + \sum_{\omega \in \Omega} \left| \frac{\gamma(\omega)}{\mu(\omega)} \cdot \xi_a(\omega) - \xi_a(\omega) \right| \\
&\leq 2 \cdot \sup_{\omega \in \Omega} \frac{\xi_a(\omega)}{\mu(\omega)} \cdot \|\gamma - \mu\|_1 \\
&\leq \frac{2\epsilon}{p_0},
\end{aligned}$$

where in the final inequality, we have used $\min_\omega \mu(\omega) \geq p_0$ to get $\sup_{\omega \in \Omega} \frac{\xi_a(\omega)}{\mu(\omega)} \leq \frac{1}{p_0}$. Since $\mu(\cdot|a) = \xi_a$, this implies that $\gamma(\cdot|a) \in \mathsf{B}_1\left(\xi_a, \frac{2\epsilon}{p_0}\right) = \mathsf{B}_1(\xi_a, \delta D) \subseteq \mathcal{P}_a$. Thus, the signal $a \in A_+$ is persuasive for the distribution $\gamma \in \mathsf{B}_1(\mu, \epsilon)$. Taken together, we obtain that the signaling mechanism $\widehat{\sigma}$ is persuasive for all $\gamma \in \mathsf{B}_1(\mu, \epsilon)$.

The persuasiveness of $\widehat{\sigma}$ for all $\gamma \in \mathsf{B}_1(\mu, \epsilon)$ implies that

$$
\begin{aligned}
\sup_{\sigma' \in \mathsf{Pers}(\mathsf{B}_1(\mu,\epsilon))} V(\mu, \sigma') &\geq V(\mu, \widehat{\sigma}) \\
&= \sum_{\omega \in \Omega} \sum_{a \in A_+} \mu(\omega) \widehat{\sigma}(\omega, a) v(\omega, a) + \sum_{\omega \in \Omega} \sum_{s \in S} \mu(\omega) \widehat{\sigma}(\omega, s) v(\omega, a_\omega) \\
&\geq \sum_{\omega \in \Omega} \sum_{a \in A_+} \bar{\rho} w_a \xi_a(\omega) v(\omega, a) \\
&= \bar{\rho} \sum_{\omega \in \Omega} \sum_{a \in A_+} w_a \left((1-\delta) \mu_a(\omega) + \delta \eta_a(\omega) \right) v(\omega, a) \\
&\geq \bar{\rho}(1-\delta) \sum_{\omega \in \Omega} \sum_{a \in A_+} w_a \mu_a(\omega) v(\omega, a) \\
&= \bar{\rho}(1-\delta) \mathsf{OPT}(\mu).
\end{aligned}
$$

Thus, we obtain

$$
\begin{aligned}
\mathsf{Gap}(\mu, \mathsf{B}_1(\mu, \epsilon)) &= \mathsf{OPT}(\mu) - \sup_{\sigma' \in \mathsf{Pers}(\mathsf{B}_1(\mu,\epsilon))} V(\mu, \sigma') \\
&\leq \left(1 - \bar{\rho}(1-\delta)\right) \mathsf{OPT}(\mu) \\
&\leq \left(\frac{4}{p_0^2 D} \right) \epsilon,
\end{aligned}
$$

where the final inequality follows from $\bar{\rho} \geq \frac{p_0}{p_0+\delta}$, $\delta = \frac{2\epsilon}{p_0 D}$ and $\mathsf{OPT}(\mu) \leq 1$. □

A.1.2 Proof of Proposition 2

In this section, we provide the proof of Proposition 2.

Proof of Proposition 2. It is straightforward to verify that the following signaling mechanism $\sigma^* \in \mathsf{Pers}(\mu^*)$ optimizes the sender's expected utility among all mechanisms in $\mathsf{Pers}(\mu^*)$:

$$\begin{aligned}
\sigma^*(\omega_0, a_1) = \sigma^*(\omega_0, a_2) &= \frac{1}{2}, \\
\sigma^*(\omega_1, a_1) = \sigma^*(\omega_2, a_2) &= \frac{1}{2} + \frac{D}{2(1-p_0)}, \\
\sigma^*(\omega_1, a_2) = \sigma^*(\omega_2, a_1) &= \frac{1}{2} - \frac{D}{2(1-p_0)}, \\
\sigma^*(\omega, a) = 0, &\quad \text{otherwise.}
\end{aligned}$$

Since the action recommendations are always in $\{a_1, a_2\}$, we obtain $\mathsf{OPT}(\mu^*) = 1$.

Recall that $\bar{\mu}_1 = \mu^* + \frac{\epsilon}{2}(e_1 - e_2), \bar{\mu}_2 = \mu^* + \frac{\epsilon}{2}(e_2 - e_1)$. By the linearity of obedience constraints and $\mu^* = (\bar{\mu}_1 + \bar{\mu}_2)/2$, it follows that $\mathsf{Pers}(\{\mu^*, \bar{\mu}_1, \bar{\mu}_2\})$ can be obtained by imposing the obedience constraints at distributions $\bar{\mu}_1$ and $\bar{\mu}_2$. The optimization problem $\max_\sigma\{V(\mu^*, \sigma) : \sigma \in \mathsf{Pers}(\{\bar{\mu}_1, \bar{\mu}_2\})\}$ can be solved to obtain the following optimal signaling mechanism:

$$\begin{aligned}
\hat{\sigma}(\omega_0, a_1) = \hat{\sigma}(\omega_0, a_2) &= \frac{1}{2}, \\
\hat{\sigma}(\omega_1, a_1) = \hat{\sigma}(\omega_2, a_2) &= \frac{X}{Z}, \\
\hat{\sigma}(\omega_1, a_2) = \hat{\sigma}(\omega_2, a_1) &= \frac{Y}{Z}, \\
\hat{\sigma}(\omega_1, a_3) = \hat{\sigma}(\omega_2, a_4) &= 1 - \hat{\sigma}(\omega_1, a_1) - \hat{\sigma}(\omega_1, a_2), \\
\hat{\sigma}(\omega, a) = 0, &\quad \text{otherwise,}
\end{aligned}$$

where

$$\begin{aligned}
X &= 2p_0(1-p_0-\epsilon)(1-p_0+D)D^2 + p_0(1-p_0+\epsilon)(1-p_0-D-2D^2), \\
Y &= p_0(1-p_0-\epsilon)(1-p_0-3D+2D^2) + 2p_0(1-p_0+\epsilon)(1-p_0-D)(1-2D)D^2, \\
Z &= (1-p_0+\epsilon)^2(1-p_0-D)(1-2D)(1-p_0-D-2D^2) \\
&\quad - (1-p_0-\epsilon)^2(1-p_0+D)(1-p_0-3D+2D^2).
\end{aligned}$$

The difference in the sender's expected utility between using the optimal persuasive signaling mechanism for the distribution $\mu^* \in \mathcal{B}$ and using the optimal signaling mechanism that is persuasive for all distributions in $\{\mu^*, \bar{\mu}_1, \bar{\mu}_2\}$ is given by

$$\begin{aligned}\mathsf{Gap}(\mu^*, \mathsf{Pers}(\mu^*, \bar{\mu}_1, \bar{\mu}_2)) &= V(\mu^*, \sigma^*) - V(\mu^*, \hat{\sigma}) \\ &\geq \frac{\epsilon}{2} \frac{1/2 + Dp_0(1 + \epsilon/2 - Dp_0 - D)}{Dp_0 + \epsilon} \\ &\geq \frac{\epsilon}{8Dp_0}.\end{aligned}$$

□

Appendix B

Appendix to Chapter 3

B.1 Proofs from Section 3.4

B.1.1 Proof of Theorem 1

Proof of Theorem 1. If $\mu^* \in \mathcal{B}_t$ for each $t \in [T]$, then since $\sigma[h_t]$ is persuasive under all distributions in $\mathcal{B}_t$, we deduce that $\sigma[h_t]$ is persuasive under the distribution μ^* for all $t \in [T]$. Thus, we obtain that the $\mathfrak{Rai}$-algorithm is β-robustly persuasive for

$$\beta = \sup_{\mu^* \in \mathcal{B}_0} \mathbf{P}_{\mu^*} \left(\cap_{t \in [T]} \mathcal{B}_t \not\ni \mu^* \right).$$

Now, for any $\mu \in \mathcal{B}_0$, using the union bound we get

$$\begin{aligned}
\mathbf{P}_{\mu} \left(\cap_{t \in [T]} \mathcal{B}_t \not\ni \mu \right) &= \mathbf{P}_{\mu} \left(\cup_{t \in [T]} \mathcal{B}_t^c \ni \mu \right) \\
&\leq \sum_{t \in [T]} \mathbf{P}_{\mu} \left(\mathcal{B}_t^c \ni \mu \right) \\
&= \sum_{t \in [T]} \mathbf{P}_{\mu} \left(\|\gamma_t - \mu\|_1 > \epsilon_t \right) \\
&= \sum_{t \in [T]} \mathbf{P}_{\mu} \left(\|\gamma_t - \mu\|_1 > \sqrt{\frac{|\Omega|}{t}} \left(1 + \sqrt{\Phi \log T} \right) \right).
\end{aligned}$$

For $t < \frac{1}{4}\Phi \log T$, we have

$$\sqrt{\frac{|\Omega|}{t}}\left(1+\sqrt{\Phi \log T}\right) > 2\sqrt{|\Omega|}\left(1+\frac{1}{\sqrt{\Phi \log T}}\right) \geq 2.$$

Hence, $\mathbf{P}_\mu\left(\|\gamma_t - \mu\|_1 > \sqrt{\frac{|\Omega|}{t}}\left(1+\sqrt{\Phi \log T}\right)\right) = 0$. On the other hand, for $t \geq \frac{1}{4}\Phi \log T$, we have $\sqrt{\Phi \log T} \leq 2\sqrt{t}$, and hence from Lemma 5, we obtain

$$\begin{aligned}
\sum_{t \geq \frac{1}{4}\Phi \log T} \mathbf{P}_\mu\left(\|\gamma_t - \mu\|_1 > \sqrt{\frac{|\Omega|}{t}}\left(1+\sqrt{\Phi \log T}\right)\right) &\leq \sum_{t \geq \frac{1}{4}\Phi \log T} \exp\left(-\frac{3\Phi \log T\sqrt{\Omega}}{56}\right) \\
&\leq T^{-\frac{3\Phi\sqrt{\Omega}}{56}}\left(T - \frac{\Phi \log T}{4}\right) \\
&\leq T^{1-\frac{3\Phi\sqrt{\Omega}}{56}}.
\end{aligned}$$

Setting $\Phi > 20$ implies that the final term is at most $T^{-0.5}$. □

Lemma 5. For each $t \in [T]$, and for any $\mu \in \Delta(\Omega)$, we have for all $0 < \Phi_t \leq 2\sqrt{t}$,

$$\mathbf{P}_\mu\left(\|\gamma_t - \mu\|_1 \geq \sqrt{\frac{|\Omega|}{t}}\left(1+\Phi_t\right)\right) \leq \exp\left(-\frac{3\Phi_t^2\sqrt{\Omega}}{56}\right)\mathbf{I}\left\{\sqrt{\frac{|\Omega|}{t}}\left(1+\Phi_t\right) \leq 2\right\}.$$

Proof. Let $X_t \in \{0,1\}^{|\Omega|}$ denote the random variable with $X_t(\omega) = \mathbf{I}\{\omega_t = \omega\}$, and define $Y_t = X_t - \mathbf{E}_\mu[X_t]$. Let $Z_t = \left\|\sum_{\tau \in [t]} Y_\tau\right\|_1$. Since $\|Y_t\|_1 \leq \|X_t - \mathbf{E}_\mu[X_t]\|_1 \leq 2$ for each $t \in [T]$, by Foucart and Rauhut (2013, Corollary 8.46), we obtain for each $t \in [T]$,

$$\mathbf{P}_\mu\left(Z_t \geq \mathbf{E}_\mu[Z_t] + s\right) \leq \exp\left(-\frac{3s^2}{4\left(6t + 6\mathbf{E}_\mu[Z_t] + s\right)}\right).$$

Next, letting $Z_{t,\omega} = |\sum_{\tau \in [t]} Y_\tau(\omega)|$ for $\omega \in \Omega$, we obtain

$$\begin{aligned}
\mathbf{E}_\mu[Z_t] &= \sum_{\omega \in \Omega} \mathbf{E}_\mu[Z_{t,\omega}] \\
&= \sum_{\omega \in \Omega} \mathbf{E}_\mu[\sqrt{Z_{t,\omega}^2}] \\
&\leq \sum_{\omega \in \Omega} \sqrt{\mathbf{E}_\mu[Z_{t,\omega}^2]} \\
&= \sum_{\omega \in \Omega} \sqrt{\sum_{\tau \in [t]} \mathbf{Var}_\mu[Y_\tau(\omega)]} \\
&= \sqrt{t} \cdot \sum_{\omega \in \Omega} \sqrt{\mu(\omega)(1 - \mu(\omega))} \\
&\leq \sqrt{|\Omega| t},
\end{aligned}$$

where the first inequality follows from Jensen's inequality, and the third equality follows from the fact that, since $\mathbf{E}_\mu[Y_t(\omega)] = 0$, we have $\mathbf{E}[Z_{t,\omega}^2] = \sum_{\tau \in [t]} \mathbf{Var}_\mu[Y_\tau(\omega)]$. The final step follows from a straightforward optimization. Thus, we obtain

$$\mathbf{P}_\mu\left(Z_t \geq \sqrt{|\Omega| t} + s\right) \leq \exp\left(-\frac{3s^2}{4\left(6t + 6\sqrt{|\Omega| t} + s\right)}\right).$$

Choosing $s = \Phi_t \sqrt{|\Omega| t}$ for $0 < \Phi_t \leq 2\sqrt{t}$, and noting that $Z_t = t\|\gamma_t - \mu\|_1$, we obtain

$$\begin{aligned}
\mathbf{P}_\mu\left(\|\gamma_t - \mu\|_1 \geq \sqrt{\frac{|\Omega|}{t}}\,(1 + \Phi_t)\right) &\leq \exp\left(-\frac{3\Phi_t^2 |\Omega| t}{4\left(6t + 6\sqrt{|\Omega| t} + \Phi_t \sqrt{|\Omega| t}\right)}\right) \\
&\leq \exp\left(-\frac{3\Phi_t^2 \sqrt{|\Omega|}}{4\left(12 + \Phi_t / \sqrt{t}\right)}\right) \\
&\leq \exp\left(-\frac{3\Phi_t^2 \sqrt{\Omega}}{56}\right).
\end{aligned}$$

The lemma statement then follows after noticing that for all $t \in [T]$, we have $\|\gamma_t - \mu\|_1 \leq \|\gamma_t\|_1 + \|\mu\|_1 \leq 2$. □

B.2 Proofs from Section 3.5

B.2.1 Proof of Theorem 2

In this section, we provide the proof of Theorem 2. In the process, we also state and prove several helper lemmas used in the proof.

Proof of Theorem 2. In Lemma 6, we obtain the following bound on the regret:

$$\begin{aligned}\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) \leq &\sum_{t\in[T]} \mathsf{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \|\mu^* - \gamma_t\|_1)) + \sum_{t\in[T]} \mathsf{Gap}(\gamma_t, \mathcal{B}_t) \\ &+ \sum_{t\in[T]} \|\mu^* - \gamma_t\|_1 + \sum_{t\in[T]} \left(\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t)\right).\end{aligned}$$

Now, from Proposition 1, we have

$$\mathsf{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \|\mu^* - \gamma_t\|_1)) \leq \left(\frac{4}{p_0^2 D}\right) \cdot \|\mu^* - \gamma_t\|_1.$$

Thus, we obtain

$$\begin{aligned}\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) \leq &\sum_{t\in[T]} \mathsf{Gap}(\gamma_t, \mathcal{B}_t) + \left(\frac{4}{p_0^2 D} + 1\right) \sum_{t\in[T]} \|\mu^* - \gamma_t\|_1 \\ &+ \sum_{t\in[T]} \mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t).\end{aligned}$$

Finally, in Lemma 9, we show that on the event $\{\mu^* \in \mathcal{B}_t\}$, we have $\mathsf{Gap}(\gamma_t, \mathcal{B}_t) \leq$

$\left(\frac{16}{p_0^2 D}\right)\epsilon_t$. Thus, on the event $\{\mu^* \in \cap_{t\in[T]}\mathcal{B}_t\}$, we obtain

$$\begin{aligned}
\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) &\le \left(\frac{20}{p_0^2 D} + 1\right)\sum_{t\in[T]}\epsilon_t + \sum_{t\in[T]}\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t) \\
&\le \left(\frac{20}{p_0^2 D} + 1\right)\left(2 + \sum_{t=1}^{T-1}\sqrt{\frac{|\Omega|}{t}}(1 + \sqrt{\Phi \log T})\right) \\
&\quad + \sum_{t\in[T]}\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t) \\
&\le 2\left(\frac{20}{p_0^2 D} + 1\right)\left(1 + \sqrt{|\Omega| T}(1 + \sqrt{\Phi \log T})\right) \\
&\quad + \sum_{t\in[T]}\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t),
\end{aligned}$$

where in the final inequality, we have used the fact that $\sum_{t=1}^{T-1} 1/\sqrt{t} \le 2\sqrt{T}$.

From Theorem 1, we have $\mathbf{P}_\mu\left(\cap_{t\in[T]}\mathcal{B}_t \not\ni \mu\right) \le T^{1-\frac{3\Phi\sqrt{\Omega}}{56}}$. For $t \in [T]$, let $X_t \triangleq \mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t)$. Observe that $\mathbf{E}_{\mu^*}[X_t|h_t] = 0$ and $|X_t| \le 1$. Thus the sequence $\{X_t : t \in [T]\}$ is a bounded martingale difference sequence. Hence, from Azuma-Hoeffding (Boucheron et al., 2013), we obtain for $z \ge 0$,

$$\mathbf{P}_{\mu^*}\left(\sum_{t\in[T]}\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t) \ge z\right) < \exp\left(-\frac{2z^2}{T}\right).$$

Choosing $z = \sqrt{\alpha T \log T}$ for $\alpha \ge 0$ for $\alpha > 0$, we have

$$\mathbf{P}_{\mu^*}\left(\sum_{t\in[T]}\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t) \ge \sqrt{\alpha T \log T}\right) < \frac{1}{T^{2\alpha}}.$$

After choosing $\alpha = 4\Phi|\Omega|$ and taking the union bound, we obtain with probability at least $1 - T^{1-\frac{3\Phi\sqrt{\Omega}}{56}} - T^{-8\Phi|\Omega|}$, we have

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) \le 2\left(\frac{20}{p_0^2 D} + 1\right)\left(1 + \sqrt{|\Omega| T}(1 + 2\sqrt{\Phi \log T})\right).$$

□

Lemma 6. The $\mathfrak{Rai}$ algorithm satisfies

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) \leq \sum_{t\in[T]} \mathsf{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \|\mu^* - \gamma_t\|_1)) + \sum_{t\in[T]} \mathsf{Gap}(\gamma_t, \mathcal{B}_t)$$
$$+ \sum_{t\in[T]} \|\mu^* - \gamma_t\|_1 + \sum_{t\in[T]} \left(\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t)\right).$$

Proof. From the definition (3.2) of regret, we have

$$\begin{aligned}
\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) &= \mathsf{OPT}(\mu^*) \cdot T - \sum_{t\in[T]} v(\omega_t, a_t) \\
&= \mathsf{OPT}(\mu^*) \cdot T - \sum_{t\in[T]} \mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] && \text{(B.1)} \\
&\quad + \sum_{t\in[T]} \left(\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t)\right) \\
&= \sum_{t\in[T]} \left(\mathsf{OPT}(\mu^*) - V(\mu^*, \sigma[h_t])\right) && \text{(B.2)} \\
&\quad + \sum_{t\in[T]} \left(\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t)\right), && \text{(B.3)}
\end{aligned}$$

where in the last equality, we have used the fact that $\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] = V(\mu^*, \sigma[h_t])$. Moreover, note that

$$\begin{aligned}
\mathsf{OPT}(\mu^*) - V(\mu^*, \sigma[h_t]) &= \mathsf{OPT}(\mu^*) - V(\gamma_t, \sigma[h_t]) + V(\gamma_t, \sigma[h_t]) - V(\mu^*, \sigma[h_t]) \\
&= (\mathsf{OPT}(\mu^*) - \mathsf{OPT}(\gamma_t)) + (\mathsf{OPT}(\gamma_t) - V(\gamma_t, \sigma[h_t])) \\
&\quad + (V(\gamma_t, \sigma[h_t]) - V(\mu^*, \sigma[h_t])) \\
&= (\mathsf{OPT}(\mu^*) - \mathsf{OPT}(\gamma_t)) + \mathsf{Gap}(\gamma_t, \mathcal{B}_t) \\
&\quad + (V(\gamma_t, \sigma[h_t]) - V(\mu^*, \sigma[h_t])),
\end{aligned}$$

where in the final equality, we have used the fact that $\mathsf{OPT}(\gamma_t) - V(\gamma_t, \sigma[h_t]) =$

$\mathsf{Gap}(\gamma_t, \mathcal{B}_t)$. Substituting the preceding expression into (B.3) yields

$$\begin{aligned}\mathsf{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) = &\sum_{t\in[T]} (\mathsf{OPT}(\mu^*) - \mathsf{OPT}(\gamma_t)) + \sum_{t\in[T]} \mathsf{Gap}(\gamma_t, \mathcal{B}_t) \\ &+ \sum_{t\in[T]} (V(\gamma_t, \sigma[h_t]) - V(\mu^*, \sigma[h_t])) \\ &+ \sum_{t\in[T]} (\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t)) .\end{aligned}$$

Now, in Lemma 7, we prove $\mathsf{OPT}(\mu^*) - \mathsf{OPT}(\gamma_t) \leq \mathsf{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \|\mu^* - \gamma_t\|_1)) + \frac{1}{2} \cdot \|\mu^* - \gamma_t\|_1$. Furthermore, in Lemma 8, we show that $V(\gamma_t, \sigma[h_t]) - V(\mu^*, \sigma[h_t]) \leq \frac{1}{2}\|\mu^* - \gamma_t\|_1$. Putting it all together yields the lemma statement. □

Lemma 7. For any $\mu_1, \mu_2 \in \Delta(\Omega)$, we have

$$\mathsf{OPT}(\mu_1) - \mathsf{OPT}(\mu_2) \leq \mathsf{Gap}(\mu_1, \mathsf{B}_1(\mu_1, \|\mu_1 - \mu_2\|_1)) + \frac{1}{2} \cdot \|\mu_1 - \mu_2\|_1.$$

Proof. Fix $\mu_1, \mu_2 \in \Delta(\Omega)$. For $i \in \{1, 2\}$, let $\sigma_i \in \arg\max_{\sigma' \in \mathsf{Pers}(\mu_i)} V(\mu_i, \sigma')$. By definition, we have $\mathsf{OPT}(\mu_i) = V(\mu_i, \sigma_i)$.

Next, among all signaling mechanisms that are persuasive for all $\mu \in \mathsf{B}_1(\mu_1, \|\mu_1 - \mu_2\|_1)$, let σ_3 maximize $V(\mu_1, \sigma)$. Since σ_3 is persuasive for μ_2, we have $\mathsf{OPT}(\mu_2) = V(\mu_2, \sigma_2) \geq V(\mu_2, \sigma_3)$. Thus, we have

$$\begin{aligned}\mathsf{OPT}(\mu_1) - \mathsf{OPT}(\mu_2) &= V(\mu_1, \sigma_1) - V(\mu_2, \sigma_2) \\ &\leq V(\mu_1, \sigma_1) - V(\mu_2, \sigma_3) \\ &= V(\mu_1, \sigma_1) - V(\mu_1, \sigma_3) + V(\mu_1, \sigma_3) - V(\mu_2, \sigma_3) \\ &\leq \mathsf{Gap}(\mu_1, \mathsf{B}_1(\mu_1, \|\mu_1 - \mu_2\|_1)) + \frac{1}{2} \cdot \|\mu_1 - \mu_2\|_1.\end{aligned}$$

Here, the inequality follows from the definition of $\mathsf{Gap}(\cdot)$, and from Lemma 8. □

Lemma 8. For any $\mu_1, \mu_2 \in \Delta(\Omega)$ and any signaling mechanism σ, we have

$$|V(\mu_1, \sigma) - V(\mu_2, \sigma)| \leq \frac{1}{2} \cdot \|\mu_1 - \mu_2\|_1.$$

Proof. Fix $\mu_1, \mu_2 \in \Delta(\Omega)$. For any signaling mechanism σ that is persuasive under

μ_1, we have for any $x \in \mathbb{R}$,

$$|V(\mu_1,\sigma) - V(\mu_2,\sigma)| = \left| \sum_{\omega\in\Omega} (\mu_1(\omega) - \mu_2(\omega)) \left(\sum_{a\in A} \sigma(\omega,a)v(\omega,a) - x \right) \right|$$
$$\leq \|\mu_1 - \mu_2\|_1 \cdot \sup_{\omega\in\Omega} \left| \sum_{a\in A} \sigma(\omega,a)v(\omega,a) - x \right|,$$

where we have used the Hölder's inequality in the last line. Optimizing over x, together with the fact that the sender's valuations lie in $[0,1]$, yields the result. □

Lemma 9. For $t \in [T]$, on the event $\{\mu^* \in \mathcal{B}_t\}$, we have

$$\mathsf{Gap}(\gamma_t, \mathcal{B}_t) \leq \left(\frac{16}{p_0^2 D} \right) \epsilon_t.$$

Proof. On the event $\{\mu^* \in \mathcal{B}_t\}$, we have

$$\gamma_t(\omega) \geq \mu(\omega) - \|\gamma_t - \mu^*\|_1$$
$$\geq p_0 - \epsilon_t.$$

Thus, for $\epsilon_t < \frac{p_0}{2}$, we have $\min_\omega \gamma_t(\omega) \geq \frac{p_0}{2}$. Using the same argument as in Proposition 1, we then obtain

$$\mathsf{Gap}(\gamma_t, \mathcal{B}_t) = \mathsf{Gap}(\gamma_t, \mathsf{B}_1(\gamma_t, \epsilon_t)) \leq \left(\frac{4}{D \min_\omega \gamma_t(\omega)^2} \right) \epsilon_t \leq \left(\frac{16}{p_0^2 D} \right) \epsilon_t.$$

For $\epsilon_t > p_0/2$, the bound holds trivially since $16\epsilon_t/p_0^2 D > 1$. □

B.2.2 Proof of Theorem 3

Proof of Theorem 3. For a distribution $\mu \in \mathcal{B}_0$, define the event $\mathcal{E}_T(\mu)$ as

$$\mathcal{E}_T(\mu) = \left\{ h_T : \sigma^{\mathfrak{a}}[h_t] \in \mathsf{Pers}(\mu), \text{ for each } t \in [T] \right\}.$$

In words, under the event $\mathcal{E}_T(\mu)$, the signaling mechanism $\sigma^{\mathfrak{a}}[h_t]$ chosen by the algorithm $\mathfrak{a}$ after any history $h_t \in \mathcal{E}_T(\mu)$ is persuasive for the distribution μ. Since the algorithm $\mathfrak{a}$ is β_T-robustly persuasive, we obtain

$$\mathbf{P}_{\mu}\left(\mathcal{E}_T(\mu)\right) \geq 1 - \beta_T, \quad \text{for all } \mu \in \mathcal{B}_0.$$

Fix an $\epsilon \in (0, \frac{1-3p_0}{2})$ to be chosen later, and consider the distributions $\mu^* = (p_0, \frac{1-p_0}{2}, \frac{1-p_0}{2})$ and $\bar{\mu}_1 = \mu^* + \frac{\epsilon}{2}(e_1 - e_2)$ and $\bar{\mu}_2 = \mu^* + \frac{\epsilon}{2}(e_2 - e_1)$, where e_j is the belief that puts all its weight on state ω_j for $j \in \{1, 2\}$. Observe that $\mathbf{P}_{\mu^*}\left(\mathcal{E}_T(\mu^*)\right) \geq 1 - \beta_T$ since $\mu^* \in \mathcal{B}_0$. Note that for each $j \in \{1, 2\}$ and for all $\epsilon \in (0, \frac{1-3p_0}{2})$, we have $\bar{\mu}_j \in \mathcal{B}_0$ and hence $\mathbf{P}_{\bar{\mu}_j}\left(\mathcal{E}_T(\bar{\mu}_j)\right) \geq 1 - \beta_T$.

Now, on the event $\mathcal{E}_T(\mu^*) \cap \mathcal{E}_T(\bar{\mu}_1) \cap \mathcal{E}_T(\bar{\mu}_2)$, the signaling mechanism $\sigma^{\mathfrak{a}}[h_t]$ chosen by the algorithm after any history h_t is persuasive for all the distributions $\mu^*, \bar{\mu}_1, \bar{\mu}_2$. Thus on the event $\mathcal{E}_T(\mu^*) \cap \mathcal{E}_T(\bar{\mu}_1) \cap \mathcal{E}_T(\bar{\mu}_2)$, we have

$$T \cdot \mathsf{OPT}(\mu^*) - \sum_{t \in [T]} V(\mu^*, \sigma^{\mathfrak{a}}[h_t]) \geq T \cdot \mathsf{Gap}(\mu^*, \{\mu^*, \bar{\mu}_1, \bar{\mu}_2\}) \geq \frac{\epsilon T}{8 D p_0},$$

where the first inequality follows from the definition of Gap in (2.3), and the second inequality follows from Proposition 2.

Now, we have

$$\begin{aligned} 2\left|\mathbf{P}_{\mu^*}\left(\mathcal{E}_T(\bar{\mu}_1)\right) - \mathbf{P}_{\bar{\mu}_1}\left(\mathcal{E}_T(\bar{\mu}_1)\right)\right|^2 &\leq \sum_{t \in [T]} \mathsf{KL}\left(\mu^* || \bar{\mu}_1\right) \\ &= \frac{1-p_0}{2} \log\left(\frac{(1-p_0)^2}{(1-p_0)^2 - \epsilon^2}\right) T \\ &= \frac{1-p_0}{2} \log\left(1 + \frac{\epsilon^2}{(1-p_0)^2 - \epsilon^2}\right) T \\ &\leq \frac{1-p_0}{2} \left(\frac{\epsilon^2}{(1-p_0)^2 - \epsilon^2}\right) T, \end{aligned}$$

where the first inequality is Pinsker's inequality, and the first equality is from the definition of the Kullback-Leibler divergence, and the final inequality follows from $\log(1+x) \leq x$ for $x \geq 0$. Thus, for $\epsilon < \frac{1-p_0}{2}$, we obtain

$$2\left|\mathbf{P}_{\mu^*}\left(\mathcal{E}_T(\bar{\mu}_1)\right) - \mathbf{P}_{\bar{\mu}_1}\left(\mathcal{E}_T(\bar{\mu}_1)\right)\right|^2 \leq \frac{2\epsilon^2 T}{3(1-p_0)} \leq \epsilon^2 T,$$

where we have used $p_0 \le \frac{1}{|\Omega|} = \frac{1}{3}$ in the final inequality. Thus, we obtain that

$$\begin{aligned}\mathbf{P}_{\mu^*}\left(\mathcal{E}_T(\bar{\mu}_1)\right) &\geq \mathbf{P}_{\mu^*}\left(\mathcal{E}_T(\mu^*)\right) - \left|\mathbf{P}_{\mu^*}\left(\mathcal{E}_T(\bar{\mu}_1)\right) - \mathbf{P}_{\bar{\mu}_1}\left(\mathcal{E}_T(\bar{\mu}_1)\right)\right| \\ &\geq 1 - \beta_T - \epsilon\sqrt{\frac{T}{2}}.\end{aligned}$$

By the same argument, we obtain $\mathbf{P}_{\mu^*}\left(\mathcal{E}_T(\bar{\mu}_2)\right) \geq 1 - \beta_T - \epsilon\sqrt{\frac{T}{2}}$.

By the linearity of the obedience constraints, we obtain that if $\sigma \in \mathsf{Pers}(\bar{\mu}_1) \cap \mathsf{Pers}(\bar{\mu}_2)$, then $\sigma \in \mathsf{Pers}(\mu^*)$. Thus, we have $\mathcal{E}_T(\bar{\mu}_1) \cap \mathcal{E}_T(\bar{\mu}_2) \subseteq \mathcal{E}_T(\mu^*)$, and hence

$$\begin{aligned}\mathbf{P}_{\mu^*}(\mathcal{E}_T(\mu^*) \cap \mathcal{E}_T(\bar{\mu}_1) \cap \mathcal{E}_T(\bar{\mu}_2)) &= \mathbf{P}_{\mu^*}(\mathcal{E}_T(\bar{\mu}_1) \cap \mathcal{E}_T(\bar{\mu}_2)) \\ &\geq \mathbf{P}_{\mu^*}(\mathcal{E}_T(\bar{\mu}_1)) + \mathbf{P}_{\mu^*}(\mathcal{E}_T(\bar{\mu}_2)) - 1 \\ &\geq 1 - 2\beta_T - \epsilon\sqrt{2T}.\end{aligned}$$

Finally, by the Azuma-Hoeffding inequality, we obtain

$$\mathbf{P}_{\mu^*}\left(\sum_{t\in[T]} V(\mu^*, \sigma^{\mathfrak{a}}[h_t]) - \sum_{t\in[T]} v(\omega_t, a_t) < -\sqrt{T}\right) < e^{-1/2}.$$

Taken together, we obtain that with probability at least $1 - 2\beta_T - \epsilon\sqrt{2T} - e^{-1/2}$, we have

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{a}, T, \mu^*) = T \cdot \mathsf{OPT}(\mu^*) - \sum_{t\in[T]} v(\omega_t, a_t) \geq \frac{\epsilon T}{8Dp_0} - \sqrt{T}.$$

For $T \geq T_0 = \frac{1}{(1-3p_0)^2}$, choosing $\epsilon = \frac{1}{32\sqrt{T}} \leq \frac{1-3p_0}{2}$, we obtain, with probability at least $\frac{1}{3} - 2\beta_T$,

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{a}, T, \mu^*) = T \cdot \mathsf{OPT}(\mu^*) - \sum_{t\in[T]} v(\omega_t, a_t) \geq \sqrt{T}\left(\frac{1}{16Dp_0} - 1\right) \geq \frac{\sqrt{T}}{32Dp_0},$$

for $Dp_0 < 1/32$. □

B.3 Concentration Inequalities

In this section, we provide some concentration inequalities used in the proofs of our main results. We use the same notation as in the main text. The following lemma provides a bound on the ℓ_1-norm of the deviation of the empirical distribution from its mean.

The following lemma is a standard application of the Azuma-Hoeffding inequality (Boucheron et al., 2013), and used to obtain bounds on the regret.

Lemma 10. For all $\mu^* \in \mathcal{B}_0$ and $\alpha > 0$, we have

$$\mathbf{P}_{\mu^*}\left(\sum_{t\in[T]} \mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t) \geq \sqrt{\alpha T \log T}\right) < \frac{1}{T^{2\alpha}}.$$

Proof. For $t \in [T]$, let $X_t \triangleq \mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t)$. Observe that $\mathbf{E}_{\mu^*}[X_t|h_t] = 0$ and $|X_t| \leq 1$. Thus the sequence $\{X_t : t \in [T]\}$ is a bounded martingale difference sequence. Hence, from Azuma-Hoeffding (Boucheron et al., 2013), we obtain for $z \geq 0$,

$$\mathbf{P}_{\mu^*}\left(\sum_{t\in[T]} \mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t) \geq z\right) < \exp\left(-\frac{2z^2}{T}\right).$$

Choosing $z = \sqrt{\alpha T \log T}$ for $\alpha \geq 0$ yields the lemma statement. □

B.4 Examples of Persuasion Instances

In this section, we provide examples of persuasion instances that illuminate various aspects of our theoretical results.

B.4.1 Failure of the Regularity Condition

We begin with an example of an instance $\mathcal{I}_1$ in which the regularity condition does not hold, and in which any β-robustly persuasive algorithm incurs a linear regret. We establish this by proving that in this instance, the cost of robust persuasion $\mathsf{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \epsilon))$ is a constant independent of ϵ for all $\epsilon > 0$.

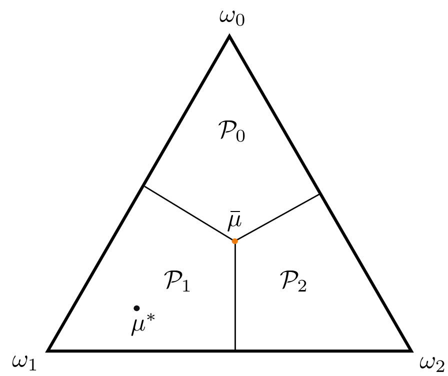


Figure B.1: The persuasion instance $\mathcal{I}_1$.

In the persuasion instance $\mathcal{I}_1$, the state space is given by $\Omega = \{\omega_0, \omega_1, \omega_2\}$ and the receiver has four actions $A = \{a_0, a_1, a_2, a_3\}$. The receiver's utility is given by $u(\omega_i, a_j) = \mathbf{I}\{i = j\} + \frac{1}{3}\mathbf{I}\{j = 3\}$ for $i \in \{0, 1, 2\}$ and $j \in \{0, 1, 2, 3\}$. The sender's payoff is given by $v(\omega_i, a_j) = \mathbf{I}\{j = 3\}$; in other words, the sender strictly prefers the receiver choosing action a_3 over any other action in all states. The sender's initial knowledge regarding the underlying state distribution is captured by $\mathcal{B}_0 = \mathsf{B}_1(\mu^*, \epsilon_0)$ for some $\epsilon_0 > 0$, where $\mu^* = (\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$.

The receiver's preferences can be depicted in Fig. B.1, with sets $\mathcal{P}_j$ for $j \in \{0, 1, 2\}$ denoting the set of beliefs for which the receiver finds it optimal to choose action a_j. On the other hand, the set of beliefs for which it is optimal for the receiver to choose the sender's preferred action a_3 is given by $\mathcal{P}_3 = \{\bar{\mu}\}$ where $\bar{\mu} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ (the orange central point in the figure). Since $\mathcal{P}_3$ has an empty interior, the first regularity condition fails for the instance $\mathcal{I}_1$.

If the distribution μ^* is known, the sender can use a signaling mechanism that induces $\bar{\mu}$ as the posterior belief with positive probability, causing the receiver to choose action a_3 leading to a positive payoff for the sender. Formally, under the distribution $\mu^* = (\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$, the optimal signaling mechanism is given by

$$\begin{aligned}
\sigma^*(\omega_1, a_3) &= 1 - \sigma^*(\omega_1, a_1) = \frac{1}{4}, \\
\sigma^*(\omega_0, a_3) &= \sigma^*(\omega_2, a_3) = 1, \\
\sigma^*(\omega, a) &= 0, \quad \text{otherwise.}
\end{aligned}$$

Under this mechanism, the sender's utility is given by $\mathsf{OPT}(\mu^*) = \frac{1}{2}$.

However, for any $\epsilon > 0$, the only recommendations that are robustly persuasive for all distributions in $\mathsf{B}_1(\mu^*, \epsilon)$ are a_0, a_1, a_2. Thus, any signaling mechanism that is persuasive for all distributions in $\mathsf{B}_1(\mu^*, \epsilon)$ can never recommend the sender's preferred action a_3, leading to the sender's payoff of zero. Hence, the difference in the sender's expected utility between using the optimal persuasive signaling mechanism for the distribution μ^* and using the optimal signaling mechanism that is persuasive for all distributions in $\mathsf{B}_1(\mu^*, \epsilon)$ is given by

$$\mathsf{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \epsilon)) = V(\mu^*, \sigma^*) - V(\mu^*, \hat{\sigma}) = \frac{1}{2}.$$

Thus, $\mathsf{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \epsilon))$ is a constant independent of $\epsilon > 0$. Using this bound on the cost of robust persuasion and an argument similar to the proof of Theorem 3, one can show that the regret of any β-robustly persuasive mechanism is of order $\Omega(T)$ with probability at least 1/3.

B.4.2 Linear Regret for 0-robustly Persuasive Mechanisms

In this section, we establish the necessity to consider β-robustly persuasive mechanisms with (small) $\beta > 0$ for obtaining meaningful regret bounds. This is demonstrated by a simple example of the persuasion instance $\mathcal{I}_2$ in which any 0-robustly persuasive algorithm necessarily incurs a linear regret.

In the persuasion instance $\mathcal{I}_2$, the state space is given by $\Omega = \{\omega_0, \omega_1\}$ and the receiver's action space is given by $A = \{a_0, a_1\}$. The receiver's utility is given by $u(\omega_i, a_j) = \mathbf{I}\{i = j\}$ for $i, j \in \{0, 1\}$, i.e., the receiver desires to "match" the action with the state. On the other hand, the sender strictly prefers the receiver choosing action a_0 over action a_1 in all states, i.e., $v(\omega_i, a_j) = \mathbf{I}\{j = 0\}$ for all $i, j \in \{0, 1\}$. The sender's initial knowledge regarding the distribution is captured by $\mathcal{B}_0 = \{(\frac{1}{2} - \alpha, \frac{1}{2} + \alpha) \colon \alpha \in [-\frac{1}{4}, +\frac{1}{4}]\}$.

For each $i \in \{0, 1\}$, the set of beliefs for which it is optimal for the receiver to choose action a_i is given by $\mathcal{P}_i$ where $\mathcal{P}_0 = \{(a, 1 - a) \colon a \in [\frac{1}{2}, 1]\}$ and $\mathcal{P}_1 = \{(a, 1 - a) \colon a \in [0, \frac{1}{2}]\}$. Note that the persuasion instance $\mathcal{I}_2$ satisfies both the regularity conditions.

Now, since all the distributions in $\mathcal{B}_0$ are absolutely continuous with respect to each other, any algorithm $\mathfrak{a}$ that is 0-robustly persuasive must select at each time $t \in [T]$ a signaling mechanism σ_t in the set $\mathsf{Pers}(\mathcal{B}_0)$. However, it is straightforward to verify that among all mechanisms that are persuasive for all distributions in $\mathcal{B}_0$, the one that maximizes sender's payoff is given by $\hat{\sigma}(\omega_0, a_0) = 1 - \hat{\sigma}(\omega_0, a_1) = 1, \hat{\sigma}(\omega_1, a_0) = 1 - \hat{\sigma}(\omega_1, a_1) = \frac{1}{3}$. For the distribution $\mu^* = (\frac{1}{2}, \frac{1}{2}) \in \mathcal{P}_0 \cap \mathcal{B}_0$, it follows that the sender's payoff under $\hat{\sigma}$ is $V(\mu^*, \hat{\sigma}) = \frac{2}{3}$.

On the other hand, since $\mu^* \in \mathcal{P}_0 \cap \mathcal{B}_0$, the signaling mechanism that recommends action a_0 in both states is persuasive for μ^*, and thus achieves an expected payoff of $\mathsf{OPT}(\mu^*) = 1$. Thus, we deduce that for the distribution μ^*, any 0-robustly persuasive algorithm must incur a constant regret of at least $\frac{1}{3}$ at each time leading to an overall regret linear in T.

Appendix C

Appendix to Chapter 4

C.1 Proofs from Section 4.4

C.1.1 Proofs from Section 4.4.1

Proof of Lemma 1. We prove the two statements corresponding to the no-history information model Φ_{no} and the full-history information model Φ_{full} separately.

1. No-history information model Φ_{no}: We prove the statement by showing that for any $\sigma \in \Sigma_k \cap \mathsf{Pers}(\Phi_{\mathsf{no}})$ for some k, there exists a $\widehat{\sigma} \in \Sigma_0 \cap \mathsf{Pers}(\Phi_{\mathsf{no}})$ with the same payoff for the sender.

Recall that, assuming all other receivers have followed their recommendations, a receiver's prior belief in the no-history information model is characterized by the invariant distribution $\pi = \mathsf{Inv}(\sigma)$. Note that π is an invariant distribution of $(\bar{h}_t^k, \bar{\omega}_t)$ under σ, assuming all receivers follow the recommendations.

Now, define $\widehat{\pi} \in \Delta(\Omega)$ as

$$\widehat{\pi}(\omega) \triangleq \sum_{h^k \in \mathcal{X}^k} \pi(h^k, \omega) \qquad \text{for all } \omega \in \Omega.$$

Define the signaling mechanism $\widehat{\sigma} \in \Sigma_0$ as follows: for all $\omega \in \Omega$ with $\widehat{\pi}(\omega) = \sum_{h^k \in \mathcal{X}^k} \pi(h^k, \omega) > 0$, let

$$\widehat{\sigma}(a|\omega) \triangleq \frac{\sum_{h^k \in \mathcal{X}^k} \pi(h^k, \omega)\sigma(a|h^k, \omega)}{\sum_{h^k \in \mathcal{X}^k} \pi(h^k, \omega)} = \frac{\sum_{h^k \in \mathcal{X}^k} \pi(h^k, \omega)\sigma(a|h^k, \omega)}{\widehat{\pi}(\omega)} \qquad \text{for all } a \in A,$$

and for any $\omega \in \Omega$ with $\widehat{\pi}(\omega) = 0$, let $\widehat{\sigma}$ recommend the receiver-optimal action at ω.

We first show that $\widehat{\pi}$ is an invariant distribution under $\widehat{\sigma}$ assuming all the receivers follow their recommendations. To see this, observe that if $\widehat{\pi}(\omega) = 0$ then $\sum_{h^k \in \mathcal{X}^k} \pi(h^k, \omega)\sigma(a|h^k, \omega) = 0$ for all (ω, a), and hence,

$$\begin{aligned}
&\sum_{(\omega_{-1},a_{-1})\in\mathcal{X}} \widehat{\pi}(\omega_{-1})\widehat{\sigma}(a_{-1}|\omega_{-1})p(\omega|\omega_{-1}, a_{-1}) \\
&= \sum_{(\omega_{-1},a_{-1})\in\mathcal{X}} \sum_{h^k\in\mathcal{X}^k} \pi(h^k, \omega_{-1})\sigma(a_{-1}|h^k, \omega_{-1})p(\omega|\omega_{-1}, a_{-1}) \\
&= \sum_{(\omega_{-1},a_{-1})\in\mathcal{X}} \sum_{h^{k-1}\in\mathcal{X}^{k-1}} \sum_{x_k\in\mathcal{X}} \pi(x_k, h^{k-1}, \omega_{-1})\sigma(a_{-1}|x_k, h^{k-1}, \omega_{-1})p(\omega|\omega_{-1}, a_{-1}) \\
&= \sum_{h^{k-1}\in\mathcal{X}^{k-1}} \sum_{(\omega_{-1},a_{-1})\in\mathcal{X}} \pi(h^{k-1}, \omega_{-1}, a_{-1}, \omega) \\
&= \widehat{\pi}(\omega).
\end{aligned}$$

Here, the first and the final equality follows from the definitions of $\widehat{\pi}$ and $\widehat{\sigma}$, the third equality follows from the balance equations (4.1) for $\pi = \mathsf{Inv}(\sigma)$, and in the second equality, we have written $h^k \in \mathcal{X}^k$ as $h^k = (x_k, h^{k-1})$ for $x_k \in \mathcal{X}$ and $h^{k-1} \in \mathcal{X}^{k-1}$. Thus, we see that $\widehat{\pi}$ satisfies the balance equations (4.1) under $\widehat{\sigma} \in \Sigma_0$, and thus $\widehat{\pi} = \mathsf{Inv}(\widehat{\sigma})$.

Given that $\widehat{\pi} = \mathsf{Inv}(\widehat{\sigma})$, under the information model Φ_{no}, a receiver's prior belief is given by $\widehat{\pi}$. Upon receiving a recommendation $\bar{s}_t = a$ from $\widehat{\sigma}$, the receiver's posterior belief is given by

$$\frac{\widehat{\pi}(\omega)\widehat{\sigma}(a|\omega)}{\sum_{\omega'\in\Omega} \widehat{\pi}(\omega')\widehat{\sigma}(a|\omega')} = \frac{\sum_{h^k\in\mathcal{X}^k} \pi(h^k, \omega)\sigma(a|h^k, \omega)}{\sum_{\omega'\in\Omega} \sum_{h^k\in\mathcal{X}^k} \pi(h^k, \omega')\sigma(a|h^k, \omega')}.$$

The latter expression is exactly the receiver's belief under the information model Φ_{no} with the signaling mechanism $\sigma \in \Sigma_k \cap \mathsf{Pers}(\Phi_{\mathsf{no}})$. Since the receiver has the same belief, and because $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}})$, it follows that $\widehat{\sigma} \in \mathsf{Pers}(\Phi_{\mathsf{no}})$ as well. Finally, the sender's payoff under $\widehat{\sigma}$ is given by

$$\sum_{\omega\in\Omega, a\in A} \widehat{\pi}(\omega)\widehat{\sigma}(a|\omega)v(\omega, a) = \sum_{h^k\in\mathcal{X}^k} \sum_{\omega\in\Omega, a\in A} \pi(h^k, \omega)\sigma(a|h^k, \omega)v(\omega, a),$$

and thus the signaling mechanism $\widehat{\sigma}$ yields the sender the same payoff as under σ.

2. Full-history information model Φ_{full}: The proof for the full-history information model follows along similar lines: we show that for any $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{full}}) \cap \Sigma_k$ for some $k \geq 2$, there exists a $\widehat{\sigma} \in \Sigma_1 \cap \mathsf{Pers}(\Phi_{\mathsf{full}})$ with the same payoff for the sender.

Recall that, assuming that all receivers have followed their recommendations, the invariant distribution of $(\bar{h}_t^k, \bar{\omega}_t)$ under σ is given by $\pi = \mathsf{Inv}(\sigma) \in \Delta(\mathcal{X}^k \times \Omega)$. Define $\widehat{\pi} \in \Delta(\mathcal{X} \times \Omega)$ as

$$\widehat{\pi}(x, \omega) \triangleq \sum_{h^{k-1} \in \mathcal{X}^{k-1}} \pi\left((h^{k-1}, x), \omega\right), \quad \text{for all } x \in \mathcal{X} \text{ and } \omega \in \Omega,$$

and define the signaling mechanism $\widehat{\sigma}$ as follows: for all $(x, \omega) \in \mathcal{X} \times \Omega$ with $\widehat{\pi}(x, \omega) > 0$, let

$$\widehat{\sigma}(a|x, \omega) = \frac{\sum_{h^{k-1} \in \mathcal{X}^{k-1}} \pi((h^{k-1}, x), \omega)\sigma(a|(h^{k-1}, x), \omega)}{\widehat{\pi}(x, \omega)} \qquad \text{for all } a \in A,$$

and for any $(x, \omega) \in \mathcal{X} \times \Omega$ with $\widehat{\pi}(x, \omega) = 0$, let $\widehat{\sigma}$ recommend the receiver-optimal action at ω.

We next show that $\widehat{\pi}$ is an invariant distribution under $\widehat{\sigma}$ assuming all the receivers follow their recommendations. To see this, observe that if $\widehat{\pi}(x, \omega) = 0$ then $\sum_{h^{k-1} \in \mathcal{X}^{k-1}} \pi\left((h^{k-1}, x), \omega\right) \sigma(a|(h^{k-1}, x), \omega) = 0$ for all (x, ω) and $a \in A$, and hence,

$$\begin{aligned}
&\sum_{x \in \mathcal{X}} \widehat{\pi}(x, \omega_{-1})\widehat{\sigma}(a_{-1}|x, \omega_{-1})p(\omega|\omega_{-1}, a_{-1}) \\
&\quad = \sum_{x \in \mathcal{X}} \sum_{h^{k-1} \in \mathcal{X}^{k-1}} \pi\left((h^{k-1}, x), \omega_{-1}\right) \sigma(a_{-1}|(h^{k-1}, x), \omega_{-1})p(\omega|\omega_{-1}, a_{-1}) \\
&\quad = \sum_{x \in \mathcal{X}} \sum_{h^{k-2} \in \mathcal{X}^{k-2}} \sum_{x' \in \mathcal{X}} \pi\left((x', h^{k-2}, x), \omega_{-1}\right) \sigma(a_{-1}|(x', h^{k-2}, x), \omega_{-1})p(\omega|\omega_{-1}, a_{-1}) \\
&\quad = \sum_{x \in \mathcal{X}} \sum_{h^{k-2} \in \mathcal{X}^{k-2}} \pi\left((h^{k-2}, x, \omega_{-1}, a_{-1}), \omega\right) \\
&\quad = \widehat{\pi}(\omega_{-1}, a_{-1}, \omega).
\end{aligned}$$

Here, the first and the final equality follows from the definitions of $\widehat{\pi}$ and $\widehat{\sigma}$, the third equality follows from the balance equations (4.1) for $\pi = \mathsf{Inv}(\sigma)$, and in the second equality, we have written $h^{k-1} \in \mathcal{X}^{k-1}$ as $h^{k-1} = (x', h^{k-2})$ for $x' \in \mathcal{X}$ and $h^{k-2} \in \mathcal{X}^{k-2}$. Thus, we see that $\widehat{\pi}$ satisfies the balance equations (4.1) under $\widehat{\sigma} \in \Sigma_1$ and thus $\widehat{\pi} = \mathsf{Inv}(\widehat{\sigma})$.

Now, under the information model Φ_{full}, after any history with $\bar{x}_{t-1} = x \in \mathcal{X}$, a receiver's prior belief that $\bar{\omega}_t = \omega \in \Omega$ is given by $p(\omega|x)$. Suppose $\widehat{\sigma}$ sends a recommendation $\bar{s}_t = a \in A$. This can only happen if there exists an $\omega' \in \Omega$ with $p(\omega'|x)\widehat{\sigma}(a|x,\omega') > 0$. Then, the receiver's posterior belief that $\bar{\omega}_t = \omega$ is given by

$$\frac{p(\omega|x)\widehat{\sigma}(a|x,\omega)}{\sum_{\omega'\in\Omega} p(\omega'|x)\widehat{\sigma}(a|x,\omega')}.$$

Thus, for $\widehat{\sigma} \in \mathsf{Pers}(\Phi_{\mathsf{full}})$, we require the following inequality to hold for all $x \in \mathcal{X}$:

$$\sum_{\omega\in\Omega} p(\omega|x)\widehat{\sigma}(a|x,\omega)\partial u(\omega,a,a') \geq 0, \quad \text{for all } a, a' \in A. \tag{C.1}$$

To see why this holds, observe first that for any $x = (\omega_{-1}, a_{-1}) \in \mathcal{X}$ and $\omega \in \Omega$, we have from the balance equations (4.1) that

$$\widehat{\pi}(x,\omega) = \sum_{x'\in\mathcal{X}} \widehat{\pi}(x',\omega_{-1})\widehat{\sigma}(a_{-1}|x',\omega_{-1})p(\omega|x) = \widehat{\mu}(x)p(\omega|x),$$

where we have defined $\widehat{\mu}(x) \triangleq \sum_{x'\in\mathcal{X}} \widehat{\pi}(x',\omega_{-1})\widehat{\sigma}(a_{-1}|x',\omega_{-1})$.

For any $x \in \mathcal{X}$ with $\widehat{\mu}(x) = 0$, we obtain that $\widehat{\pi}(x,\omega) = 0$ for all ω. Then, by the definition of $\widehat{\sigma}$, we have $\widehat{\sigma}(a|x,\omega) > 0$ only if action a is receiver-optimal at ω, and hence $\partial u(\omega,a,a') \geq 0$ for all $a' \in A$. Thus, we get (C.1) holds for any $x \in \mathcal{X}$ with $\widehat{\mu}(x) = 0$.

For any $x = (\omega_{-1}, a_{-1}) \in \mathcal{X}$ with $\widehat{\mu}(x) > 0$, we have for all $a, a' \in A$,

$$
\begin{aligned}
&\widehat{\mu}(x)\left(\sum_{\omega\in\Omega} p(\omega|x)\widehat{\sigma}(a|x,\omega)\partial u(\omega,a,a')\right)\\
=&\sum_{\omega\in\Omega}\widehat{\pi}(x,\omega)\widehat{\sigma}(a|x,\omega)\partial u(\omega,a,a')\\
=&\sum_{\omega\in\Omega}\sum_{h^{k-1}\in\mathcal{X}^{k-1}}\pi\left((h^{k-1},x),\omega\right)\sigma(a|(h^{k-1},x),\omega)\partial u(\omega,a,a')\\
=&\sum_{\omega\in\Omega}\sum_{h^{k-1}\in\mathcal{X}^{k-1}}\left(\sum_{x'\in\mathcal{X}}\pi((x',h^{k-1}),\omega_{-1})\sigma(a_{-1}|(x',h^{k-1}),\omega_{-1})p(\omega|x)\right)\\
&\sigma(a|(h^{k-1},x),\omega)\partial u(\omega,a,a')\\
=&\sum_{h^{k-1}\in\mathcal{X}^{k-1}}\sum_{x'\in\mathcal{X}}\pi((x',h^{k-1}),\omega_{-1})\sigma(a_{-1}|(x',h^{k-1}),\omega_{-1})\\
&\left(\sum_{\omega\in\Omega}p(\omega|x)\sigma(a|(h^{k-1},x),\omega)\partial u(\omega,a,a')\right)
\end{aligned}
$$

Here, we have used the definitions of $\widehat{\pi}$ and $\widehat{\sigma}$ in the second equality along with the fact that if $\widehat{\pi}(x,\omega) = 0$ then so is $\sum_{h^{k-1}\in\mathcal{X}^{k-1}} \pi((h^{k-1},x),\omega)\sigma(a|(h^{k-1},x),\omega)$ for all $a \in A$. The third equality follows from (4.1) for $\pi = \mathsf{Inv}(\sigma)$. Now, for any $h^{k-1} \in \mathcal{X}^{k-1}$ with $\sum_{x'\in\mathcal{X}} \pi((x',h^{k-1}),\omega_{-1})\sigma(a_{-1}|(x',h^{k-1}),\omega_{-1}) > 0$, we must have $\sum_{\omega\in\Omega} p(\omega|x)\sigma(a|(h^{k-1},x),\omega)\partial u(\omega,a,a') \geq 0$ since $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{full}}) \cap \Sigma_k$. From this, we conclude that (C.1) also holds for any $x \in \mathcal{X}$ with $\widehat{\mu}(x) > 0$.

Taken together, we obtain that (C.1) holds for all $x \in \mathcal{X}$, and hence $\widehat{\sigma} \in \mathsf{Pers}(\Phi_{\mathsf{full}})$. Finally, the sender's payoff under $\widehat{\sigma}$ is given by

$$
\begin{aligned}
&\sum_{x\in\mathcal{X}}\sum_{\omega\in\Omega,a\in A}\widehat{\pi}(x,\omega)\widehat{\sigma}(a|x,\omega)v(\omega,a)\\
=&\sum_{h^{k-1}\in\mathcal{X}^{k-1}}\sum_{x\in\mathcal{X}}\sum_{\omega\in\Omega,a\in A}\pi((h^{k-1},x),\omega)\sigma(a|(h^{k-1},x),\omega)v(\omega,a),
\end{aligned}
$$

and thus the signaling mechanism $\widehat{\sigma}$ yields the sender the same payoff as under σ. □

Proof of Proposition 4. We prove the statement for $\mathsf{MPP}(\Phi_{\mathsf{full}})$. A similar argument, with minor modifications, obtains the equivalence of $\mathsf{MPP}(\Phi_{\mathsf{no}})$ and $\mathsf{LP}(\mathsf{no})$; we omit it for brevity.

From Lemma 2, we know that there exists an optimal signaling mechanism for $\mathsf{MPP}(\Phi_{\mathsf{full}})$ within the set Σ_1. The proof shows that for any $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{full}}) \cap \Sigma_1$, there exists a corresponding feasible solution z to $\mathsf{LP}(\mathsf{full})$ whose objective value equals the sender's payoff under σ, and conversely, for any feasible solution z to $\mathsf{LP}(\mathsf{full})$, there exists a signaling mechanism $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{full}}) \cap \Sigma_1$ with sender's payoff equaling the objective value at z.

To begin, fix $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{full}}) \cap \Sigma_1$, and let $\pi = \mathsf{Inv}(\sigma) \in \Delta(\mathcal{X} \times \Omega)$. The balance equations (4.1) for σ are given by

$$\sum_{x \in \mathcal{X}} \pi(x, \omega)\sigma(a|x, \omega)p(\omega'|\omega, a) = \pi(\omega, a, \omega'), \quad \text{for all } (\omega, a) \in \mathcal{X} \text{ and } \omega' \in \Omega.$$

Define $z(x, \omega, a) \triangleq \pi(x, \omega)\sigma(a|x, \omega)$ for x and $(\omega, a) \in \mathcal{X}$, and note that z constitutes the joint distribution of two consecutive state-action pairs under π. From the balance equations for π, it is straightforward to verify that z satisfies the first equality in $\mathsf{LP}(\mathsf{full})$. The second equality follows because z is distribution. Finally, since $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{full}}) \cap \Sigma_1$, we obtain for all $x_1 \in \mathcal{X}$ and $a, a' \in A$,

$$\sum_{\omega \in \Omega} p(\omega|x_{-1})\sigma(a|x_{-1}, \omega)\partial u(\omega, a, a') \geq 0.$$

Thus, we obtain for all $x_{-1} = (\omega_{-1}, a_{-1}) \in \mathcal{X}$ and for all $a, a' \in A$,

$$\begin{aligned}
&\sum_{\omega \in \Omega} z(x_{-1}, \omega, a)\partial u(\omega, a, a') \\
&\quad = \sum_{\omega \in \Omega} \pi(x_{-1}, \omega)\sigma(a|x_{-1}, \omega)\partial u(\omega, a, a') \\
&\quad = \sum_{\omega \in \Omega} \left(\sum_{x' \in \mathcal{X}} \pi(x', \omega_{-1})\sigma(a_{-1}|x', \omega_{-1})p(\omega|\omega_{-1}, a_{-1}) \right) \sigma(a|x_{-1}, \omega)\partial u(\omega, a, a') \\
&\quad = \left(\sum_{x' \in \mathcal{X}} \pi(x', \omega_{-1})\sigma(a_{-1}|x', \omega_{-1}) \right) \left(\sum_{\omega \in \Omega} p(\omega|x_{-1})\sigma(a|x_{-1}, \omega)\partial u(\omega, a, a') \right) \geq 0.
\end{aligned}$$

Thus, we obtain that z satisfies the inequality in $\mathsf{LP}(\mathsf{full})$. Finally, since z represents the joint distribution of two consecutive state-action pairs under π, we conclude that the $\mathsf{LP}(\mathsf{full})$ objective evaluated at z equals the sender's payoff under σ. This concludes the first part of the statement.

Conversely, suppose z is a feasible solution for $\mathsf{LP}(\mathsf{full})$. Define $\pi(x,\omega) \triangleq \sum_{a \in A} z(x,\omega,a)$ for $x \in \mathcal{X}$ and $\omega \in \Omega$. Observe that the second equality of $\mathsf{LP}(\mathsf{full})$ implies that $\pi \in \Delta(\mathcal{X} \times \Omega)$. Next, define the signaling mechanism $\sigma \in \Sigma_1$ as follows: for all $x \in \mathcal{X}$ and $\omega \in \Omega$ with $\pi(x,\omega) > 0$, let

$$\sigma(a|x,\omega) \triangleq \frac{z(x,\omega,a)}{\sum_{a' \in A} z(x,\omega,a')} = \frac{z(x,\omega,a)}{\pi(x,\omega)}.$$

For $x \in \mathcal{X}$ and $\omega \in \Omega$ with $\pi(x,\omega) = 0$, let $\sigma(\cdot|x,\omega)$ recommend any receiver-optimal action at ω.

We first show that π is invariant under σ. To see this, observe that if $\pi(x,\omega) = 0$ then so is $z(x,\omega,a)$ for all $a \in A$. Thus, we have

$$\begin{aligned}
\sum_{x \in \mathcal{X}} \pi(x,\omega)\sigma(a|x,\omega)p(\omega'|\omega,a) &= \sum_{x \in \mathcal{X}} z(x,\omega,a)p(\omega'|\omega,a) \\
&= \sum_{a' \in A} z(\omega,a,\omega',a') \\
&= \pi(\omega,a,\omega'),
\end{aligned}$$

where the first and third equality follow from the definitions of π and σ, and the second equality follows because z is feasible for $\mathsf{LP}(\mathsf{full})$. Hence, π satisfies the balance equations (4.1) under σ, and thus $\pi = \mathsf{Inv}(\sigma)$.

Next, for any $x_{-1} = (\omega_{-1}, a_{-1}) \in \mathcal{X}$ and all $a, a' \in A$, we have

$$\begin{aligned}
&\sum_{\omega \in \Omega} z(x_{-1}, \omega, a) \partial u(\omega, a, a') \\
&= \sum_{\omega \in \Omega} \pi(x_{-1}, \omega) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a') \\
&= \sum_{\omega \in \Omega} \left(\sum_{x' \in \mathcal{X}} \pi(x', \omega_{-1}) \sigma(a_{-1}|x', \omega_{-1}) p(\omega|\omega_{-1}, a_{-1}) \right) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a') \\
&= \left(\sum_{x' \in \mathcal{X}} \pi(x', \omega_{-1}) \sigma(a_{-1}|x', \omega_{-1}) \right) \left(\sum_{\omega \in \Omega} p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a') \right) \\
&= \left(\sum_{x' \in \mathcal{X}} z(x', \omega_{-1}, a_{-1}) \right) \left(\sum_{\omega \in \Omega} p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a') \right)
\end{aligned}$$

Here, the first and final equality follows from the definitions of π and σ and the second equality follows from the balance equations for π. Thus, for all $x_{-1} = (\omega_{-1}, a_{-1}) \in \mathcal{X}$ with $\sum_{x' \in \mathcal{X}} z(x', \omega_{-1}, a_{-1}) > 0$, from the feasibility of z for $\mathsf{LP}(\mathsf{full})$, we obtain

$$\sum_{\omega \in \Omega} p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a') \geq 0.$$

and hence a receiver, after observing $\bar{x}_{t-1} = x_{-1}$, would find it optimal to adopt action a if recommended by σ. On the other hand, if $\sum_{x' \in \mathcal{X}} z(x', \omega_{-1}, a_{-1}) = 0$ for some $x_{-1} = (\omega_{-1}, a_{-1}) \in \mathcal{X}$, then for all $\omega \in \Omega$, we have

$$\pi(x_{-1}, \omega) = \sum_{a \in A} z(x_{-1}, \omega, a) = \sum_{x' \in \mathcal{X}} z(x', \omega_{-1}, a_{-1}) p(\omega|\omega_{-1}, a_{-1}) = 0.$$

Thus, we obtain that if $\bar{x}_{t-1} = x_{-1}$, then σ recommends a receiver-optimal action for each value of $\bar{\omega}_t = \omega$. Thus, again, the receiver finds it optimal to follow the recommendation. Taken together, we conclude that $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{full}})$. Finally, we notice that the sender's payoff under σ equals the $\mathsf{LP}(\mathsf{full})$ objective evaluated at z. This completes the proof of the converse statement.

Summarizing the two parts, we obtain that the sender's problem $\mathsf{MPP}(\Phi_{\mathsf{full}})$ can be equivalently formulated as $\mathsf{LP}(\mathsf{full})$.

□

C.1.2 Proofs from Section 4.4.2

Proof of Lemma 2. We first show that for any $\ell \geq 1$, $\mathsf{Pers}(\Phi_{\mathsf{full}}) \subseteq \mathsf{Pers}(\Phi_\ell)$. To see this, let $\sigma \in \Sigma_k \cap \mathsf{Pers}(\Phi_{\mathsf{full}})$. Define $\widehat{\sigma}$ to be the signaling mechanism that, at each time t, in addition to recommending an action according to σ also truthfully reveals $\bar{h}_t^\ell$. Since the information of the receiver under $\widehat{\sigma}$ in the model Φ_ℓ is the same as that under σ in the model Φ_{full}, we conclude that it is optimal for the receiver to follow the recommended action. Since this is true no matter the realization of the slice $\bar{h}_t^\ell$, the receiver should find it optimal to follow the recommendation even without being informed about the realization. In other words, the receiver should find it optimal to follow the recommendations of σ in the model Φ_ℓ, and hence $\sigma \in \mathsf{Pers}(\Phi_\ell)$. Thus, we conclude $\mathsf{Pers}(\Phi_{\mathsf{full}}) \subseteq \mathsf{Pers}(\Phi_\ell)$ for $\ell \geq 1$. A similar argument yields $\mathsf{Pers}(\Phi_\ell) \subseteq \mathsf{Pers}(\Phi_{\ell+1}) \subseteq \mathsf{Pers}(\Phi_{\mathsf{no}})$. This in turn implies the ordering of the sender's optimal payoff in the corresponding information models.

□

Proof of Theorem 4. Fix $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$. We begin by proving (1) $\implies$ (2). Suppose $\widehat{\sigma} \in \mathsf{Pers}(\Phi_{\mathsf{full}}) \cap \Sigma_1$ results in the same invariant distribution over the state-action pairs as under σ, and induces the posterior belief μ_a after any action recommendation $a \in A_+$. Let $\pi = \mathsf{Inv}(\sigma) \in \Delta(\Omega)$ and $\widehat{\pi} = \mathsf{Inv}(\widehat{\sigma}) \in \Delta(\mathcal{X} \times \Omega)$ be the corresponding invariant distributions.

Now, under $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$, the joint distribution of the state and the action in steady state is given by

$$\eta(x) \coloneqq \pi(\omega)\sigma(a|\omega) = \tau_a \mu_a(\omega), \quad \text{for all } x = (\omega, a) \in \mathcal{X}.$$

Since the joint distribution of the state and action is the same under $\widehat{\pi}$ and π, we obtain $\widehat{\pi}(x, \omega) = \eta(x) p(\omega|x)$ for all $x \in \mathcal{X}$ and $\omega \in \Omega$. In particular, only the state-action pairs $x \in \mathcal{X}_+$ occur with positive probability under $\widehat{\sigma}$, and only actions in the set A_+ are recommended by it. Consider a receiver in the full-history information model Φ_{full}, who observes the past state action pair $\bar{x}_{t-1} = x \in \mathcal{X}_+$ and receives an action recommendation of $a \in A_+$ from $\widehat{\sigma}$. The receiver's posterior belief that the state equals $\omega \in \Omega$ is given by

$$\frac{\widehat{\pi}(x,\omega)\widehat{\sigma}(a|x,\omega)}{\sum_{\omega'\in\Omega}\widehat{\pi}(x,\omega')\widehat{\sigma}(a|x,\omega')} = \frac{\eta(x)p(\omega|x)\widehat{\sigma}(a|x,\omega)}{\sum_{\omega'\in\Omega}\eta(x)p(\omega'|x)\widehat{\sigma}(a|x,\omega')} = \frac{p(\omega|x)\widehat{\sigma}(a|x,\omega)}{\sum_{\omega'\in\Omega}p(\omega'|x)\widehat{\sigma}(a|x,\omega')}.$$

Since $\widehat{\sigma}$ induces the same posteriors as σ, this preceding expression equal $\mu_a(\omega)$. Thus, we obtain for all $\omega \in \Omega$, $a \in A_+$ and $x \in \mathcal{X}_+$,

$$\frac{p(\omega|x)\widehat{\sigma}(a|x,\omega)}{\sum_{\omega'\in\Omega}p(\omega'|x)\widehat{\sigma}(a|x,\omega')} = \mu_a(\omega).$$

Define $\lambda(a|x) \coloneqq \sum_{\omega\in\Omega} p(\omega|x)\widehat{\sigma}(a|x,\omega) \geq 0$ for all $a \in A_+$ and $x \in \mathcal{X}_+$. Thus, we obtain for $\omega \in \Omega$, $a \in A_+$, and $x \in \mathcal{X}_+$,

$$p(\omega|x)\widehat{\sigma}(a|x,\omega) = \lambda(a|x)\mu_a(\omega). \tag{C.2}$$

Summing over all $a \in A_+$, we obtain for all $\omega \in \Omega$, and $x \in \mathcal{X}_+$,

$$p(\omega|x) = \sum_{a\in A_+} p(\omega|x)\widehat{\sigma}(a|x,\omega) = \sum_{a\in A_+} \lambda(a|x)\mu_a(x).$$

Thus, the first equation of (2) in the theorem statement holds. Next, the balance equations for $\widehat{\pi} = \mathsf{Inv}(\widehat{\sigma})$ yield for all $\omega, \bar{\omega} \in \Omega$ and $a \in A$,

$$\sum_{x\in\mathcal{X}} \widehat{\pi}(x,\omega)\widehat{\sigma}(a|x,\omega)p(\bar{\omega}|\omega,a) = \widehat{\pi}(\omega,a,\bar{\omega}).$$

Substituting $\widehat{\pi}(x,\omega) = \eta(x)p(\omega|x)$ into the balance equations, and after canceling common terms on both sides, we obtain

$$\sum_{x\in\mathcal{X}} \eta(x)p(\omega|x)\widehat{\sigma}(a|x,\omega) = \eta(\omega,a), \quad \text{for all } (\omega,a) \in \mathcal{X}.$$

Noticing that $\eta(x) = 0$ for $x \notin \mathcal{X}_+$ and using (C.2), we obtain

$$\sum_{x\in\mathcal{X}_+} \eta(x)\lambda(a|x)\mu_a(\omega) = \eta(\omega,a), \quad \text{for all } \omega \in \Omega \text{ and } a \in A_+.$$

Finally, substituting $\eta(\omega, a) = \tau_a \mu_a(\omega)$ for $\omega \in \Omega$ and $a \in A_+$, we obtain

$$\sum_{x=(\omega',a')\in\mathcal{X}_+} \tau_{a'}\mu_{a'}(\omega')\lambda(a|x)\mu_a(\omega) = \tau_a\mu_a(\omega), \quad \text{for all } \omega \in \Omega \text{ and } a \in A_+.$$

After canceling common terms from both sides, we conclude

$$\sum_{x=(\omega',a')\in\mathcal{X}_+} \tau_{a'}\mu_{a'}(\omega')\lambda(a|x) = \tau_a, \quad \text{for all } a \in A_+.$$

Thus, the second equation of (2) in the theorem statement also holds. Thus, we obtain the forward implication.

To prove the converse, i.e., (2) $\implies$ (1), assume for $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$, there exists weights $\{\lambda(a|x) \geq 0 : a \in A_+, x \in \mathcal{X}_+\}$ satisfying the two equations in the theorem statement. Define the mechanism $\widehat{\sigma} \in \Sigma_1$ as follows: for each $x \in \mathcal{X}_+$, let

$$\widehat{\sigma}(a|x,\omega) \triangleq \frac{\lambda(a|x)\mu_a(\omega)}{\sum_{a'\in A_+}\lambda(a'|x)\mu_{a'}(\omega)} = \frac{\lambda(a|x)\mu_a(\omega)}{p(\omega|x)}, \quad \text{for all } a \in A_+,$$

whenever the denominator is positive. In all other cases, let $\widehat{\sigma}$ recommend a receiver-optimal action at ω.

We first show that the distribution $\widehat{\pi} \in \Delta(\mathcal{X} \times \Omega)$ defined below is invariant under $\widehat{\sigma}$, assuming the receivers follow their recommendations:

$$\widehat{\pi}(x,\omega) \triangleq \eta(x)p(\omega|x),$$

where $\eta(\omega, a) = \tau_a\mu_a(\omega)$ is the joint distribution of the state and action in steady state under σ. To see this, observe that for $a \in A_+$ and $\omega, \bar{\omega} \in \Omega$,

$$\begin{aligned}
\sum_{x \in \mathcal{X}} \widehat{\pi}(x,\omega)\widehat{\sigma}(a|x,\omega)p(\bar{\omega}|\omega,a) &= \sum_{x \in \mathcal{X}} \eta(x)p(\omega|x)\widehat{\sigma}(a|x,\omega)p(\bar{\omega}|\omega,a) \\
&= \sum_{x \in \mathcal{X}_+} \eta(x)p(\omega|x)\widehat{\sigma}(a|x,\omega)p(\bar{\omega}|\omega,a) \\
&= \sum_{x \in \mathcal{X}_+} \eta(x)\lambda(a|x)\mu_a(\omega)p(\bar{\omega}|\omega,a) \\
&= \sum_{x=(\omega',a') \in \mathcal{X}_+} \tau_{a'}\mu_{a'}(\omega')\lambda(a|x)\mu_a(\omega)p(\bar{\omega}|\omega,a) \\
&= \tau_a\mu_a(\omega)p(\bar{\omega}|\omega,a) \\
&= \eta(\omega,a)p(\bar{\omega}|\omega,a) \\
&= \widehat{\pi}(\omega,a,\bar{\omega}).
\end{aligned}$$

Here, we use the definition of $\widehat{\sigma}$ in the third equality, and the fact that $\lambda(a|x)$ satisfies the second equation of the theorem statement in the fifth equality. For any $a \notin A_+$, it is straightforward to verify that both sides of the above equality equal 0 and hence are equal. Thus, we see that $\widehat{\pi}$ satisfies the balance equations for $\widehat{\sigma}$. From the definition of $\widehat{\pi}$ we see that the joint distribution of state and action in steady state under $\widehat{\sigma}$ is the same as that under σ, and thus only the states in $\mathcal{X}_+$ occur with positive probability under $\widehat{\pi}$.

Finally, we show that $\widehat{\sigma} \in \mathsf{Pers}(\Phi_{\mathsf{full}})$. For any $x \notin \mathcal{X}_+$, $\widehat{\sigma}$ recommends a receiver-optimal action at each state ω, and hence it is optimal for any receiver to follow the recommendation upon seeing the past state-action pair to be x. For any $x \in \mathcal{X}_+$, the mechanism $\widehat{\sigma}$ only sends recommendations in the set A_+, and the posterior belief of a receiver upon obtaining a recommendation $a \in A_+$ is given by

$$\begin{aligned}
\frac{\widehat{\pi}(x,\omega)\widehat{\sigma}(a|x,\omega)}{\sum_{\omega' \in \Omega} \widehat{\pi}(x,\omega')\widehat{\sigma}(a|x,\omega')} &= \frac{\eta(x)p(\omega|x)\widehat{\sigma}(a|x,\omega)}{\sum_{\omega' \in \Omega} \eta(x)p(\omega'|x)\widehat{\sigma}(a|x,\omega')} \\
&= \frac{p(\omega|x)\widehat{\sigma}(a|x,\omega)}{\sum_{\omega' \in \Omega} p(\omega'|x)\widehat{\sigma}(a|x,\omega')} \\
&= \frac{\lambda(a|x)\mu_a(\omega)}{\sum_{\omega' \in \Omega} \lambda(a|x)\mu_a(\omega')} \\
&= \mu_a(\omega).
\end{aligned}$$

Thus, we obtain that for any $x \in \mathcal{X}_+$, a receiver's posterior belief upon receiving a recommendation $a \in A_+$ is the same as that under σ. Since σ is persuasive, we conclude it is optimal for the receiver to follow the recommendation. Thus, we have $\widehat{\sigma} \in \mathsf{Pers}(\Phi_{\mathsf{full}})$.

Finally, we prove the last statement in the theorem by showing the existence of weights satisfying the conditions in part (2). First, since $p(\omega|x) \in \mathsf{Conv}(\mathcal{B}_{\mathsf{no}})$ for each $x \in \mathcal{X}_+$, there exists a set of non-negative weights $\{\lambda(a|x) \geq 0 : a \in A_+, x \in \mathcal{X}_+\}$ such that

$$\sum_{a \in A_+} \lambda(a|x)\mu_a(\omega) = p(\omega|x), \quad \text{for all } x \in \mathcal{X}_+ \text{ and } \omega \in \Omega.$$

Now, from the balance equations for $\pi = \mathsf{Inv}(\sigma)$, we obtain

$$\sum_{\omega \in \Omega} \sum_{a \in A} \pi(\omega)\sigma(a|\omega)p(\bar{\omega}|\omega, a) = \pi(\bar{\omega}).$$

By definition, we have $\pi(\omega)\sigma(a|\omega) = \tau_a \mu_a(\omega)$ for each $(\omega, a) \in \mathcal{X}_+$ and $\pi(\omega)\sigma(a|\omega) = 0$ for $x \notin \mathcal{X}_+$. Substituting, we obtain

$$\sum_{(\omega,a) \in \mathcal{X}_+} \tau_a \mu_a(\omega) p(\bar{\omega}|\omega, a) = \sum_{a \in A_+} \tau_a \mu_a(\bar{\omega}).$$

Writing the transition kernel as a convex combination of the posterior beliefs, we obtain

$$\sum_{(\omega,a) \in \mathcal{X}_+} \tau_a \mu_a(\omega) \left(\sum_{a' \in A_+} \lambda(a'|\omega, a)\mu_{a'}(\bar{\omega}) \right) = \sum_{a \in A_+} \tau_a \mu_a(\bar{\omega}).$$

After collecting like-terms and rearranging, we obtain

$$\sum_{a \in A_+} \mu_a(\bar{\omega}) \left(\tau_a - \sum_{(\omega,a') \in \mathcal{X}_+} \tau_{a'} \mu_{a'}(\omega) \lambda(a|\omega, a') \right) = 0.$$

From the linear independence of $\{\mu_a : a \in A_+\}$, we conclude that for all $a \in A_+$

$$\tau_a = \sum_{(\omega,a')\in\mathcal{X}_+} \tau_{a'}\mu_{a'}(\omega)\lambda(a|\omega,a').$$

Thus, the weights satisfy both the conditions in (2) of the theorem statement. □

C.2 Proofs from the Section 4.5

In this section, we provide the missing proofs from Section 4.5. Throughout, we use the same notation as in that section.

Proof of Proposition 5. The proof of the proposition is similar to that of Proposition 4, and we only highlight the parts that are different.

Consider a signaling mechanism $\sigma \in \Sigma_k \cap \mathsf{Pers}(\Phi_1)$. Assuming all (other) receivers follow the recommendations provided by σ, let $\pi \in \Delta(\mathcal{X}^{k+1} \times \Omega)$ denote the steady state distribution of $(\bar{h}_t^{k+1}, \bar{\omega}_t)$. Furthermore, define $z : \mathcal{X}^{k+2} \to [0,1]$ as follows:

$$z(x,h,\omega,a) \triangleq \pi(x,h,\omega)\sigma(a|h,\omega), \quad \text{for all } x \in \mathcal{X},\ h \in \mathcal{X}^k,\ \omega \in \Omega \text{ and } a \in A.$$

It is straightforward to see that z denotes the joint-distribution in steady state of $(\bar{x}_{t-k-1}, \bar{h}_t^k, \bar{\omega}_t, \bar{a}_t)$, and hence satisfies the final equality in (4.4). Next, we have

$$\begin{aligned}\sum_{x'\in\mathcal{X}} z(x',h,\omega,a)p(\omega'|\omega,a) &= \sum_{x'\in\mathcal{X}} \pi(x',h,\omega)\sigma(a|h,\omega)p(\omega'|\omega,a)\\ &= \pi(h,\omega,a,\omega')\\ &= \sum_{a'\in A} z(h,\omega,a,\omega',a')\end{aligned}$$

for all $h \in \mathcal{X}^k$, $(\omega,a) \in \mathcal{X}$ and $\omega' \in \Omega$. Here, we have used the definition of z in the first and the final equality, and the second equality follows from the steady-state balance equations. This yields the first equality in (4.4). Next, under Φ_1, the receiver at time t observes $\bar{h}_{t-1}^k = h \in \mathcal{X}^k$, and hence her posterior belief after being recommended an action $a \in A$ is given by

$$\frac{\sum_{x\in\mathcal{X}} z(h,x,\omega,a)}{\sum_{x\in\mathcal{X}}\sum_{\omega'\in\Omega} z(h,x,\omega',a)}.$$

Writing the persuasiveness constraint under this posterior belief yields the inequality in (4.4). Similarly, the value of the objective is readily seen to equal the sender's payoff in steady-state under σ. Finally, the second (bilinear) equality is obtained (after some algebra) from the following equalities:

$$\begin{aligned} z(x,h,\omega,a)\cdot\left(\sum_{a'\in A} z(x',h,\omega,a)\right) &= \pi(x,h,\omega)\sigma(a|h,\omega)\cdot\pi(x',h,\omega) \\ &= \pi(x',h,\omega)\sigma(a|h,\omega)\cdot\pi(x,h,\omega) \\ &= z(x',h,\omega,a)\cdot\left(\sum_{a'\in A} z(x,h,\omega,a)\right). \end{aligned}$$

for all $x,x' \in \mathcal{X}$, $h \in \mathcal{X}^k$, $\omega \in \Omega$ and $a \in A$. Thus, we conclude that for any $\sigma \in \Sigma_k \cap \mathsf{Pers}(\Phi_1)$, there exists a feasible solution to the bilinear program whose objective value equals the sender's payoff under σ.

Conversely, let z denote any feasible solution to (4.4). Define $\pi(x,h,\omega) := \sum_{a\in A} z(x,h,\omega,a)$ for $x \in \mathcal{X}, h \in \mathcal{X}^k$ and $\omega \in \Omega$. Observe that the final equality of (4.4) implies that $\pi \in \Delta(\mathcal{X}^{k+1}\times\Omega)$. Next, define the signaling mechanism $\sigma \in \Sigma_k$ as follows: for any $h \in \mathcal{X}^k$ and $\omega \in \Omega$, if there exists an $x \in \mathcal{X}$ with $\pi(x,h,\omega) > 0$, then let

$$\sigma(a|h,\omega) := \frac{z(x,h,\omega,a)}{\sum_{a'\in A} z(x,h,\omega,a')} = \frac{z(x,h,\omega,a)}{\pi(x,h,\omega)}.$$

On the other hand, if $\pi(x,h,\omega) = 0$ for all $x \in \mathcal{X}$, then let $\sigma(\cdot|h,\omega)$ recommend any receiver-optimal action at ω. Note that the bilinear constraint on z implies that the definition of σ in the former case does not depend on the choice of $x \in \mathcal{X}$, and thus, σ is indeed an element of the set Σ_k. Using a similar argument as in Proposition 4, it can be shown that (1) π is an invariant distribution of $(\bar{h}_t^{k+1}, \bar{\omega}_t)$ under $\sigma \in \Sigma_k$, assuming all receivers follow their recommendations; and (2) furthermore, $\sigma \in \mathsf{Pers}(\Phi_1)$. Finally, it is straightforward to see that the sender's payoff under σ equals the objective value at z of (4.4). Thus, we conclude that for any feasible solution of (4.4), there exists a signaling mechanism $\sigma \in \Sigma_k \cap \mathsf{Pers}(\Phi_1)$ with the sender's payoff equal to the objective value at z.

Taken together, we conclude that the sender's problem $\mathsf{MPP}(\Phi_1, \Sigma_k)$ is equivalent to the bilinear program (4.4).

□

Proof of Lemma 3. Let $\sigma \in \mathsf{RP}(\epsilon)$ for some fixed $\epsilon > 0$, and let $\ell \geq 0$ be such that $d_\ell(\sigma) \leq \epsilon$. Consider the information model Φ_ℓ, and assume the receivers follow the action recommendations. From the perspective of a receiver at time t, the relevant information about the history $\bar{h}_{t-\ell}$ is the value $\bar{x}_{t-\ell-1}$, as earlier state-action pairs do not affect the subsequent transitions. If $\bar{x}_{t-\ell-1} = x \in \mathcal{X}$, the distribution of $\bar{\omega}_t$ (and hence the receiver's belief) is given by $Q^\sigma(x, \omega)$. Thus, the receiver's belief lies within $d_\ell(\sigma)$ of the invariant distribution $\pi = \mathsf{Inv}(\sigma)$. Since $\sigma \in \mathsf{RP}(\epsilon)$ and $d_\ell(\epsilon) \leq \epsilon$, we obtain that it is optimal for this receiver to follow the recommendation made by σ. Thus, we obtain $\sigma \in \mathsf{Pers}(\Phi_\ell)$.

To prove the bound in the lemma statement, we note that since $\sigma \in \mathsf{RP}(\epsilon) \subseteq \Sigma_0$, it corresponds to a stationary Markov policy, and hence the induced Markov chain over the states is ergodic by Assumption 2. The result is then obtained using the following bound on the mixing time of this chain (Levin and Peres, 2017, Theorem 12.4):

$$\ell \geq \frac{1}{\gamma_\star} \log\left(\frac{2}{\epsilon \pi_{\min}(\sigma)}\right) \implies d_\ell(\sigma) \leq \epsilon,$$

where $\gamma_\star(\sigma)$ is the absolute spectral gap of the underlying Markov chain (i.e., the smallest value of $1 - |\lambda|$ over all non-unit eigenvalues λ of the transition kernel matrix under σ), and $\pi_{\min}(\sigma) = \min_\omega \pi(\omega)$. Note that $\pi(\omega) > 0$ for all $\omega \in \Omega$ from Assumption 2, and hence $\pi_{\min}(\sigma)$ is well defined. □

Proof of Lemma 4. We begin by proving the first part of the lemma statement. Given a set S, beliefs $\{\mu_s : s \in S\}$ and the weights $w_s \geq 0$ as in the lemma statement, define the distribution $\pi \in \Delta(\Omega)$ as $\pi(\omega) = \sum_{s \in S} w_s \mu_s(\omega)$. Furthermore, define the signaling mechanism σ as follows: for any ω with $\pi(\omega) > 0$, let σ send signal $s \in S$

at state ω with probability

$$\sigma(s|\omega) := \frac{w_s \mu_s(\omega)}{\pi(\omega)} = \frac{w_s \mu_s(\omega)}{\sum_{s' \in S} w_{s'} \mu_{s'}(\omega)},$$

and for any ω with $\pi(\omega) = 0$, let σ send an arbitrary fixed signal $s_0 \in S$ at state ω.

We claim that π is an invariant distribution under the signaling mechanism σ, assuming each receiver chooses the action a_s after receiving a signal $s \in S$. (We show below that this is indeed optimal for the receiver in the information model Φ_{no}.) This follows from observing that π satisfies the corresponding balance equations, as we show next. For each $\omega \in \Omega$, we have

$$\begin{aligned}
\sum_{\omega' \in \Omega} \pi(\omega') \left(\sum_{s \in S} \sigma(s|\omega') p(\omega|\omega', a_s) \right) &= \sum_{\omega' \in \Omega} \sum_{s \in S} \pi(\omega') \sigma(s|\omega') p(\omega|\omega', a_s) \\
&= \sum_{\omega' \in \Omega} \sum_{s \in S} w_s \mu_s(\omega') p(\omega|\omega', a_s) \\
&= \sum_{s \in S} w_s \mu_s(\omega) \\
&= \pi(\omega).
\end{aligned}$$

Here, the second equality follows from the definition of σ as well as the fact that if $\pi(\omega') = 0$, then so is $w_s \mu_s(\omega')$ for all $s \in S$. The third equality follows the assumptions on the weights w_s, and the final equality follows from the definition of π. We observe that the left-hand side of the preceding equation denotes the state distribution after a transition from a state distributed according to the π, with a signal s being shared according to σ and the receiver choosing action a_s. The equality states that this distribution equals π, and hence π is an invariant under σ.

Thus, under the information model Φ_{no}, assuming all other receivers follow the recommendations from σ, each receiver's prior belief about the state equals π. The posterior belief of such a receiver that the state is ω upon receiving a signal $s \in S$ is given by Bayes' rule as

$$\frac{\pi(\omega)\sigma(s|\omega)}{\sum_{\omega' \in \Omega} \pi(\omega')\sigma(s|\omega')} = \frac{w_s \mu_s(\omega)}{\sum_{\omega'} w_s \mu_s(\omega')} = \mu_s(\omega).$$

Here, the first equality holds because, using the definitions of π and σ, we have

$\pi(\omega)\sigma(s|\omega) = w_s\mu_s(\omega)$ for all ω. Thus, each receiver's posterior belief upon receiving a signal $s \in S$ is μ_s. Since $\mu_s \in \mathcal{P}_{a_s}$, we conclude that choosing action a_s after receiving the signal s is indeed optimal for the receiver. Thus, we obtain that the signaling mechanism σ is indeed persuasive under the no-history information model. This concludes the proof of the first part of the lemma statement.

To show the converse, let $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$ and let $\pi = \mathsf{Inv}(\sigma)$ denote the invariant distribution under π. For each $a \in A$, let $w_a := \sum_{\omega\in\Omega} \pi(\omega)\sigma(a|\omega)$. By construction, we have $\sum_{a\in A} w_a = 1$.

For $a \in A$ with $w_a = 0$, let μ_a be an arbitrary belief in $\mathcal{P}_a$. For $a \in A$ with $w_a > 0$, define $\mu_a(\omega) := \frac{\pi(\omega)\sigma(a|\omega)}{w_a} = \frac{\pi(\omega)\sigma(a|\omega)}{\sum_{\omega'\in\Omega}\pi(\omega')\sigma(a|\omega')}$ for all $\omega \in \Omega$. Because $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}}) \cap \Sigma_0$ is persuasive under the no-history information model, we have $\sum_{\omega\in\Omega} \pi(\omega)\sigma(a|\omega)\partial u(\omega, a, a') \geq 0$ for all $a, a' \in A$. From the definitions of w_a and μ_a, we have $\pi(\omega)\sigma(a|\omega) = w_a\mu_a(\omega)$ for all $\omega \in \Omega$ and $a \in A$. Consequently, we obtain $\sum_{\omega\in\Omega} w_a\mu_a(\omega)\partial u(\omega, a, a') \geq 0$ for all $a, a' \in A$. This in turn implies that $\mu_a \in \mathcal{P}_a$ holds also for any $a \in A$ with $w_a > 0$.

Next, for all $\omega \in \Omega$, we have

$$\begin{aligned}\sum_{a\in A} w_a\mu_a(\omega) = \pi(\omega)& \\ &= \sum_{\omega'\in\Omega}\sum_{a\in A} \pi(\omega')\sigma(a|\omega')p(\omega|\omega', a) \\ &= \sum_{a\in A}\sum_{\omega'\in\Omega} w_a\mu_a(\omega')p(\omega|\omega', a).\end{aligned}$$

Here, the first and the third equality follows from the definitions of w_a and μ_a and the second equality follows from the balance equations (4.1) because $\pi = \mathsf{Inv}(\sigma)$.

Finally, in the no-history information model and under the signaling mechanism σ, a receiver's prior belief is given by π. For any $a \in A$ with $w_a = 0$, the signaling mechanism never sends signal $a \in A$ under π. For each $a \in A$ with $w_a > 0$, a receiver's posterior belief upon receiving signal $a \in A$ equals $\frac{\pi(\omega)\sigma(a|\omega)}{\sum_{\omega'\in\Omega}\pi(\omega')\sigma(a|\omega')} = \mu_a(\omega)$. Thus, we conclude that $\{w_a, \mu_a : a \in A\}$ satisfies all the conditions in the lemma statement.

□

Proof of Theorem 5. To begin, let $\sigma \in \mathsf{Pers}(\Phi_{\mathsf{no}})$ denote the optimal signaling mechanism in the no-history information model Φ_{no}. Let $\pi = \mathsf{Inv}(\sigma)$ denote the invariant distribution under σ, and let $\pi(\omega) = \sum_a \pi(\omega, a)$ denote the marginal over the states. From Lemma 4 we know there exist weights $w_a \geq 0$, with $\sum_{a \in A} w_a = 1$, and beliefs $\mu_a \in \mathcal{P}_a$ for $a \in A$, satisfying $\pi(\omega, a) = w_a \mu_a(\omega)$ and

$$\sum_{\omega, a} w_a \mu_a(\omega) p(\cdot|\omega, a) = \sum_a w_a \mu_a. \tag{C.3}$$

Let $A_+ = \{a \in A : \sum_{\omega \in \Omega} \pi(\omega)\sigma(a|\omega) > 0\}$ denote the set of actions that are recommended with positive probability under σ. It is straightforward to show that $A_+ = \{a : w_a > 0\}$.

Construction of a signaling mechanism: We begin by constructing a signaling mechanism $\widehat{\sigma}$ and show it to be persuasive in the no-information model Φ_{no}; subsequently, we prove the stronger claim of ϵ-robust persuasiveness. First, using Assumption 3, for any $a \in A_+$, let $\eta_a \in \mathcal{P}_a$ be such that $\mathsf{B}_1(\eta_a, D) \subseteq \mathcal{P}_a$. For some small $\delta \in [0, 1]$, whose exact value we will set later to obtain robustness, define $\xi_a = (1 - \delta)\mu_a + \delta\eta_a$ for all $a \in A_+$. Since $\mu_a, \eta_a \in \mathcal{P}_a$ and the latter set is convex, we obtain that $\mathsf{B}_1(\xi_a, \delta D) \subseteq \mathcal{P}_a$. Next, let e_ω denote the belief that puts all its weight on the state $\omega \in \Omega$.

We seek to construct a signaling mechanism $\widehat{\sigma}$ which sends signals in the set $S = A_+ \cup \Omega$, such that in the model Φ_{no}, the posterior belief upon receiving a signal $s = a \in A_+$ is ξ_a, whereas upon receiving a signal $s = \omega \in \Omega$, the posterior belief is e_ω. Let $a_s = a$ if $s = a \in A_+$ and $a_s = a_\omega$ for $s = \omega \in \Omega$, where a_ω denotes an optimal action for the receiver at state ω. Using Lemma 4, there exists a signaling mechanism $\widehat{\sigma}$ inducing the aforementioned beliefs in steady state if there exist weights $\{\hat{w}_s : s \in S\}$ with $\sum_{s \in S} \hat{w}_s = 1$, such that

$$\sum_{a \in A_+} \sum_{\omega} \hat{w}_a \xi_a(\omega) p(\cdot|\omega, a) + \sum_{\omega} \hat{w}_\omega p(\cdot|\omega, a_\omega) = \sum_{a \in A_+} \hat{w}_a \xi_a + \sum_{\omega} \hat{w}_\omega e_\omega. \tag{C.4}$$

To produce such weights, we first define $\{\hat{w}_a\}$ in terms of $\{\hat{w}_\omega\}$. Let $\hat{w}_a = (1 - \sum_\omega \hat{w}_\omega) w_a$ for each $a \in A_+$. Since $\sum_a w_a = 1$, it follows that the weights $\{\hat{w}_s\}$ sum to one as well. Further, to simplify expressions, let $\rho \triangleq \sum_\omega \hat{w}_\omega$. Then, after moving all terms containing $\hat{w}_\omega$ on one side, the condition (C.4) becomes

$$\frac{1}{1-\rho}\left(\sum_\omega \hat{w}_\omega e_\omega - \sum_\omega \hat{w}_\omega p(\cdot|\omega, a_\omega)\right) = \sum_{a\in A_+}\sum_\omega w_a \xi_a(\omega) p(\cdot|\omega, a) - \sum_{a\in A_+} w_a \xi_a$$
$$= \delta\left(\sum_{a\in A_+}\sum_\omega w_a \eta_a(\omega) p(\cdot|\omega, a) - \sum_{a\in A_+} w_a \eta_a\right), \tag{C.5}$$

where, in the second equality, we have used $\xi_a = (1-\delta)\mu_a + \delta\eta_a$, along with the fact that $\{\mu_a, w_a\}_a$ satisfy (C.3).

In Lemma 11, we show that there exists $y = (y_\omega \geq 0 : \omega \in \Omega)$ satisfying

$$\sum_\omega y_\omega e_\omega - \sum_\omega y_\omega p(\cdot|\omega, a_\omega) = \sum_{a\in A_+}\sum_\omega w_a \eta_a(\omega) p(\cdot|\omega, a) - \sum_{a\in A_+} w_a \eta_a. \tag{C.6}$$

For any such y, we obtain that $\hat{w}_\omega = \frac{\delta y_\omega}{1+\delta\|y\|_1}$ and $\rho = \sum_\omega \hat{w}_\omega = \frac{\delta\|y\|_1}{1+\delta\|y\|_1}$ form a solution to (C.5), and hence, there exist weights satisfying (C.4).

Thus, by Lemma 4, we obtain the existence of a history-independent signaling mechanism $\widehat{\sigma}$ sending signals $s \in S = A_+ \cup \Omega$, such that in the no-history information model Φ_{no}, the posterior beliefs lie in the set $\{\xi_a : a \in A_+\} \cup \{e_\omega : \omega \in \Omega\}$. The mechanism $\widehat{\sigma}$ sends signals with the following probabilities: for each $\omega \in \Omega$:

$$\widehat{\sigma}(s|\omega) \triangleq \begin{cases} \frac{\hat{w}_a \xi_a(\omega)}{\sum_{a'} \hat{w}_{a'} \xi_{a'}(\omega) + \hat{w}_\omega} & \text{for } s \in A_+; \\ \frac{\hat{w}_\omega}{\sum_{a'} \hat{w}_{a'} \xi_{a'}(\omega) + \hat{w}_\omega} & \text{if } s = \omega; \\ 0, & \text{otherwise.} \end{cases} \tag{C.7}$$

Here, we interpret the signal $s = a \in A_+$ as an direct recommendation to choose action a. On the other hand, the signal $s = \omega \in \Omega$ fully reveals the state and is interpreted as a recommendation to choose the action a_ω. Since $\xi_a \in \mathcal{P}_a$ for each $a \in A_+$ and a_ω is an optimal action for the receiver at state $\omega \in \Omega$, we conclude that the mechanism $\widehat{\sigma}$ is persuasive, in the sense that, the receiver finds it optimal to follow the recommendation.

Note that Lemma 4 implies that the invariant distribution $\widehat{\pi}$ under $\widehat{\sigma}$ is given by $\widehat{\pi}(\omega, a) = \hat{w}_a \xi_a(\omega) \mathbf{I}\{a \in A_+\} + \hat{w}_\omega \mathbf{I}\{a = a_\omega\}$ for each $\omega \in \Omega$ and $a \in A$. Let

$\widehat{\pi}(\omega) \triangleq \sum_{a\in A} \widehat{\pi}(\omega, a) = \sum_{a\in A_+} \hat{w}_a \xi_a(\omega) + \hat{w}_\omega$.

Robustness: Next, we show that $\widehat{\sigma} \in \mathsf{RP}(\epsilon)$. Suppose a receiver's belief about $\bar{\omega}_t$ is given by a distribution $\pi' \in \Delta(\Omega)$, with $\|\pi' - \widehat{\pi}\|_1 \leq \epsilon$. Upon receiving a signal $s = \omega \in S$ from the signaling mechanism $\widehat{\sigma}$, it is straightforward to see that the receiver's belief about $\bar{\omega}_t$ continues to update to e_ω, and hence a_ω is still optimal for the receiver on receiving signal $s = \omega$.

On the other hand, upon receiving a signal $s = a \in A_+$, the receiver's belief about $\bar{\omega}_t$ updates to ξ'_a, obtained via Bayes' rule as $\xi'_a(\omega) = \frac{\mu'(\omega)\widehat{\sigma}(a|\omega)}{\sum_{\omega'} \mu'(\omega')\widehat{\sigma}(a|\omega')}$. Using a similar argument as in the proof of Proposition 1, we obtain that the following bound on the ℓ_1 distance between ξ'_a and ξ_a:

$$\|\xi'_a - \xi_a\|_1 \leq 2 \left(\sup_{\omega\in\Omega} \frac{\xi_a(\omega)}{\widehat{\pi}(\omega)} \right) \cdot \|\pi' - \widehat{\pi}\|_1.$$

Since $\widehat{\pi}(\omega) = \sum_{a\in A} \widehat{\pi}(\omega, a) = \sum_{a\in A_+} \hat{w}_a \xi_a(\omega) + \hat{w}_\omega$ for each $\omega \in \Omega$, we have for each $a \in A_+$,

$$\sup_{\omega\in\Omega} \frac{\xi_a(\omega)}{\widehat{\pi}(\omega)} = \sup_{\omega\in\Omega} \frac{\xi_a(\omega)}{\sum_{a'\in A_+} \hat{w}_{a'} \xi_{a'}(\omega) + \hat{w}_\omega} \leq \frac{1}{\hat{w}_a} \leq \frac{1}{(1-\rho)w_{\min}},$$

where we use the fact that $\hat{w}_a = (1-\rho)w_a$ for $a \in A_+$, and define $w_{\min} \triangleq \min_{a\in A_+} w_a$. Thus, we obtain

$$\|\xi'_a - \xi_a\|_1 \leq \frac{2\|\pi' - \widehat{\pi}\|_1}{(1-\rho)w_{\min}} \leq \frac{2(1+\delta\|y\|_1)\epsilon}{w_{\min}},$$

where we have substituted $\rho = \frac{\delta\|y\|_1}{1+\delta\|y\|_1}$, and used $\|\pi' - \widehat{\pi}\|_1 \leq \epsilon$.

For $\epsilon \leq \frac{w_{\min} D}{2(1+\|y\|_1)}$, choosing $\delta = \frac{2\epsilon}{w_{\min} D - 2\epsilon\|y\|_1} \in [0, 1]$, we obtain that $\|\xi'_a - \xi_a\|_1 \leq \delta D$ and hence $\xi'_a \in \mathsf{B}_1(\eta_a, \delta D) \subseteq \mathcal{P}_a$. Hence, starting with a prior π' with $\|\pi' - \widehat{\pi}\| \leq \epsilon$, the posterior belief upon receiving a signal $s = a \in A_+$ lies in the set $\mathcal{P}_a$, implying that the action a continues to be optimal for the receiver. Taken together, the signaling mechanism $\widehat{\sigma}$ is persuasive for all beliefs $\pi' \in \mathsf{B}_1(\widehat{\pi}, \epsilon)$, and hence is ϵ-robustly persuasive.

Bound on sender's payoff: Finally, we provide a bound on the sender's expected utility under the signaling mechanism $\widehat{\sigma}$, as follows:

$$\begin{aligned}
\sum_{\omega\in\Omega}\sum_{a\in A}\widehat{\pi}(\omega,a)v(\omega,a) &= \sum_{\omega}\sum_{a\in A_+}\hat{w}_a\xi_a(\omega)v(\omega,a)+\sum_{\omega}\hat{w}_\omega v(\omega,a_\omega)\\
&\geq \sum_{\omega}\sum_{a\in A_+}\hat{w}_a\xi_a(\omega)v(\omega,a)\\
&\geq (1-\rho)(1-\delta)\sum_{\omega}\sum_{a\in A_+}w_a\mu_a(\omega)v(\omega,a)\\
&= (1-\rho)(1-\delta)\mathsf{OPT}(\Phi_{\mathsf{no}}),
\end{aligned}$$

where in the second inequality, we use $\hat{w}_a=(1-\rho)w_a$ and $\xi_a\geq(1-\delta)\mu_a$.

Substituting for δ and ρ, we obtain the sender's payoff is lower-bounded by

$$\frac{1-\delta}{1+\delta\|y\|_1}\mathsf{OPT}(\Phi_{\mathsf{no}})=\left(1-\frac{2(1+\|y\|_1)}{w_{\min}D}\epsilon\right)\mathsf{OPT}(\Phi_{\mathsf{no}}).$$

In Lemma 11, we show that there exists a solution $y\geq 0$ to (C.6) satisfying $\|y\|_1\leq\frac{2(1+\tau)\sqrt{|\Omega|}}{s_f}$. Thus, we obtain that the sender's expected payoff under $\widehat{\sigma}$ is lower-bounded by

$$\left(1-\frac{2\epsilon}{w_{\min}D}\left(1+\frac{2(1+\tau)\sqrt{|\Omega|}}{s_f}\right)\right)\cdot\mathsf{OPT}(\Phi_{\mathsf{no}}).$$

This completes the proof. □

The following lemma is used in the proof of Theorem 5.

Lemma 11. Consider the following linear program:

$$\min_{y\geq 0}\quad\sum_{\omega\in\Omega}y_\omega$$

$$\sum_{\omega\in\Omega}y_\omega e_\omega-\sum_{\omega\in\Omega}y_\omega p(\cdot|\omega,a_\omega)=\sum_{a\in A_+}\sum_{\omega\in\Omega}w_a\eta_a(\omega)p(\cdot|\omega,a)-\sum_{a\in A_+}w_a\eta_a. \tag{C.8}$$

The preceding linear program is feasible, and its optimal solution is upper bounded by $\frac{2(1+\tau)\sqrt{|\Omega|}}{s_f}$, where $\tau=\max_\omega 1/\nu_f(\omega)$, and s_f is the smallest positive singular value of $I-P_f$.

Proof. To show the feasibility of the linear program (C.8), we begin by casting it into a matrix form. Let $P_a \in \mathbb{R}^{|\Omega|\times|\Omega|}$ be the matrix with $P_a(\omega,\omega') \triangleq p(\omega'|\omega,a)$. Then, the LP (C.8) can be recast as

$$\begin{aligned} \min_{y\in\mathbb{R}^{|\Omega|}} \quad & \mathbf{1}^T \cdot y \\ y^T(I-P_f) = & \sum_{a\in A_+} w_a \eta_a^T (P_a - I) \\ y \geq & \ 0. \end{aligned}$$

where y, and $\mathbf{1} \in \mathbb{R}^{\Omega}$ is the all-one vector.

First, note that since P_f is ergodic (from Assumption 2), there is a unique stationary distribution ν_f such that $\nu_f^T(I-P_f) = 0$ and $\nu_f(\omega) > 0$ for all $\omega \in \Omega$. Thus, if y is a solution to the set of linear equalities above, then so is $y + c\nu_f$ for any $c \in \mathbb{R}$. By choosing $c > 0$ large enough, we obtain a non-negative solution $y + c\nu_f$ from any such y. Thus, to prove LP feasibility, it suffices to show that the linear equation has a solution. To see this, observe that since P_f and $\{P_a\}_{a\in A}$ are transition kernels, we have $(I-P_f)\mathbf{1} = (I-P_a)\mathbf{1} = 0$, and hence $\mathbf{1}$ does not lie in the row span of $I-P_f$ and $I-P_a$ for any $a \in A$. This implies that $\mathbf{1}$ is orthogonal to vector $\sum_{a\in A_+} w_a\eta_a^T(P_a - I)$. Since P_f is ergodic, we have $\mathsf{rank}(I-P_f) = |\Omega|-1$ and hence the vector $\sum_{a\in A_+} w_a\eta_a^T(P_a - I)$ lies in the row span of $I-P_f$. Therefore, the equation $y^T(I-P_f) = \sum_{a\in A_+} w_a\eta_a^T(P_a-I)$ has a feasible solution. This completes the proof of feasibility of LP (C.8).

To obtain a bound on the optimal value, let y denote an optimal solution to LP (C.8), and we decompose y as $y = u + k\nu_f$, where $u \in \mathbb{R}^{|\Omega|}$ satisfies $u^T\nu_f = 0$. Because $\nu_f^T(I-P_f) = 0$, we note that u lies in the column-space of $I - P_f$.

Letting $n \coloneqq \mathsf{rank}(I-P_f) = |\Omega| - 1$, consider the singular value decomposition $I - P_f = \sum_{i=1}^n s_i U_i V_i^T$, where $s_1 \geq s_2 \geq \cdots \geq s_n$ denote the (positive) singular values of $I - P_f$ and $\{U_i\}$ and $\{V_i\}$ are (respectively) orthonormal sets of vectors. Since u is in the column-space of $I - P_f$, we have $u = \sum_{i=1}^n \alpha_i U_i$ for some $\alpha_i \in \mathbb{R}$. Thus, we obtain

$$u^T(I-P_f) = \sum_{i=1}^n \alpha_i U_i^T \cdot \sum_{j=1}^n s_j U_j V_j^T = \sum_{i=1}^n \alpha_i s_i V_i^T.$$

Thus, we obtain

$$\|u^T(I-P_f)\|_1 \geq \|u^T(I-P_f)\|_2 = \sqrt{\sum_{i=1}^{n} s_i^2\alpha_i^2} \geq s_n\sqrt{\sum_{i=1}^{n}\alpha_i^2} = s_n\|u\|_2 \geq \frac{s_n}{\sqrt{|\Omega|}}\|u\|_1.$$

Here, the first and the final inequality follows from the relationship between ℓ_1 and ℓ_2 norms. Moreover, we have

$$\begin{aligned}
\left\|u^T(I-P_f)\right\|_1 &= \left\|\sum_{a\in A_+} w_a\eta_a^T(P_a-I)\right\|_1 \\
&\leq \sum_{a\in A_+} w_a\left\|\eta_a^T(P_a-I)\right\|_1 \\
&= \sum_{a\in A_+} w_a\left|\sum_{j=1}^{|\Omega|}\left(\sum_{i\neq j}\eta_a^{(i)}p_{ij}-\eta_a^{(j)}(1-p_{jj}))\right)\right| \\
&\leq \sum_{a\in A_+} w_a\sum_{j=1}^{|\Omega|}\left(\sum_{i\neq j}\eta_a^{(i)}p_{ij}+\eta_a^{(j)}(1-p_{jj}))\right) \\
&= \sum_{a\in A_+} w_a\left(\sum_{j=1}^{|\Omega|}\sum_{i\neq j}\eta_a^{(i)}p_{ij}+\sum_{j=1}^{|\Omega|}\eta_a^{(j)}(1-p_{jj})\right) \\
&= \sum_{a\in A_+} w_a\left(\sum_{i=1}^{|\Omega|}\eta_a^{(i)}\sum_{j\neq i}p_{ij}+\sum_{j=1}^{|\Omega|}\eta_a^{(j)}(1-p_{jj})\right) \\
&= \sum_{a\in A_+} w_a\left(\sum_{j=1}^{|\Omega|}\eta_a^{(j)}\sum_{i\neq j}p_{ji}+\sum_{j=1}^{|\Omega|}\eta_a^{(j)}(1-p_{jj})\right) \\
&= \sum_{a\in A_+} w_a\sum_{j=1}^{|\Omega|}\eta_a^{(j)}\left(\sum_{i\neq j}p_{ji}+1-p_{jj}\right) \\
&= 2\sum_{a\in A_+} w_a\sum_{j=1}^{|\Omega|}\eta_a^{(j)}\left(1-p_{jj}\right) \\
&\leq 2\sum_{a\in A_+} w_a\sum_{j=1}^{|\Omega|}\eta_a^{(j)} \\
&= 2,
\end{aligned}$$

where the first and the second inequality follows from triangle inequality. Taken together, we thus conclude that $\|u\|_1 \leq \frac{2\sqrt{|\Omega|}}{s_n}$.

Next, observe that since $y = u + k\nu_f$ is an optimal solution to the LP (C.8), the value of k must equal the smallest value c that makes $u + c\nu_f$ non-negative. Since $u^T\nu_f = 0$ and $\nu_f^{(i)} > 0$ for each i, there exists an i with $u^{(i)} \leq 0$, and hence $k \geq 0$. Furthermore, we obtain

$$k = \max_{i:u^{(i)}\leq 0} \frac{|u^{(i)}|}{\nu_f^{(i)}} \leq \left(\max_i \frac{1}{\nu_f^{(i)}}\right) \cdot \max_{i:u^{(i)}\leq 0} |u^{(i)}| \leq \left(\max_i \frac{1}{\nu_f^{(i)}}\right) \cdot \|u\|_1 = \tau\|u\|_1.$$

Finally, using the fact that $\|\nu_f\|_1 = 1$, we have

$$\begin{aligned} 1^T y = \|y\|_1 = \|u + k\nu_f\|_1 &\leq \|u\|_1 + k\|\nu_f\|_1 \leq \|u\|_1 + \tau\|u\|_1 \leq (1+\tau)\|u\|_1 \\ &\leq \frac{2(1+\tau)\sqrt{|\Omega|}}{s_n}. \end{aligned}$$

The result then follows from noticing that $s_n = s_{\min}$.

□