

Markov Persuasion Processes with Endogenous Agent Beliefs

Krishnamurthy Iyer¹, Haifeng Xu², and You Zu¹

¹ Industrial and Systems Engineering, University of Minnesota, {kriyer,zu000002}@umn.edu

² Department of Computer Science, University of Chicago, haifengxu@uchicago.edu

Abstract. We consider a dynamic Bayesian persuasion setting where a single long-lived sender persuades a stream of “short-lived” agents (receivers) by sharing information about a payoff-relevant state. The state transitions are Markovian conditional on the receivers’ actions, and the sender seeks to maximize the long-run average reward by committing to a (possibly history-dependent) signaling mechanism. Such problems are common in platform markets, where the platform seeks to achieve desirable long-term revenue and welfare outcomes by influencing the actions of users. While most previous studies of Markov persuasion consider exogenous agent beliefs that are independent of the chain, we study a more natural variant with *endogenous* agent beliefs that depend on the chain’s realized history. A key challenge to analyzing such settings is to model the agents’ partial knowledge about the history information. To address this challenge, we analyze a Markov persuasion process (MPP) under various information models that differ in the amount of information the receivers have about the history of the process. Specifically, we formulate a general partial-information model where each receiver observes the history with an ℓ periods lag (for $\ell \geq 0$). Our technical contribution starts with analyzing two benchmark models, i.e., the full-history information model (i.e., $\ell = 0$) and the no-history information model (i.e., $\ell = \infty$). We establish an ordering of the sender’s payoff as a function of the informativeness of agent’s information model (with no-history as the least informative), and develop efficient algorithms to compute optimal solutions for these two benchmarks. For the information model with general ℓ , we present the technical challenges in finding an optimal signaling mechanism, where even determining the right dependency on the history becomes difficult. Restricting the dependence on the history to a given length, we formulate the sender’s problem as a bilinear optimization program. To bypass the resulting computational complexity, we use a robustness framework to design a “simple” *history-independent* signaling mechanism that approximately achieves optimal payoff when ℓ is reasonably large.

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 platforms such as Uber and Lyft share real-time demand information with drivers to enable them to make repositioning decisions. In such settings, the participants’ actions not only affect their and the platform’s immediate rewards, but also impact the evolution of the system state. Given this dependence, a central question is to understand how the platform can make these recommendations taking into account the participants’ incentives as well as long-term objectives like welfare and revenue.

To study such settings, we consider a model of Markovian persuasion [Wu et al., 2022, Gan et al., 2022, Ely, 2017, Farhadi and Teneketzi, 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], 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].³ 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. 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 paper, 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 are 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 leaves open even the question of 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.

³ Few works assume that the receivers observe past signals but not past states [Ely, 2017, Farhadi and Teneketzis, 2022, Renault et al., 2017, Ashkenazi-Golan et al., 2022]; Lehrer and Shaiderman [2022] allows for stochastic revelation of past states. However, all these papers study the specialized setting where the state evolves independently of the receivers' actions.

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 [Zu et al., 2021], 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., 1995, 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.

2 Literature Survey

Our work contributes to the study of Bayesian persuasion [Kamenica and Gentzkow, 2011, Bergemann and Morris, 2016, 2019, Dughmi, 2017] in dynamic settings. Specifically, our work relates to the following streams of literature.

Markov persuasion. A number of papers have looked at persuasion problems where the state evolves according to Markov chain. We discuss a few that are close to our setting. Gan et al. [2022] study an infinite-horizon dynamic persuasion setting where the sender can observe the payoff-relevant parameter and recommends actions to the uninformed receivers to maximize the sender's cumulative rewards. They consider two types of receivers (myopic and far-sighted) and show that when the receivers are myopic, the optimal signaling strategy can be computed in polynomial time by solving a linear program. But in the setting where the receivers are far-sighted, it is NP-hard to find an approximately optimal policy. A crucial difference between our model and theirs is that in our model, the receivers' belief is endogenously determined by the sender's signaling mechanism, while in their model, the receiver's belief about the external parameter is exogenous and known to the sender.

Wu et al. [2022] focus on a finite-horizon Markov persuasion process where a single long-lived sender seeks to persuade a stream of myopic receivers to maximize the cumulative rewards. The state of the world is seen by both the sender and receiver while the uncertain outcome that affects the transition probability is only known to the sender. The authors use a reinforcement learning approach to design an online learning algorithm that achieves $O(\sqrt{T})$ regret. Similar to the work [Gan et al., 2022], the receiver's prior belief about the state is an exogenous common prior distribution. They study a finite-horizon Markov persuasion process while we consider the infinite time horizon.

Bernasconi et al. [2022] study sequential persuasion problem where the sender seeks to persuade the far-sighted receiver by sharing the payoff-relevant state. The sender can observe the realization of the state, but neither the sender nor the receiver knows the state distribution. They show that without the knowledge of the state distribution, no algorithm can be persuasive at each round with

high probability. The setting is different from ours because the transition probability is common knowledge, and the receivers are myopic and short-lived in our setting.

Also relevant to us is the recent line of work on dynamic Bayesian persuasion, where the state evolves according to a Markov chain. [Ely \[2017\]](#), [Renault et al. \[2017\]](#), [Farhadi and Teneketzis \[2022\]](#) study the dynamic Bayesian persuasion problem between two long-lived players. To reiterate, in these works, the receiver's actions do not affect the state transitions. [Ely \[2017\]](#) show that the sender's optimal strategy is myopic. Namely, the sender's optimal strategy ignores the effect of the sender's signals on the receiver's future belief. A generalization is the work by [\[Ashkenazi-Golan et al., 2022\]](#). They study the dynamic persuasion problem with binary states and any finite number of actions. The authors show that the sender's optimal strategy involves only two types of distribution of induced beliefs depending on the receiver's belief at each round. [Renault et al. \[2017\]](#), who consider a similar setting propose a greedy disclosure policy and prove that it is optimal when the initial state is sufficiently close to the invariant distribution of the Markov chain. [Farhadi and Teneketzis \[2022\]](#) study the setting in the finite time horizon and propose a time-varying optimal strategy for the sender. [Lehrer and Shaiderman \[2022\]](#) study the setting where the sender observes the state but the receivers randomly observe it. In contrast to these works, our model assumes that the receiver's actions affect the state evolution. Other papers also study dynamic persuasion for various application contexts, such as [\[Li and Norman, 2021, Wu, 2021, Board and Lu, 2018, Orlov et al., 2020, Bizzotto et al., 2021, Alizamir et al., 2020\]](#).

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. Because our proposed signaling mechanism in the partial-history information model relies on the robust persuasion framework, our work also relates to robust persuasion. [Zu et al. \[2021\]](#) 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 and achieves $O(\sqrt{T} \log T)$ regret. For our robustness results, we extend their approach to settings where the receivers' beliefs are endogenous.

[Kosterina \[2022\]](#) study a persuasion setting without 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. The sender evaluates the expected utility under each signaling mechanism concerning the worst-case prior of the receiver. [Ui \[2022\]](#) study the optimal robust public information sharing where the sender discloses public information with receivers who also acquire costly private information. The sender is uncertain about the precision and the cost of the private information. Similarly, [Hu and Weng \[2021\]](#) study the problem where the receiver may have exogenous private information unknown to the sender. The sender seeks to maximize her expected payoff under the worst-case payoff across the receiver's possible private belief distributions and then, among them, chooses the one that maximizes the expected utility under her conjectured prior. [Dworcak and Pavan \[2022\]](#) share a similar angle. Both works focus on static persuasion models with robustness to exogenous receiver beliefs whereas our model focuses on robust persuasion with the 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.

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, $p: \Omega \times \Omega \times A \rightarrow [0, 1]$ 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 $v(\bar{\omega}_t, \bar{a}_t) \in [0, 1]$.

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, 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: $\bar{h}_t^k = (\bar{x}_{t-k}, \dots, \bar{x}_{t-1}) \in \mathcal{X}^k$. We denote a generic element of \mathcal{X}^k by $h^k = (x_{-k}, \dots, x_{-1})$. 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 $\sigma: \mathcal{X}^k \times \Omega \rightarrow \Delta(A)$ (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.

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 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 $\partial u(\omega, a, a') := u(\omega, a) - u(\omega, a')$ denotes the incremental payoff for the receiver for choosing 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 $\text{Pers}(\Phi)$ 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 $\text{Pers}(\Phi)$ is non-empty, since the mechanism that recommends the receivers' preferred action at each state is persuasive for sequence Φ .

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 $\sigma \in \text{Pers}(\Phi) \cap \Sigma_k$ for some $k \geq 0$ and

that the receivers follow the sender's recommendations. For $k \geq 1$, the induced process dynamics can be described as a Markov chain with states given by slices $\bar{h}_t^k \in \mathcal{X}^k$. An invariant distribution $\pi \in \Delta(\mathcal{X}^k)$ of this chain satisfies the following balance equations:

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

for each $h^{k-1} = (x_{-(k-1)}, \dots, x_{-1}) \in \mathcal{X}^{k-1}$ and $(\omega, a) \in \mathcal{X}$. Here, the left-hand side expression gives the probability that the slice \bar{h}_{t+1}^k equals (h^{k-1}, ω, a) after a Markovian transition if the slice $\bar{h}_t^k = (x_{-k}, h^{k-1})$ is distributed as π and the receiver at time t follows the sender's recommendation. The equality then just states the fact that for an invariant distribution, this distribution must be π itself. For $k = 0$, the induced process dynamics can be described as Markov chain with states $(\bar{\omega}_{t-1}, \bar{a}_{t-1}) = \bar{h}_t^1 \in \mathcal{X}$, and the balance equation for an invariant distribution $\pi \in \Delta(\mathcal{X})$ given by

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

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 sends signal $\bar{s}_t = 0$ if $(\bar{\omega}_{t-1}, \bar{a}_{t-1}, \bar{\omega}_t) = (0, 0, 0)$ and sends signal $\bar{s}_t = 1$ if $(\bar{\omega}_{t-1}, \bar{a}_{t-1}, \bar{\omega}_t) = (0, 1, 0)$. Then, any distribution over \mathcal{X} 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 expected reward is maximized (with ties broken arbitrarily). We denote such a distribution by $\text{Inv}(\sigma)$. 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 $\pi = \text{Inv}(\sigma)$ also denote the distribution of the 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 . Furthermore, for any $\ell \geq 1$, we let $\pi(h^\ell)$ denote the (marginal) distribution of a slice of history \bar{h}_t^ℓ .

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, the 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 $t \in \mathbb{Z}$,

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

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}^σ 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 approach, motivated by the requirement of consistency,⁴ is to let each belief ϕ_t equal the invariant distribution $\text{Inv}(\sigma)$. Specifically, we have for each $t \in \mathbb{Z}$,

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

Here, $\text{Inv}(\sigma) \otimes P$ encodes the distribution of $(\bar{h}_t, \bar{\omega}_t)$ where the history \bar{h}_t is distributed as $\text{Inv}(\sigma)$, 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_t would have a complex dependence on the history \bar{h}_t . Although a comprehensive analysis of all such models is beyond the scope of this paper, we focus on a particular sequence of information models to capture realistic scenarios where the receivers may have some stale information about the process.⁵

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 $t \in \mathbb{Z}$,

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

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

⁴ 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 w.r.t. the invariant distribution when the latter is unique.

⁵ Such stale information about the process could plausibly arise from the receivers having interacted with the process in the past; however, we do not consider such repeated interactions in our model.

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

$$\begin{aligned} \text{MPP}(\Phi) := \max_{\sigma, \pi} \quad & \mathbf{E}^\pi[v(\omega, a)] \\ \text{subject to,} \quad & \sigma \in \text{Pers}(\Phi) \cap \Sigma, \quad \pi = \text{Inv}(\sigma), \end{aligned} \quad (2)$$

and let $\text{OPT}(\Phi)$ denote its optimal value. Furthermore, for $k \geq 0$, we analogously define $\text{MPP}(\Phi, \Sigma_k)$ (and $\text{OPT}(\Phi, \Sigma_k)$) 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 1 (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 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 (2) under the benchmark full-history and no-history information models. As we show later, the sender's optimal payoff 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.

4.1 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 A (as are the proofs of all results in this section). Most importantly, the lemma allows us to show that the sender's problem (2) in the benchmark settings can be formulated as a polynomially-sized linear program. This LP formulation plays a key role in particular in Proposition 2, 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 begin by recalling that under the information model Φ_{no} and with a signaling mechanism $\sigma \in \Sigma_0$, the states $\bar{\omega}_t \in \Omega$ form a Markov chain. On the other hand, for a signaling mechanism $\sigma \in \Sigma_1$ under the model Φ_{full} , the induced Markov chain can be described with states $(\bar{x}_{t-1}, \bar{\omega}_t) \in \mathcal{X} \times \Omega$. We now introduce some notation to unify the presentation. First, define $\mathcal{X}_{\text{full}} := \mathcal{X}$ and $\mathcal{X}_{\text{no}} = \{\star\}$. For $i \in \{\text{full}, \text{no}\}$, define the corresponding state space $\mathcal{W}_i := \mathcal{X}_i \times \Omega$. Let $v(w, a) = v(\omega, a)$ for $w = (x, \omega) \in \mathcal{W}_i$, and extend $u(w, a)$ and $\partial u(w, a, a')$ for $w \in \mathcal{W}_i$ similarly. Next, for $w, \hat{w} \in \mathcal{W}_i$, with $w = (x_{-1}, \omega)$ and $\hat{w} = (\hat{x}_{-1}, \hat{\omega})$, we define the transition kernel $p(w|\hat{w}, \hat{a}) := p(\omega|\hat{\omega}, \hat{a})\mathbf{I}\{x_{-1} = (\hat{\omega}, \hat{a}) \text{ or } x_{-1} = \star\}$. Finally, for $x \in \mathcal{X}_i$ and $w = (x_{-1}, \omega) \in \mathcal{W}_i$, define $D(x, w) := \mathbf{I}\{x = x_{-1}\}$.

With these notation in place, for $i \in \{\text{full}, \text{no}\}$, we consider the following linear program $\text{LP}(i)$ with variables $z(w, a)$ with $w \in \mathcal{W}_i$ and $a \in A$.

$$\begin{aligned} \text{LP}(i) := \max_{z \geq 0} \quad & \sum_{w \in \mathcal{W}_i} \sum_{a \in A} z(w, a) v(w, a) \\ & \sum_{w \in \mathcal{W}_i} z(w, a) D(x, w) \partial u(w, a, a') \geq 0, \quad \text{for all } a, a' \in A \text{ and } x \in \mathcal{X}_i \\ & \sum_{\hat{w} \in \mathcal{W}_i} \sum_{\hat{a} \in A} z(\hat{w}, \hat{a}) p(w|\hat{w}, \hat{a}) = \sum_a z(w, a), \quad \text{for all } w \in \mathcal{W}_i \\ & \sum_{w \in \mathcal{W}_i} \sum_a z(w, a) = 1. \end{aligned} \tag{3}$$

To interpret the linear program $\text{LP}(i)$, we focus on the full-history information model Φ_{full} , and consider a persuasive signaling mechanism $\sigma \in \Sigma_1$. Then, writing the balance equations (1) for the invariant distribution $\pi \in \text{Inv}(\sigma)$, we obtain

$$\sum_{\omega_{-1}, a_{-1}} \pi(\omega_{-1}, a_{-1}) p(\omega|\omega_{-1}, a_{-1}) \sigma(a|\omega_{-1}, a_{-1}, \omega) = \pi(\omega, a).$$

Since this equation is bilinear in π and σ , we introduce the variables $z(w, a)$ for $w = (\omega_{-1}, a_{-1}, \omega)$ to denote the summands on the left-hand side of the preceding equation, and note that z constitutes the joint distribution of two consecutive state-action pairs $(\bar{x}_{t-1}, \bar{x}_t)$ under the signaling mechanism σ . These variables are readily seen to satisfy the two equalities in $\text{LP}(\text{full})$. Assuming σ is persuasive under Φ_{full} then yields the inequality. Thus, the variables $z(w, a)$ defined above are feasible for $\text{LP}(\text{full})$. We obtain the following proposition upon showing the converse, i.e., for any z feasible for $\text{LP}(\text{full})$, there exists a signaling mechanism $\sigma \in \Sigma_1$ persuasive under the full-history information model Φ_{full} satisfying the equation $z(w, a) = \pi(\omega_{-1}, a_{-1}) p(\omega|\omega_{-1}, a_{-1}) \sigma(a|\omega_{-1}, a_{-1}, \omega)$ (and a similar statement for the Φ_{no} case).

Proposition 1. *For $i \in \{\text{full}, \text{no}\}$, the sender's problem $\text{MPP}(\Phi_i)$ can be equivalently formulated as the corresponding linear program $\text{LP}(i)$. In particular, for any optimal solution z_i^* of $\text{LP}(i)$ the signaling mechanism σ_i , defined as,*

$$\sigma_i(a|w) := \frac{z_i^*(w, a)}{\sum_{a'} z_i^*(w, a')}, \quad \text{for all } w \in \mathcal{W}_i \text{ and } a \in A,$$

(if the denominator is positive; otherwise, recommending the receivers' preferred action), is optimal for the problem $\text{MPP}(\Phi_i)$.

Note that the linear program $\text{LP}(i)$ has $\mathcal{O}(|\mathcal{W}_i| \cdot |A|)$ variables and $\mathcal{O}(|A|^2 \cdot |\mathcal{X}_i| + |\mathcal{W}_i|)$ constraints. Thus, for the no-history information model Φ_{no} , the linear program $\text{LP}(\text{no})$ 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 $\text{LP}(\text{full})$ has $O(|\Omega|^2 \cdot |A|^2)$ variables and $O(|\Omega|^2 \cdot |A|^3)$ constraints. Together, the preceding result establishes that, not only the sender's problem in the benchmark models has a simple LP formulation, but also that they can be solved efficiently. As we discuss later in Section 5.1, this is in stark contrast to the case under partial-history information models.

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 $\text{Pers}(\Phi_{\text{full}}) \subseteq \text{Pers}(\Phi_\ell) \subseteq \text{Pers}(\Phi_{\ell+1}) \subseteq \text{Pers}(\Phi_{\text{no}})$ and consequently, $\text{OPT}(\Phi_{\text{full}}) \leq \text{OPT}(\Phi_\ell) \leq \text{OPT}(\Phi_{\ell+1}) \leq \text{OPT}(\Phi_{\text{no}})$.*

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 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 0.8. By solving the LP formulations in the preceding section, we find that the sender's optimal payoff in the no-history information model equals $\text{OPT}(\Phi_{\text{no}}) = 1$, 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 $\text{OPT}(\Phi_{\text{full}}) = 0.52$.

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. Let $\sigma \in \Sigma_0$ denote an optimal signaling mechanism in the no-history information model Φ_{no} . Note that under Φ_{no} , the receivers' prior beliefs equal the invariant distribution $\pi = \text{Inv}(\sigma)$. For any action $a \in A$ recommended by the optimal signaling mechanism, let μ_a denote the resulting posterior belief of such a receiver, and let \mathcal{B}_{no} denote the set of all posterior beliefs so induced. Finally, let $\text{Conv}(\mathcal{B}_{\text{no}})$ denote the convex hull of \mathcal{B}_{no} .

Proposition 2. Suppose the set of posterior beliefs \mathcal{B}_{no} induced by the optimal signaling mechanism in the no-history information model Φ_{no} is linearly independent. Furthermore, suppose the transition kernels lie in the convex hull of these beliefs, i.e., we have $p(\cdot|x) \in \text{Conv}(\mathcal{B}_{\text{no}})$ for all $x \in \mathcal{X}$. Then, we have $\text{OPT}(\Phi_{\text{no}}) = \text{OPT}(\Phi_{\text{full}})$.

The proposition, together with Lemma 2, implies that when the conditions in the proposition 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 $\text{OPT}(\Phi_\ell) = \text{OPT}(\Phi_{\text{no}})$ for all $\ell \geq 0$. The proof of the proposition starts by writing the transition probabilities $p(\omega|x)$ as a convex combination $\sum_{a \in A} \lambda(a|x) \mu_a(\omega)$ of the beliefs in \mathcal{B}_{no} . Then, we use these weights $\lambda(a|x)$ to explicitly construct a signaling mechanism $\hat{\sigma} \in \Sigma_1$ which is persuasive for the model Φ_{full} , and induces the same set \mathcal{B}_{no} of posterior beliefs for the receivers with the same distribution, resulting in the same payoff for the sender. Finally, observe that the beliefs $\mu_a \in \mathcal{B}_{\text{no}}$ can be easily computed from the optimal solution of the linear program $\text{LP}(\text{no})$; thus, the sufficient conditions in the proposition statement are straightforward to verify.

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 $\text{MPP}(\Phi_\ell)$, and the associated computational challenges for the problem $\text{MPP}(\Phi_\ell, \Sigma_k)$, which we formulate as a bilinear optimization program. Given these challenges, we design an approximately optimal signaling mechanism for $\text{MPP}(\Phi_\ell)$ 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 .

5.1 Intricacies of Persuasion in Partial-history Information Models

Consider the sender’s persuasion problem $\text{MPP}(\Phi_\ell)$ 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 $\Phi_{\text{full}} = \Phi_0$, 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

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

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 $\text{MPP}(\Phi_1, \Sigma_1)$ as a bilinear program.

Proposition 3. *The sender's problem $\text{MPP}(\Phi_1, \Sigma_1)$ can be formulated as the following bilinear program:*

$$\begin{aligned} & \max_{z \geq 0} \sum_{x_{-2}, x_{-1} \in \mathcal{X}} \sum_{\omega, a} z(x_{-2}, x_{-1}, \omega, a) v(\omega, a) \\ & \sum_{x_{-1} \in \mathcal{X}} \sum_{\omega} z(x_{-2}, x_{-1}, \omega, a) \partial u(\omega, a, a') \geq 0, \quad \text{for all } x_{-2} \in \mathcal{X} \text{ and } a, a' \in A. \\ & \sum_{x_{-3} \in \mathcal{X}} z(x_{-3}, x_{-2}, x_{-1}) p(\omega | x_{-1}) = \sum_{a \in A} z(x_{-2}, x_{-1}, \omega, a), \quad \text{for all } x_{-2}, x_{-1} \in \mathcal{X} \text{ and } \omega \in \Omega \\ & \sum_{x_{-2}, x_{-1}, \omega, a} z(x_{-2}, x_{-1}, \omega, a) = 1 \\ & z(x_{-2}, x_{-1}, \omega, a) \cdot \sum_{a' \in A} z(x'_{-2}, x_{-1}, \omega, a') = z(x'_{-2}, x_{-1}, \omega, a) \cdot \sum_{a' \in A} z(x_{-2}, x_{-1}, \omega, a'), \\ & \quad \text{for all } x'_{-2}, x_{-2}, x_{-1}, (\omega, a) \in \mathcal{X}. \end{aligned} \tag{4}$$

To elaborate, for any $\sigma \in \Sigma_1$, assuming the receivers follow the recommendation and given our preceding discussion, the underlying process dynamics can be described as a Markov chain with states given by slices $\bar{h}_t^2 = (\bar{x}_{t-2}, \bar{x}_{t-1})$. The balance equation for this chain's invariant distribution $\pi \in \text{Inv}(\sigma)$ is given by

$$\sum_{x_{-2} \in \mathcal{X}} \pi(x_{-2}, x_{-1}) p(\omega | x_{-1}) \sigma(a | x_{-1}, \omega) = \pi(x_{-1}, \omega, a),$$

for all $x_{-1} \in \mathcal{X}$ and $(\omega, a) \in \mathcal{X}$. As in the case for Φ_{full} and Φ_{no} , this equation is non-linear in π and σ . However, introducing the variables $z(x_{-2}, x_{-1}, \omega, a) := \pi(x_{-2}, x_{-1}) p(\omega | x_{-1}) \sigma(a | x_{-1}, \omega)$ no longer yields a linear program, because of the restriction that σ cannot depend on x_{-2} . In particular, the non-linear equality constraint in (4) explicitly encodes the requirement that $z(x_{-2}, x_{-1}, \omega, a) / \sum_{a' \in A} z(x_{-2}, x_{-1}, \omega, a')$ is independent of x_{-2} .

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 $(\bar{x}_{t-\ell-k}, \dots, \bar{x}_{t-\ell-1})$, but the signaling mechanism σ does not base its recommendation

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

Fig. 1. (Example 1 contd.) Sender’s optimal payoff in $\text{MPP}(\Phi_1, \Sigma_k)$ for different values of k . The optimal values are obtained by numerically solving bilinear optimization programs analogous to (4) for different values of k . Here, $\text{OPT}(\Phi_{\text{full}}) = 0.52$ and $\text{OPT}(\Phi_{\text{no}}) = 1$.

on $(\bar{x}_{t-\ell-k}, \dots, \bar{x}_{t-k-1})$. Again, the sender’s problem $\text{MPP}(\Phi_\ell, \Sigma_k)$ 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.

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) 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 1.) In the following section, we provide an approach to overcome this trade-off, as long as one is satisfied with approximate optimality.

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 an approach inspired from robust persuasion [Zu et al., 2021]. 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 $\pi = \text{Inv}(\sigma)$. 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 $\pi = \text{Inv}(\sigma)$; 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 $\mathbf{B}_1(\pi, \epsilon)$ denote the set of all distributions $\mu \in \Delta(\Omega)$ that are ϵ -close to π in ℓ_1 -norm: $\mathbf{B}_1(\pi, \epsilon) := \{\mu \in \Delta(\Omega) : \|\mu - \pi\|_1 \leq \epsilon\}$.

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 $\mathbf{B}_1(\pi, \epsilon)$:

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

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

Our next result describes the relation between $\text{RP}(\epsilon)$ and the set $\text{Pers}(\Phi_\ell)$ for large ℓ . For $\sigma \in \Sigma_0$ and $\ell \geq 1$, let $Q_\sigma^\ell(x, \omega) := \mathbf{P}^\sigma(\bar{\omega}_\ell = \omega | \bar{x}_{-1} = x)$ 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 $\pi = \text{Inv}(\sigma)$ over $x \in \mathcal{X}$:

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

Finally, let $\gamma_*(\sigma)$ 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 $\sigma \in \text{Pers}(\Phi_\ell)$. In particular, $\sigma \in \text{Pers}(\Phi_\ell)$ for all $\ell \geq \frac{1}{\gamma_*(\sigma)} \log \left(\frac{2}{\epsilon \pi_{\min}(\sigma)} \right)$.*

The proof of the bound in the lemma statement uses the unichain assumption (Assumption 1) to bound the mixing time of the underlying Markov chain. The result implies that in order to find a signaling mechanism in $\text{Pers}(\Phi_\ell)$, it suffices to find a history-independent signaling mechanism in the set $\text{RP}(\epsilon)$ 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:

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

In other words, \mathcal{P}_a is the set of beliefs for which the receiver finds it optimal to choose action a . Similar to Zu et al. [2021], we make the following regularity assumption on the receivers' utility function.

Assumption 2 (Regularity). *There exists a positive constant $D > 0$ and beliefs $\eta_a \in \mathcal{P}_a$ for $a \in A$ such that $B_1(\eta_a, D) \subseteq \mathcal{P}_a$ for each $a \in A$, where $B_1(\eta, D)$ is an ℓ_1 -ball of size D 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, Zu et al. [2021] establish 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., $a_\omega \in \operatorname{argmax}_{a \in A} u(\omega, a)$ for each $\omega \in A$. Let $P_f(\omega, \omega') := p(\omega' | \omega, a_\omega)$ 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 $\tau := \max_{\omega} 1/\nu_f(\omega)$ 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 3. *For $\epsilon < \frac{s_f w_{\min} D}{2(s_f + 2(1+\tau)\sqrt{|\Omega|})}$, there exists a signaling mechanism $\hat{\sigma} \in \text{RP}(\epsilon)$ with the sender's payoff bounded below by*

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

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 $\hat{\sigma} \in \text{RP}(\epsilon)$. From a computational perspective, constructing such a mechanism requires solving $\text{MPP}(\Phi_{\text{no}})$ (equivalently the linear program $\text{LP}(\text{no})$), and solving a separate linear program (13) with $\mathcal{O}(|\Omega|)$ variables and constraints (see Lemma 5 in Appendix B for details). Thus, not only the proposed mechanism obtains approximately optimal payoffs, but it also can be computed efficiently.

To construct the mechanism $\hat{\sigma} \in \text{RP}(\epsilon)$ with good guarantees on the sender's payoff, we use a similar approach as in [Zu et al., 2021], 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 of [Zu et al., 2021], 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., 1995, 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 $\{w_s \geq 0 : s \in S\}$ such that $\sum_{s \in S} w_s = 1$ and*

$$\sum_{s \in S} \sum_{\omega} w_s \mu_s(\omega) p(\cdot | \omega, a_s) = \sum_{s \in S} w_s \mu_s.$$

Then, there exists a signaling mechanism $\sigma \in \Sigma_0$, which sends signals $s \in S$ with probability $\sigma(s | \omega) = \frac{w_s \mu_s(\omega)}{\sum_{s' \in S} w_{s'} \mu_{s'}(\omega)}$, such that under the no-history information model Φ_{no} , the posterior belief of a receiver on receiving signal s equals μ_s .

Conversely, for any signaling mechanism $\sigma \in \text{Pers}(\Phi_{\text{no}}) \cap \Sigma_0$, there exists weights $w_a \geq 0$ and beliefs $\mu_a \in \mathcal{P}_a$ with $\sum_{a \in A} w_a = 1$ and $\sum_{\omega, a} w_a \mu_a(\omega) p(\cdot | \omega, a) = \sum_a w_a \mu_a$, such that $\sigma(a | \omega) = \frac{w_a \mu_a(\omega)}{\sum_{a'} w_{a'} \mu_{a'}(\omega)}$.

With this 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 B.

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 highlights 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.

References

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

Jerry Anunrojwong, Krishnamurthy Iyer, and Vahideh Manshadi. Information design for congested social services: Optimal need-based persuasion. *Management Science*, 2022.

Galit Ashkenazi-Golan, Penélope Hernández, Zvika Neeman, and Eilon Solan. Markovian persuasion with two states. *arXiv preprint arXiv:2209.06536*, 2022.

Robert J Aumann, Michael Maschler, and Richard E Stearns. *Repeated games with incomplete information*. MIT press, 1995.

Yakov Babichenko, Inbal Talgam-Cohen, Haifeng Xu, and Konstantin Zabarnyi. Regret-minimizing bayesian persuasion. *Games and Economic Behavior*, 136:226–248, 2022.

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

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

Martino Bernasconi, Matteo Castiglioni, Alberto Marchesi, Nicola Gatti, Francesco 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Jérôme Renault, Eilon Solan, and Nicolas Vieille. Optimal dynamic information provision. *Games and Economic Behavior*, 104:329–349, 2017.

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

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

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

Jibang Wu, Zixuan Zhang, Zhe Feng, Zhaoran Wang, Zhuoran Yang, Michael I Jordan, and Haifeng 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.

Wenhao Wu. Sequential bayesian persuasion. 2021.

You Zu, Krishnamurthy Iyer, and Haifeng Xu. Learning to persuade on the fly: Robustness against ignorance. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, EC '21, page 927–928, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450385541. <https://doi.org/10.1145/3465456.3467593>. URL <https://doi.org/10.1145/3465456.3467593>.

A Proofs from Section 4

A.1 Proofs from Section 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 $\sigma \in \Sigma_k \cap \text{Pers}(\Phi_{\text{no}})$ for some k , there exists a $\widehat{\sigma} \in \Sigma_0 \cap \text{Pers}(\Phi_{\text{no}})$ with the same payoff for the sender.

First, recall that in the no-history information model Φ_{no} and under the signaling mechanism $\sigma \in \Sigma_k \in \text{Pers}(\Phi_{\text{no}})$, the receiver's prior belief is given by the invariant distribution $\pi = \text{Inv}(\sigma)$. Note that π describes the invariant distribution of the slice $\bar{h}_t^k \in \mathcal{X}^k$ under σ . By abusing the notation, we let $\pi(\omega, a) := \sum_{h^{k-1}} \pi(h^{k-1}, \omega, a)$ also denote the marginal distribution of $(\bar{\omega}_t, \bar{a}_t)$ under π .

Now, define the signaling mechanism $\widehat{\sigma} \in \Sigma_0$ as follows: for $\omega \in \Omega$ with $\sum_a \pi(\omega, a) > 0$, let

$$\widehat{\sigma}(a|\omega) := \frac{\pi(\omega, a)}{\sum_{a'} \pi(\omega, a')} \quad \text{for } a \in A$$

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

We first show that $\widehat{\pi} \in \Delta(\mathcal{X})$ with $\widehat{\pi}(\omega, a) := \pi(\omega, a)$ is an invariant distribution under $\widehat{\sigma}$. To see this, observe that if $\sum_{a \in A} \pi(\omega, a) > 0$, we have

$$\begin{aligned} \sum_{x_{-1} \in \mathcal{X}} \widehat{\pi}(x_{-1}) p(\omega|x_{-1}) \widehat{\sigma}(a|\omega) &= \left(\sum_{x_{-1} \in \mathcal{X}} \pi(x_{-1}) p(\omega|x_{-1}) \right) \widehat{\sigma}(a|\omega) \\ &= \left(\sum_{a'} \pi(\omega, a') \right) \widehat{\sigma}(a|\omega) \\ &= \pi(\omega, a) \\ &= \widehat{\pi}(\omega, a). \end{aligned}$$

Here, the first and fourth equality follows from the definition of $\widehat{\pi}$, the second follows from the fact that $\pi = \text{Inv}(\sigma)$, and the third equality follows from the definition of $\widehat{\sigma}$. Moreover, if $\sum_{a \in A} \pi(\omega, a) = 0$, then $\sum_{x_{-1} \in \mathcal{X}} \pi(x_{-1}) p(\omega|x_{-1}) = 0$, and $\widehat{\pi}(\omega, a) = \pi(\omega, a) = 0$, and hence the equality continues to hold. Thus, $\widehat{\pi}$ satisfies the balance equations (1) under $\widehat{\sigma}$, and thus, $\widehat{\pi} = \text{Inv}(\widehat{\sigma})$.

Finally, in the information model Φ_{no} and under the mechanism σ_k , a receiver's posterior belief that $\bar{\omega}_t = \omega$ after being recommended $\bar{s}_t = a$ is given by

$$F(\omega|a; \phi_t, \sigma) = \frac{\sum_{h^k} \pi(h^k) p(\omega|x_{-1}) \sigma(a|h^k, \omega)}{\sum_{h^k} \sum_{\omega'} \pi(h^k) p(\omega'|x_{-1}) \sigma(a|h^k, \omega')} \quad (5)$$

where $\pi(h^k)$ denotes the probability that the slice $\bar{h}_t^k = h^k \in \mathcal{X}^k$ under the invariant distribution $\pi = \text{Inv}(\sigma)$. Summing the balance equation (1) for π over all values of $h^{k-1} \in \mathcal{X}^{k-1}$, we obtain

$$\sum_{h^k} \pi(h^k) p(\omega|x_{-1}) \sigma(a|h^k, \omega) = \sum_{h^{k-1}} \pi(h^{k-1}, \omega, a) = \pi(\omega, a).$$

Substituting in (5), we obtain $F(\omega|a; \phi_t, \sigma) = \frac{\pi(\omega, a)}{\sum_{\omega'} \pi(\omega', a)} = \frac{\widehat{\pi}(\omega, a)}{\sum_{\omega'} \widehat{\pi}(\omega', a)} = F(\omega|a; \phi_t, \widehat{\sigma})$, where the last equality follows by a similar argument for $\widehat{\sigma}$ in the model Φ_{no} . Since the receivers have the same belief under σ and $\widehat{\sigma}$, and further $\sigma \in \text{Pers}(\Phi_{\text{no}})$, we conclude that $\widehat{\sigma} \in \text{Pers}(\Phi_{\text{no}})$. The result then follows from the fact that the sender's payoffs under the two mechanisms are equal.

2. Full-history information model Φ_{full} : The proof for the full-history information model follows along similar lines. Fix $\sigma \in \text{Pers}(\Phi_{\text{full}}) \cap \Sigma_k$, and let $\pi = \text{Inv}(\sigma)$ denote its invariant distribution. As before, we define a mechanism $\widehat{\sigma} \in \Sigma_1$ and show that it is persuasive under Φ_{full} and achieves the same payoff for the sender. Towards that end, let $\pi(x_{-1}, \omega, a)$ denote the marginal distribution under π that $(\bar{x}_{t-1}, \bar{\omega}_t, \bar{a}_t) = (x_{-1}, \omega, a)$ and define for $x_{-1} \in \mathcal{X}$ and $\omega \in \Omega$,

$$\widehat{\sigma}(a|x_{-1}, \omega) := \frac{\pi(x_{-1}, \omega, a)}{\sum_{a'} \pi(x_{-1}, \omega, a')},$$

if the denominator is positive, and otherwise let $\widehat{\sigma}$ recommend the receiver-optimal action at ω . Similarly, define $\widehat{\pi} \in \Delta(\mathcal{X})$ to be $\widehat{\pi}(\omega, a) := \sum_{x_{-1}} \pi(x_{-1}, \omega, a)$. Whenever $\sum_{a'} \pi(x_{-1}, \omega, a) > 0$, we have

$$\begin{aligned} \sum_{x_{-1} \in \mathcal{X}} \widehat{\pi}(x_{-1}) p(\omega|x_{-1}) \widehat{\sigma}(a|x_{-1}, \omega) &= \sum_{x_{-1} \in \mathcal{X}} \left(\sum_{x_{-2} \in \mathcal{X}} \pi(x_{-2}, x_{-1}) p(\omega|x_{-1}) \right) \widehat{\sigma}(a|x_{-1}, \omega) \\ &= \sum_{x_{-1} \in \mathcal{X}} \left(\sum_{a'} \pi(x_{-1}, \omega, a') \right) \widehat{\sigma}(a|x_{-1}, \omega) \\ &= \sum_{x_{-1} \in \mathcal{X}} \pi(x_{-1}, \omega, a) \\ &= \widehat{\pi}(\omega, a). \end{aligned}$$

On the other hand, if $\sum_{a'} \pi(x_{-1}, \omega, a) = 0$, we obtain both sides of the equations are zero. Thus, we conclude that $\widehat{\pi} \in \text{Inv}(\widehat{\sigma})$.

Finally, in the information model Φ_{full} and under the signaling mechanism σ , we have

$$F(\omega|a; \phi_t, \widehat{\sigma}) = \frac{p(\omega|x_{-1}) \sigma(a|h^k, \omega)}{\sum_{\omega'} p(\omega'|x_{-1}) \sigma(a|h^k, \omega')}.$$

As $\sigma \in \text{Pers}(\Phi_{\text{full}})$, we obtain for any $a, a' \in A$ and all $h^k \in \mathcal{X}^k$,

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

After multiplying by $\pi(h^k)$, summing up over (x_{-k}, \dots, x_{-2}) , and using the fact that $\pi(x_{-1}, \omega, a) = \sum_{(x_{-k}, \dots, x_{-2})} \pi(h^k) p(\omega|x_{-1}) \sigma(a|h^k, \omega)$ from the balance equations for $\pi = \text{Inv}(\sigma)$, we obtain for all $x_{-1} \in \mathcal{X}$ and $a, a' \in A$,

$$\sum_{\omega} \pi(x_{-1}, \omega, a) \partial u(\omega, a, a') \geq 0.$$

Now, from the definition of $\widehat{\sigma}$, we have $\pi(x_{-1}, \omega, a) = \widehat{\pi}(x_{-1}) p(\omega|x_{-1}) \widehat{\sigma}(a|x_{-1}, a)$ if $\sum_{a'} \pi(x_{-1}, \omega, a') > 0$. Furthermore, under this condition and using the fact that $\pi \in \text{Inv}(\sigma)$, we have $\widehat{\pi}(x_{-1}) = \sum_{x_{-2}} \pi(x_{-2}, x_{-1}) = \sum_{x_{-2}} \pi(x_{-1}, x_{-2}) > 0$. Thus, we conclude that

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

On the other hand, if $\sum_{a'} \pi(x_{-1}, \omega, a') > 0$, then $\widehat{\sigma}$ recommends the receiver-optimal action. Thus, we conclude that $\widehat{\sigma} \in \text{Pers}(\Phi_{\text{full}})$. Once again, the result then follows as the sender's payoffs under the two mechanisms are equal. \square

Proof of Proposition 1. We prove the statement for $\text{MPP}(\Phi_{\text{full}})$. A similar argument, with minor modifications, obtains the equivalence of $\text{MPP}(\Phi_{\text{no}})$ and $\text{LP}(\text{no})$; we omit it for brevity.

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

To begin, fix $\sigma \in \text{Pers}(\Phi_{\text{full}}) \cap \Sigma_1$, and let $\pi = \text{Inv}(\sigma) \in \Delta(\mathcal{X})$. The balance equations (1) are given by

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

Define $z(w, a) := \pi(x_{-1}) p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega)$ for $w = (x_{-1}, \omega) \in \mathcal{W}_{\text{full}}$ and $a \in A$. It is straightforward to check that z satisfies the second equality in $\text{LP}(\text{full})$. For $w = (\omega_{-1}, a_{-1}, \omega) \in \mathcal{W}_{\text{full}}$, we obtain

$$\begin{aligned} \sum_a z(w, a) &= \pi(x_{-1}) p(\omega|x_{-1}) \\ &= \sum_{x_{-2}} \pi(x_{-2}) p(\omega_{-1}|x_{-2}) \sigma(a_{-1}|x_{-2}, \omega_{-1}) p(\omega|x_{-1}) \\ &= \sum_{x_{-2}} z(x_{-2}, \omega_{-1}, a_{-1}) p(\omega|x_{-1}) \\ &= \sum_{x_{-2}} \sum_{\hat{w}, \hat{a}} z(\hat{w}, \hat{a}) p(\omega|x_{-1}) \mathbf{I}\{\hat{w} = (x_{-2}, \omega_{-1}), \hat{a} = a_{-1}\} \\ &= \sum_{\hat{w}, \hat{a}} z(\hat{w}, \hat{a}) \sum_{x_{-2}} p(\omega|x_{-1}) \mathbf{I}\{\hat{w} = (x_{-2}, \omega_{-1}), \hat{a} = a_{-1}\} \\ &= \sum_{\hat{w}, \hat{a}} z(\hat{w}, \hat{a}) p(w|\hat{w}, \hat{a}). \end{aligned}$$

Here, the second equality follows from (1), the second follows from the definition of z , and the final equality follows from the definition of $p(w|\hat{w}, \hat{a})$. Thus, we conclude that z satisfies both the equalities in $\text{LP}(\text{full})$. Finally, since $\sigma \in \text{Pers}(\Phi_{\text{full}}) \cap \Sigma_1$, we obtain for all $x_1 \in \mathcal{X}$ and $a, a' \in A$,

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

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

$$\begin{aligned} \sum_{w \in \mathcal{W}_{\text{full}}} z(w, a) D(x_{-1}, w) \partial u(w, a, a') &= \sum_{x \in \mathcal{X}} \sum_{\omega \in \Omega} z(x, \omega, a) \mathbf{I}\{x = x_{-1}\} \partial u(\omega, a, a') \\ &= \sum_{\omega} z(x_{-1}, \omega, a) \partial u(\omega, a, a') \\ &= \pi(x_{-1}) \left(\sum_{\omega} p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a') \right) \geq 0. \end{aligned}$$

Thus, we obtain that z satisfies the inequality in $\text{LP}(\text{full})$. Finally, since $\sum_{x \in \mathcal{X}} z(x, \omega, a) = \pi(\omega, a)$, we conclude that the sender's payoff under σ equals the $\text{LP}(\text{full})$ objective at z . This concludes the first part of the statement.

Conversely, suppose z is a feasible solution for $\text{LP}(\text{full})$. Define the signaling mechanism $\sigma \in \Sigma_1$ as follows: for all $w = (x_{-1}, \omega) \in \mathcal{W}_{\text{full}}$ with $\sum_{a'} z(w, a') > 0$, let

$$\sigma(a|x_{-1}, \omega) := \frac{z(w, a)}{\sum_{a'} z(w, a')}.$$

For $w = (x_{-1}, \omega) \in \mathcal{W}_{\text{full}}$ with $\sum_{a'} z(w, a') = 0$, let σ recommend the receiver-optimal action at ω . We note that $\pi \in \Delta(\mathcal{X})$ defined as $\pi(\omega, a) := \sum_{x \in \mathcal{X}} z(x, \omega, a)$ for $(\omega, a) \in \mathcal{X}$ is invariant under σ . To see this, observe

$$\begin{aligned} \sum_{x_{-1}} \pi(x_{-1}) p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega) &= \sum_{x_{-1}} \left(\sum_{x_{-2}} z(x_{-2}, x_{-1}) \right) p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega) \\ &= \sum_{x_{-1}} \left(\sum_{a'} z(x_{-1}, \omega, a') \right) \sigma(a|x_{-1}, \omega) \\ &= \sum_{x_{-1}} z(x_{-1}, \omega, a) \\ &= \pi(\omega, a) \end{aligned}$$

where the first and the fourth equality follows from the definition of π . The second equality follows from the first equality constraint of $\text{LP}(\text{full})$ (and from the feasibility of z), and the third equality follows from the definition of σ . Thus, $\pi = \text{Inv}(\sigma)$. Finally, for any $x_{-1} \in \mathcal{X}$ and all $a, a' \in A$, we have

$$\begin{aligned} \sum_{w \in \mathcal{W}_{\text{full}}} z(w, a) D(x_{-1}, w) \partial u(w, a, a') &= \sum_{\omega \in \Omega} z(x_{-1}, \omega, a) \partial u(\omega, a, a') \\ &= \sum_{\omega} \left(\sum_{a'} z(x_{-1}, \omega, a') \right) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a') \\ &= \sum_{\omega} \left(\sum_{x_{-2}} z(x_{-2}, x_{-1}) p(\omega|x_{-1}) \right) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a') \\ &= \pi(x_{-1}) \cdot \sum_{\omega} p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega) \partial u(\omega, a, a'). \end{aligned}$$

Here, the second equality follows from the definition of σ , the third equality follows from the feasibility of z to $\text{LP}(\text{full})$, and in the final equality, we have used the definition of π . Thus, for all $x_{-1} \in \mathcal{X}$ with $\pi(x_{-1}) > 0$, from the feasibility of z , we obtain

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

and hence a receiver, after observing $\bar{x}_{t-1} = x_{-1}$, would find it optimal to adopt action a if recommended by σ . Finally, from the fact that $\sum_x z(x, \omega, a) = \sum_x z(\omega, a, x)$, we obtain that if $\pi(x_{-1}) = 0$, then $\sum_{a'} z(x_{-1}, \omega, a') = 0$, and hence, σ recommends the receiver-optimal action at each ω . Thus, again, the receiver finds it optimal to follow the recommendation. Taken together, we conclude that $\sigma \in \text{Pers}(\Phi_{\text{full}})$. The converse follows from noticing that the sender's payoff under σ equals the $\text{LP}(\text{full})$ objective at z .

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

A.2 Proofs from Section 4.2

Proof of Lemma 2. We first show that for any $\ell \geq 1$, $\text{Pers}(\Phi_{\text{full}}) \subseteq \text{Pers}(\Phi_\ell)$. To see this, let $\sigma \in \Sigma_k \cap \text{Pers}(\Phi_{\text{full}})$. Define $\hat{\sigma}$ 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^ℓ . Since the information of the receiver under $\hat{\sigma}$ in the model Φ_ℓ is 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^ℓ , 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 $\sigma \in \text{Pers}(\Phi_\ell)$. Thus, we conclude $\text{Pers}(\Phi_{\text{full}}) \subseteq \text{Pers}(\Phi_\ell)$ for $\ell \geq 1$. A similar argument yields $\text{Pers}(\Phi_\ell) \subseteq \text{Pers}(\Phi_{\ell+1}) \subseteq \text{Pers}(\Phi_{\text{no}})$. \square

Proof of Proposition 2. From Lemma 2, we have $\text{OPT}(\Phi_{\text{full}}) \leq \text{OPT}(\Phi_{\text{no}})$; thus, it remains to show that under the conditions of the lemma, $\text{OPT}(\Phi_{\text{full}}) \geq \text{OPT}(\Phi_{\text{no}})$. To show this inequality, we construct a signaling mechanism $\hat{\sigma} \in \text{Pers}(\Phi_{\text{full}}) \cap \Sigma_1$ that achieves the same payoff as the optimal signaling mechanism $\sigma \in \text{Pers}(\Phi_{\text{no}}) \cap \Sigma_0$.

To begin, note that since $p(\cdot|x) \in \text{Conv}(\mathcal{B}_{\text{no}})$ for each $x \in \mathcal{X}$, there exists a set of non-negative weights $\{\lambda(a|x) : a \in A, x \in \mathcal{X}\}$ such that

$$\begin{aligned} p(\omega|x) &= \sum_{a'} \lambda(a'|x) \mu_{a'}(\omega), \quad \text{for all } x \in \mathcal{X}, \omega \in \Omega \\ \sum_{a \in A} \lambda(a|x) &= 1, \quad \text{for all } x \in \mathcal{X}. \end{aligned}$$

Define the mechanism $\hat{\sigma} \in \Sigma_1$ as follows: for each $x_{-1} \in \mathcal{X}$, let

$$\hat{\sigma}(a|x_{-1}, \omega) := \frac{\lambda(a|x_{-1}) \mu_a(\omega)}{\sum_{a'} \lambda(a'|x_{-1}) \mu_{a'}(\omega)} = \frac{\lambda(a|x_{-1}) \mu_a(\omega)}{p(\omega|x_{-1})},$$

whenever the denominator is positive, and otherwise let $\hat{\sigma}$ recommend the receiver-optimal action at ω .

We first show that $\hat{\sigma} \in \text{Pers}(\Phi_{\text{full}})$. For each $x_{-1} \in \mathcal{X}$, we have

$$\begin{aligned} &\sum_{\omega} p(\omega|x_{-1}) \hat{\sigma}(a|x_{-1}, \omega) \partial u(\omega, a, a') \\ &= \sum_{\omega} \left(\sum_{a' \in A} \lambda(a'|x_{-1}) \mu_{a'}(\omega) \right) \frac{\lambda(a|x_{-1}) \mu_a(\omega)}{\sum_{a' \in A} \lambda(a'|x_{-1}) \mu_{a'}(\omega)} \partial u(\omega, a, a') \\ &= \lambda(a|x_{-1}) \left(\sum_{\omega} \mu_a(\omega) \partial u(\omega, a, a') \right). \end{aligned}$$

Since $\mu_a(\omega)$ is the posterior belief induced by $\sigma \in \text{Pers}(\Phi_{\text{no}})$, we have $\sum_{\omega} \mu_a(\omega) \partial u(\omega, a, a') \geq 0$. As $\lambda(a|x_{-1}) \geq 0$, we have for all $x_{-1} \in \mathcal{X}$,

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

If $\sum_{\omega'} p(\omega'|x_{-1}) \hat{\sigma}(a|x_{-1}, \omega') > 0$, then upon dividing by it, we obtain action a is optimal for the receiver if it is recommended by $\hat{\sigma}$ in the model Φ_{full} . On the other hand, if $\sum_{\omega'} p(\omega'|x_{-1}) \hat{\sigma}(a|x_{-1}, \omega') = 0$, then $\hat{\sigma}$ recommends the receiver-optimal action. Thus, we conclude that it is always optimal for the receiver to follow the recommendations by $\hat{\sigma}$ in the model Φ_{full} , and hence $\hat{\sigma} \in \text{Pers}(\Phi_{\text{full}})$.

We now show that σ and $\widehat{\sigma}$ induce the same marginal distribution over \mathcal{X} . Let $\pi = \text{Inv}(\sigma)$. Let $\tau_a := \sum_{\omega} \pi(\omega, a)$, and note that $\pi(\omega, a) = \tau_a \mu_a(\omega)$. Using the definition of $\widehat{\sigma}$, we have for $\omega \in \Omega$ and $a \in A$,

$$\begin{aligned} \sum_{x_{-1} \in \mathcal{X}} \pi(x_{-1}) p(\omega|x_{-1}) \widehat{\sigma}(a|x_{-1}) &= \sum_{x_{-1} \in \mathcal{X}} \pi(x_{-1}) \lambda(a|x_{-1}) \mu_a(\omega) \\ &= \sum_{x_{-1} \in \mathcal{X}} \tau_{a_{-1}} \mu_{a_{-1}}(\omega_{-1}) \lambda(a|x_{-1}) \mu_a(\omega), \end{aligned} \quad (6)$$

where we have used the fact that $\pi(x_{-1}) = \tau_{a_{-1}} \mu_{a_{-1}}(\omega_{-1})$. Summing both sides over a , we obtain

$$\sum_{x_{-1} \in \mathcal{X}} \pi(x_{-1}) p(\omega|x_{-1}) = \sum_{x_{-1} \in \mathcal{X}} \tau_{a_{-1}} \mu_{a_{-1}}(\omega_{-1}) \left(\sum_a \lambda(a|x_{-1}) \mu_a(\omega) \right).$$

Moreover, from the balance equation (1), we have

$$\sum_{x_{-1} \in \mathcal{X}} \pi(x_{-1}) p(\omega|x_{-1}) = \sum_a \pi(\omega, a) = \sum_a \tau_a \mu_a(\omega).$$

Equating the right-hand sides of the two preceding equations, we obtain

$$\sum_a \left(\tau_a - \sum_{x_{-1} \in \mathcal{X}} \tau_{a_{-1}} \mu_{a_{-1}}(\omega_{-1}) \lambda(a|x_{-1}) \right) \mu_a(\omega) = 0$$

Because $\{\mu_a\}$ are linearly independent, we have for all $a \in A$,

$$\tau_a = \sum_{x_{-1} \in \mathcal{X}} \tau_{a_{-1}} \mu_{a_{-1}}(\omega_{-1}) \lambda(a|x_{-1}).$$

Substituting back in (6), we obtain

$$\sum_{x_{-1} \in \mathcal{X}} \pi(x_{-1}) p(\omega|x_{-1}) \widehat{\sigma}(a|x_{-1}) = \tau_a \mu_a(\omega) = \pi(\omega, a). \quad (7)$$

Thus, π is also an invariant distribution under $\widehat{\sigma}$, and thus, the two mechanisms induce the same marginal distribution over $\Delta(\mathcal{X})$. \square

B Proofs from the Section 5

In this section, we provide the missing proofs from Section 5. Throughout, we use the same notation as in that section.

Proof of Proposition 3. The proof of the proposition is similar to that of Proposition 1, and we only highlight the parts that are different.

First, consider a signaling mechanism $\sigma \in \Sigma_1 \cap \text{Pers}(\Phi_1)$, and let $\pi = \text{Inv}(\sigma) \in \Delta(\mathcal{X}^2)$ denote the invariant distribution under σ . Define z as follows:

$$z(x_{-2}, x_{-1}, \omega, a) := \pi(x_{-2}, x_{-1}) p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega).$$

Then, similar arguments to Proposition 1 shows that z satisfies the linear equalities and the inequality in (4). Similarly, the value of the objective is readily seen to equal the sender's payoff under σ . Finally, the non-linear equality holds because, for any $x'_{-2}, x_{-2}, x_{-1}, (\omega, a) \in \mathcal{X}$, we have

$$\begin{aligned} z(x_{-2}, x_{-1}, \omega, a) \cdot \sum_{a' \in A} z(x'_{-2}, x_{-1}, \omega, a') &= \pi(x_{-2}, x_{-1}) p(\omega|x_{-1}) \sigma(a|x_{-1}, \omega) \pi(x'_{-2}, x_{-1}) p(\omega|x_{-1}) \\ &= z(x'_{-2}, x_{-1}, \omega, a) \cdot \sum_{a' \in A} z(x_{-2}, x_{-1}, \omega, a'). \end{aligned}$$

Conversely, let z be any feasible solution to (4). We define $\sigma \in \Sigma_1$ as follows: For any $x_{-1} \in \mathcal{X}$ and $\omega \in \Omega$, if there exists an $x_2 \in \mathcal{X}$ such that $\sum_{a' \in A} z(x_{-2}, x_{-1}, \omega, a') > 0$, let

$$\sigma(a|x_{-1}, \omega) = \frac{z(x_{-2}, x_{-1}, \omega, a)}{\sum_{a' \in A} z(x_{-2}, x_{-1}, \omega, a')}.$$

Note that the non-linear constraint on z implies that the right-hand side does not depend on x_{-2} , and thus σ is well-defined. On the other hand, if $\sum_{a' \in A} z(x_{-2}, x_{-1}, \omega, a') = 0$ for all $x_{-2} \in \mathcal{X}$, let $\sigma(\cdot|x_{-1}, \omega)$ recommend the receiver-optimal action at ω . Furthermore, define $\pi \in \Delta(\mathcal{X}^2)$ as $\pi(x_{-1}, \omega, a) := \sum_{x_{-2}} z(x_{-2}, x_{-1}, \omega, a)$. Using similar arguments in Proposition 1, it follows that $\pi = \text{Inv}(\sigma)$ and furthermore that $\sigma \in \text{Pers}(\Phi_1)$. The final step is to see that the objective of the non-linear program at z equals the sender's payoff under σ . \square

Proof of Lemma 3. Let $\sigma \in \text{RP}(\epsilon)$ 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 $\pi = \text{Inv}(\sigma)$. Since $\sigma \in \text{RP}(\epsilon)$ 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 $\sigma \in \text{Pers}(\Phi_\ell)$.

To prove the bound in the lemma statement, we note that since $\sigma \in \text{RP}(\epsilon) \subseteq \Sigma_0$, it corresponds to a stationary Markov policy, and hence the induced Markov chain over the states is ergodic by Assumption 1. The result is then obtained using the following bound on the mixing time of this chain [Levin and Peres, 2017, Theorem 12.4]:

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

where $\gamma_*(\sigma)$ 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 1, and hence $\pi_{\min}(\sigma)$ is well defined. \square

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 $\pi \in \Delta(\Omega \times A)$ as $\pi(\omega, a) = \sum_{s \in S} w_s \mu_s(\omega) \mathbf{I}\{a_s = a\}$. We claim that π is the steady state distribution under σ , when each receiver chooses the action a_s after receiving signal $s \in S$. (We show below that this is indeed optimal for the receiver in the information model Φ_{no} .) This follows from noticing that π satisfies

the balance equations, as we show next. For each $\omega \in \Omega$, we have

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

Here, the final equality follows from the assumption made on the weights $\{w_s\}$ in the lemma statement. This, in turn implies that

$$\begin{aligned}
\sum_{\omega', a'} \pi(\omega', a') p(\omega | \omega', a') \sum_{s' \in S} \sigma(s' | \omega) \mathbf{I}\{a_{s'} = a\} &= \sum_{s \in S} w_s \mu_s(\omega) \sum_{s' \in S} \sigma(s' | \omega) \mathbf{I}\{a_{s'} = a\} \\
&= \sum_{s' \in S} \mathbf{I}\{a_{s'} = a\} \sigma(s' | \omega) \left(\sum_{s \in S} w_s \mu_s(\omega) \right) \\
&= \sum_{s' \in S} \mathbf{I}\{a_{s'} = a\} w_{s'} \mu_{s'}(\omega) \\
&= \pi(\omega, a),
\end{aligned}$$

where the penultimate inequality follows from the definition of $\sigma(s | \omega)$. Thus, we conclude that π satisfies the balance equations.

Since π is the invariant distribution under σ , the marginal distribution of the state (and hence the receivers' prior belief in the model Φ_{no}) equals $\pi(\omega) = \sum_{a \in A} \pi(\omega, a) = \sum_{s \in S} w_s \mu_s(\omega)$. From the definition of σ , we have $\pi(\omega) \sigma(s | \omega) = w_s \mu_s(\omega)$. Thus, in the information model Φ_{no} , the posterior belief of a receiver that the state is ω upon receiving a signal $s \in S$ is given by Bayes' rule as

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

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. This concludes the proof of the first part of the lemma statement.

To show the converse, let $\sigma \in \text{Pers}(\Phi_{\text{no}}) \cap \Sigma_0$ and let $\pi = \text{Inv}(\sigma)$. For any $a \in A$ with $\sum_{\omega'} \pi(\omega', a) > 0$, define $w_a := \sum_{\omega'} \pi(\omega', a)$, and $\mu_a := \frac{\pi(\cdot, a)}{w_a}$. For any $a \in A$ with $\sum_{\omega'} \pi(\omega', a) = 0$, define $w_a = 0$ and μ_a be any belief in \mathcal{P}_a . By construction, we have $\sum_{a \in A} w_a = 1$ and $\pi(\omega, a) = w_a \mu_a(\omega)$ for all $(\omega, a) \in \Omega \times A$. As $\sigma \in \text{Pers}(\Phi_{\text{no}}) \cap \Sigma_0$, we have $\sum_{\omega} \pi(\omega, a) \partial u(\omega, a, a') \geq 0$ for all $a, a' \in A$. Consequently, we obtain $w_a \sum_{\omega} \mu_a(\omega) \partial u(\omega, a, a') \geq 0$, implying that $\mu_a \in \mathcal{P}_a$ also holds

for any a with $w_a > 0$. Finally, for all $\omega \in \Omega$, we obtain

$$\begin{aligned} \sum_a w_a \mu_a(\omega) &= \sum_a \pi(\omega, a) \\ &= \sum_a \sum_{\omega', a'} \pi(\omega', a') p(\omega | \omega', a') \sigma(a | \omega) \\ &= \sum_{\omega', a'} \pi(\omega', a') p(\omega | \omega', a') \\ &= \sum_{\omega', a'} w_{a'} \mu_{a'}(\omega') p(\omega | \omega', a'). \end{aligned}$$

Here, the second equality follows because π is the steady-state distribution induced by σ . Thus, we conclude that $\{w_a, \mu_a\}_a$ satisfies all the conditions in the lemma statement. \square

Proof of Theorem 3. To begin, let $\sigma \in \text{Pers}(\Phi_{\text{no}})$ denote the optimal signaling mechanism in the no-history information model Φ_{no} . Let $\pi = \text{Inv}(\sigma)$ 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

$$\sum_{\omega, a} w_a \mu_a(\omega) p(\cdot | \omega, a) = \sum_a w_a \mu_a. \quad (8)$$

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 $\hat{\sigma}$ 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 2, for any $a \in A_+$, let $\eta_a \in \mathcal{P}_a$ be such that $\mathcal{B}_1(\eta_a, D) \subseteq \mathcal{P}_a$. 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 $\mathcal{B}_1(\xi_a, \delta D) \subseteq \mathcal{P}_a$. 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 $\hat{\sigma}$ 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 $\hat{\sigma}$ 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

$$\sum_{a \in A_+} \sum_{\omega} \hat{w}_a \xi_a(\omega) p(\cdot | \omega, a) + \sum_{\omega} \hat{w}_\omega p(\cdot | \omega, a_\omega) = \sum_{a \in A_+} \hat{w}_a \xi_a + \sum_{\omega} \hat{w}_\omega e_\omega. \quad (9)$$

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 $\rho := \sum_{\omega} \hat{w}_\omega$. Then, after moving all terms containing \hat{w}_ω on one side, the condition (9) becomes

$$\begin{aligned} \frac{1}{1 - \rho} \left(\sum_{\omega} \hat{w}_\omega e_\omega - \sum_{\omega} \hat{w}_\omega p(\cdot | \omega, a_\omega) \right) &= \sum_{a \in A_+} \sum_{\omega} w_a \xi_a(\omega) p(\cdot | \omega, a) - \sum_{a \in A_+} w_a \xi_a \\ &= \delta \left(\sum_{a \in A_+} \sum_{\omega} w_a \eta_a(\omega) p(\cdot | \omega, a) - \sum_{a \in A_+} w_a \eta_a \right), \end{aligned} \quad (10)$$

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 (8).

In Lemma 5, we show that there exists $y = (y_\omega \geq 0 : \omega \in \Omega)$ satisfying

$$\sum_{\omega} y_\omega e_{\omega} - \sum_{\omega} y_{\omega} p(\cdot | \omega, a_{\omega}) = \sum_{a \in A_+} \sum_{\omega} w_a \eta_a(\omega) p(\cdot | \omega, a) - \sum_{a \in A_+} w_a \eta_a. \quad (11)$$

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 (10), and hence, there exist weights satisfying (9).

Thus, by Lemma 4, we obtain the existence of a history-independent signaling mechanism $\hat{\sigma}$ 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 $\hat{\sigma}$ sends signals with the following probabilities: for each $\omega \in \Omega$:

$$\hat{\sigma}(s | \omega) := \begin{cases} \frac{\hat{w}_a \xi_a(\omega)}{\sum_{a'} \hat{w}_{a'} \xi_{a'}(\omega) + \hat{w}_{\omega}} & \text{for } s \in A_+; \\ \frac{\hat{w}_{\omega}}{\sum_{a'} \hat{w}_{a'} \xi_{a'}(\omega) + \hat{w}_{\omega}} & \text{if } s = \omega; \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

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 $\hat{\sigma}$ is persuasive, in the sense that, the receiver finds it optimal to follow the recommendation.

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

Robustness: Next, we show that $\hat{\sigma} \in \text{RP}(\epsilon)$. Suppose a receiver's belief about $\bar{\omega}_t$ is given by a distribution $\pi' \in \Delta(\Omega)$, with $\|\pi' - \hat{\pi}\|_1 \leq \epsilon$. Upon receiving a signal $s = \omega \in S$ from the signaling mechanism $\hat{\sigma}$, 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 $\xi'_a(\omega) = \frac{\mu'(\omega) \hat{\sigma}(a | \omega)}{\sum_{\omega'} \mu'(\omega') \hat{\sigma}(a | \omega')}$. Using a similar argument as in [Zu et al., 2021], we obtain that the following bound on the ℓ_1 distance between ξ'_a and ξ_a (for the sake of completeness, we include the algebraic steps in Lemma 6):

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

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

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

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

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

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

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 $\xi'_a \in \mathcal{B}_1(\eta_a, \delta D) \subseteq \mathcal{P}_a$. Hence, starting with a prior π' with $\|\pi' - \hat{\pi}\| \leq \epsilon$, 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 $\hat{\sigma}$ is persuasive for all beliefs $\pi' \in \mathcal{B}_1(\hat{\pi}, \epsilon)$, 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 $\hat{\sigma}$, as follows:

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

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

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

In Lemma 5, we show that there exists a solution $y \geq 0$ to (11) satisfying $\|y\|_1 \leq \frac{2(1+\tau)\sqrt{|\Omega|}}{s_f}$. Thus, we obtain that the sender's expected payoff under $\hat{\sigma}$ is lower-bounded by

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

This completes the proof. \square

The following lemma is used in the proof of Theorem 3.

Lemma 5. *Consider the following linear program:*

$$\begin{aligned} \min_{y \geq 0} \quad & \sum_{\omega \in \Omega} y_{\omega} \\ \sum_{\omega} y_{\omega} e_{\omega} - \sum_{\omega} y_{\omega} p(\cdot | \omega, a_{\omega}) &= \sum_{a \in A_+} \sum_{\omega} w_a \eta_a(\omega) p(\cdot | \omega, a) - \sum_{a \in A_+} w_a \eta_a. \end{aligned} \quad (13)$$

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 (13), we first cast it into a matrix form. Let $P_a \in \mathbb{R}^{|\Omega| \times |\Omega|}$ be the matrix with $P_a(\omega, \omega') := p(\omega' | \omega, a)$. Then, the LP (13) can be recasted as

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

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

Since P_f and $\{P_a\}_{a \in A}$ are transition kernels, we have that $(I - P_f)\mathbf{1} = (I - P_a)\mathbf{1} = 0$ for all $a \in A$. This implies that $\mathbf{1}$ does not lie in the row span of $I - P_f$ and $I - P_a$ for any $a \in A_+$, and hence $\mathbf{1}$ is orthogonal to vector $\sum_{a \in A_+} w_a \eta_a^T (P_a - I)$. Since P_f is ergodic (from Assumption 1), we have $\text{rank}(I - P_f) = |\Omega| - 1$ and thus 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. Thus, it remains to be shown that there exists one that is non-negative.

Let u be any solution to the equality in the linear program. Because P_f is ergodic, there is a unique stationary distribution ν_f such that $\nu_f^T (I - P_f) = 0$ and each element $\nu_f(\omega) > 0$ for all $\omega \in \Omega$. Let $y = u + k\nu_f$ where $k = \max_{\omega: u(\omega) < 0} \frac{|u(\omega)|}{\nu_f(\omega)}$ is chosen so that $y \geq 0$. Hence, we obtain that y is feasible for the LP (13).

The proof is complete upon showing that $\|y\|_1 = \|u + k\nu_f\|_1 \leq \frac{2(1+\tau)\sqrt{|\Omega|}}{s_f}$. To see this, note that since $\text{rank}(P_f - I) = |\Omega| - 1$, the matrix $I - P_f$ has a singular value decomposition $Q\Lambda\tilde{Q}^T$, where $\Lambda \in \mathbb{R}^{(|\Omega|-1) \times (|\Omega|-1)}$ is the diagonal matrix of singular values and $Q, \tilde{Q} \in \mathbb{R}^{|\Omega| \times (|\Omega|-1)}$ are composed of orthogonal column vectors. Because multiplying by an orthogonal matrix preserves the ℓ_2 norm, we obtain

$$\|u^T (I - P_f)\|_2 = \|u^T Q\Lambda\tilde{Q}^T\|_2 = \|u^T Q\Lambda\|_2 \geq s_f \|u^T Q\|_2 = s_f \|u\|_2 \geq \frac{s_f}{\sqrt{|\Omega|}} \|u\|_1,$$

where s_f is the smallest diagonal element (i.e., the smallest singular value) of Λ , and the final inequality follows from the relationship between ℓ_1 and ℓ_2 norms. Note that from Assumption 1, we have $s_f > 0$. On the other hand,

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

where the second inequality follows triangle inequality. Taken together, $\|u\|_1 \leq \frac{2\sqrt{|\Omega|}}{s_f}$. Finally observe that $k = \max_{\omega: u(\omega) < 0} \frac{|u(\omega)|}{\nu_f(\omega)} \leq \max_{\omega} \frac{1}{\nu_f(\omega)} \cdot \max_{\omega: u(\omega) < 0} |u(\omega)| \leq \tau \|u\|_1$. Hence, using the fact that $\|\nu_f\|_1 = 1$, we obtain $\|y\|_1 = \|u + k\nu_f\|_1 \leq \|u\|_1 + k\|\nu_f\|_1 = \|u\|_1 + k \leq (1 + \tau)\|u\|_1 \leq \frac{2(1+\tau)\sqrt{|\Omega|}}{s_f}$. \square

The following lemma establishes the continuity of the Bayes' update. The proof is from [Zu et al., 2021]; we include it here for completeness. We use the same notation as in the proof of Theorem 3.

Lemma 6. *For each $a \in A_+$, we have*

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

Proof. We obtain

$$\begin{aligned}
\|\xi'_a - \xi_a\|_1 &= \sum_{\omega \in \Omega} |\xi'_a(\omega) - \xi_a(\omega)| \\
&= \sum_{\omega \in \Omega} \left| \frac{\pi'(\omega)\hat{\sigma}(a|\omega)}{\sum_{\omega'} \pi'(\omega')\hat{\sigma}(a|\omega')} - \frac{\hat{\pi}(\omega)\hat{\sigma}(a|\omega)}{\sum_{\omega'} \hat{\pi}(\omega')\hat{\sigma}(a|\omega')} \right| \\
&\leq \sum_{\omega \in \Omega} \left| \frac{\pi'(\omega)\hat{\sigma}(a|\omega)}{\sum_{\omega'} \pi'(\omega')\hat{\sigma}(a|\omega')} - \frac{\pi'(\omega)\hat{\sigma}(a|\omega)}{\sum_{\omega'} \hat{\pi}(\omega')\hat{\sigma}(a|\omega')} \right| \\
&\quad + \sum_{\omega \in \Omega} \left| \frac{\pi'(\omega)\hat{\sigma}(a|\omega)}{\sum_{\omega'} \hat{\pi}(\omega')\hat{\sigma}(a|\omega')} - \frac{\hat{\pi}(\omega)\hat{\sigma}(a|\omega)}{\sum_{\omega'} \hat{\pi}(\omega')\hat{\sigma}(a|\omega')} \right| \\
&\leq \left| \sum_{\omega} \frac{\hat{\sigma}(a|\omega)}{\sum_{\omega'} \hat{\pi}(\omega')\hat{\sigma}(a|\omega')} (\hat{\pi}(\omega) - \pi'(\omega)) \right| \\
&\quad + \sum_{\omega \in \Omega} \frac{\hat{\sigma}(a|\omega)}{\sum_{\omega'} \hat{\pi}(\omega')\hat{\sigma}(a|\omega')} |\pi'(\omega) - \hat{\pi}(\omega)| \\
&\leq 2 \sum_{\omega \in \Omega} \frac{\hat{\sigma}(a|\omega)}{\sum_{\omega'} \hat{\pi}(\omega')\hat{\sigma}(a|\omega')} |\pi'(\omega) - \hat{\pi}(\omega)| \\
&\leq 2 \left(\sup_{\omega \in \Omega} \frac{\hat{\sigma}(a|\omega)}{\sum_{\omega'} \hat{\pi}(\omega')\hat{\sigma}(a|\omega')} \right) \cdot \|\pi' - \hat{\pi}\|_1 \\
&= 2 \left(\sup_{\omega \in \Omega} \frac{\xi_a(\omega)}{\hat{\pi}(\omega)} \right) \cdot \|\pi' - \hat{\pi}\|_1,
\end{aligned}$$

where the last equality follows from the definition of $\xi_a(\omega)$. \square