Operations Research

Learning to Persuade on the Fly: Robustness Against Ignorance

You Zu, Krishnamurthy Iyer, Haifeng Xu

Please scroll down for article—it is on subsequent pages

Crosscutting Areas

Learning to Persuade on the Fly: Robustness Against Ignorance

You Zu,[a] Krishnamurthy Iyer,[a,*] Haifeng Xu[b]

[a] Industrial and Systems Engineering, University of Minnesota, Minneapolis, Minnesota 55455; [b] Department of Computer Science, University of Chicago, Chicago, Illinois 60637
*Corresponding author
Contact: zu000002@umn.edu, https://orcid.org/0000-0002-0091-4123 (YZ); kriyer@umn.edu, https://orcid.org/0000-0002-5538-1432 (KI); haifengxu@uchicago.edu, https://orcid.org/0000-0001-6371-4906 (HX)

Abstract. Motivated by information sharing in online platforms, we study repeated persuasion between a sender and a stream of receivers, where, at each time, the sender observes a payoff-relevant state drawn independently and identically from an *unknown* distribution and shares state information with the receivers, who each choose an action. The sender seeks to persuade the receivers into taking actions aligned with the sender's preference by selectively sharing state information. However, in contrast to the standard models, neither the sender nor the receivers know the distribution, and the sender has to persuade while learning the distribution on the fly. We study the sender's learning problem of making *persuasive* action recommendations to achieve low regret against the optimal persuasion mechanism with the knowledge of the distribution. To do this, we first propose and motivate a persuasiveness criterion for the unknown distribution setting that centers *robustness* as a requirement in the face of uncertainty. Our main result is an algorithm that, with high probability, is robustly persuasive and achieves $O(\sqrt{T \log T})$ regret, where T is the horizon length. Intuitively, at each time, our algorithm maintains a set of candidate distribution and chooses a signaling mechanism that is simultaneously persuasive for all of them. Core to our proof is a tight analysis about the cost of robust persuasion, which may be of independent interest. We further prove that this regret order is optimal (up to logarithmic terms) by showing that no algorithm can achieve regret better than $\Omega(\sqrt{T})$.

Keywords: no-regret learning • robustness • persuasion • prior-independence

1. Introduction

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

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

Example 1 (Content Recommendations by Online Media Platforms). Consider a media platform, like YouTube or TikTok, that recommends content created by independent creators ("channels") to its users. New channels regularly join the platform and start producing

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

Example 2 (Recommendations on Hiring Platforms). Consider a hiring platform, where employers receive recommendations about candidates for recruitment (e.g., "recommended matches" in LinkedIn Recruiter). These recommendations are typically tailored to the employer's project requirements. However, within the set of candidates satisfying the requirements, there would be a range of capabilities/fit, whose *distribution* would be unknown to the platform or the employer. Nevertheless, for any particular candidate who might be interested in the position, the platform may be able to assess the candidate's capability based on various candidate features, such as her endorsements, references, etc.; by using this information, the platform decides whether to recommend the candidate. Similarly, the employer, through the course of interviewing different candidates, may learn about the capability distribution. Although the employer would prefer to be matched with few high-capability candidates to interview, the platform may have additional incentives from having to cater to the candidates' side of the market, such as increasing the overall number of interviews.

Once again, if the distribution of the candidates' capabilities is known to the platform and the employer, the platform could reliably recommend candidates to optimize its goals, while simultaneously meeting the employer's preferences. But, without such information, the platform needs to learn to recommend candidates as they apply over time.

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

1.2. Modeling Contributions

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

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

Certainly, the sender cannot freely make arbitrary recommendations, if the expectation is that these recommendations would influence the receivers' actions. A

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

Our main modeling contribution addresses this issue by proposing a natural criteria for persuasiveness when neither the sender nor the receivers know the state distribution. The starting point of our approach is the observation that any persuasiveness criteria directly corresponds to a model of receivers' response on receiving a recommendation (just as in the case of known state distribution). Thus, by considering reasonable behavioral models for the receiver, we develop in Section 2.2 a persuasiveness criterion that centers *robustness* as a requirement in the face of uncertainty. Specifically, our criterion requires that the sender's recommendations are persuasive under all state distributions in a set of "confidence regions," which contain the true distribution with a given degree of confidence; these confidence regions shrink over time as the sender observes more state realizations. This is in line with the approach in statistics that uses confidence regions to address the uncertainty in parameter estimates. Furthermore, this robustness requirement naturally leads to conservative recommendations, thereby making it likely that the recommendations will be accepted. We refer to this notion as β-robustly persuasiveness, where $1 - \beta$ denotes the confidence level.

1.3. Algorithmic Contribution and Regret Characterization

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

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

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

To characterize the regret of the $\mathfrak{Rai}$ algorithm, we next undertake a brief digression, in Section 4, into studying the (static) problem of robust persuasion. Specifically, we study a static persuasion setting with known state distribution, but impose the restriction that the signaling mechanism must be persuasive for all distributions in the neighborhood of the actual state distribution. For this problem, we define and analyze a quantity Gap that measures the sender's *cost of robust persuasion*. Formally, $\mathsf{Gap}(\mu, \mathcal{B})$ captures the loss in the sender's expected utility (under distribution μ) from using a signaling mechanism that is persuasive for all distributions in the set $\mathcal{B}$, as opposed to using one that is persuasive only for the distribution μ. In Proposition 2, we establish that, under some regularity conditions, the sender's cost of robust persuasion $\mathsf{Gap}(\mu, \mathcal{B})$ is at most linear in the radius of the set $\mathcal{B}$. This is achieved via an explicit construction of a signaling mechanism that is persuasive for all distributions in $\mathcal{B}$ and achieves sender's utility close to the optimum. Further, we provide a matching lower bound in Proposition 3 by carefully crafting a persuasion instance and using its geometry to prove a linear cost of robust persuasion; this instance thus serves as a lower-bound example for robust persuasion. The characterization of the *cost of robust persuasion* provides useful insight about the problem of robust persuasion, which may be of independent interest.

Using this characterization of the cost of robust persuasion, we perform a tight regret analysis of persuasion under unknown state distribution in Section 5. Our positive result, Theorem 2, establishes that for any persuasion setting satisfying the aforementioned regularity conditions, the $\mathfrak{Rai}$ algorithm achieves $O(\sqrt{T \log T})$ regret with high probability. Furthermore, in Theorem 3, we provide a matching lower bound (up to $\log T$ terms)

for the regret of any algorithm that makes persuasive recommendations. In addition to the characterization of Gap and the custom persuasion instance from Propositions 2 and 3, the proofs of these theorems rely on concentration results for sums of independent random vectors in Banach spaces.

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

1.4. Literature Survey

Our paper contributes to the burgeoning literature on Bayesian persuasion and information design in economics, operations research, and computer science. We refer readers to Kamenica and Gentzkow (2011) and Bergemann and Morris (2019), as well as Candogan (2020), for a general overview of the recent developments and Dughmi (2017) for a survey from algorithmic perspective.

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

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

Dispensing with the common prior itself, Camara et al. (2020) study an adversarial online learning model, where both a mechanism designer and the agent learn about the states over time. The agent is long-lived and

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

Recent works by Hahn et al. (2020, 2022) study information design in online optimization problems, such as the secretary problem (Hahn et al. 2022) and the prophet inequalities (Hahn et al. 2020), and propose constant-approximation persuasive schemes. These online optimization problems often take the adversarial approach, which is different from our stochastic setup and learning-focused tasks. Therefore, our results are not comparable.

1.4.2. Robust Persuasion. The algorithm we propose relies crucially on robust persuasion due to the ignorance of the prior, and, as a part of establishing the regret bounds for the algorithm, we quantify the sender's cost of robustness. Kosterina (2022) studies a persuasion setting in the absence of the common prior assumption. In particular, the sender has a known prior, whereas only the set in which the receiver's prior lies is known to the sender. Furthermore, the sender evaluates the expected utility under each signaling mechanism with respect to the worst-case prior of the receiver. Similarly, Hu and Weng (2021) study the problem of the sender persuading a privately informed receiver, where the sender seeks to maximize her expected payoff under the worst-case information of the receiver. Finally, Dworczak and Pavan (2020) study a related setting and propose a lexicographic solution concept, where the sender first identifies the signaling mechanisms that maximize her worst-case payoff and then, among them, chooses the one that maximizes the expected utility under her conjectured prior. In contrast to these works, our model focuses on a setting with a common, but unknown, prior and where the receiver has no private information. Instead, our notion of robustness is with respect to this unknown (common) prior.

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

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

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

Recent works by Pacchiano et al. (2021), Khezeli and Bitar (2019), and Moradipari et al. (2020, 2021) study a similar safe learning problem in different contexts. Pacchiano et al. (2021) require that, at each time, the chosen action has an expected cost below a certain threshold. Khezeli and Bitar (2019) and Moradipari et al. (2020) study safe learning, where, in addition to maximizing the expected reward, one requires the reward to be above a threshold with high probability. In these settings, the objective and the constraint are aligned. Our

setup is different because the sender's and the receivers' preferences, corresponding, respectively, to the objective and constraints, need not be aligned with each other. Most importantly, all these works impose a single constraint at each round, whereas our persuasiveness condition requires multiple constraints at each round.

1.4.4. Online Linear/Convex Optimization. Because the persuasion problem can be posed as a linear program, our work also relates to the online convex optimization problem. Mostly, the focus here is on adversarial setting, where the loss function (objective) is adversarially chosen and revealed at the end of each time period. Some papers (Mahdavi et al. 2013, Cao et al. 2019) focus on the stochastic setting, but either study an unconstrained problem (Cao et al. 2019) or study a batch algorithm rather than an online algorithm (Mahdavi et al. 2013). Focusing on the constraints, and using the terminology of Kim and Lee (2023), these works typically consider either a *long-term constraint* formulation (Mahdavi et al. 2011, Neely and Yu 2017, Yu et al. 2017, Cao and Liu 2018, Yi et al. 2021, Kim and Lee 2023) or consider a *cumulative constraint* formulation (Yuan and Lamperski 2018, Yi et al. 2023, Guo et al. 2024). The *long-term constraint* formulation requires feasibility, on average, in the long run. Such constraints are reasonable in applications where the constraints are on aggregate quantities, such as budgets in online advertising (Liakopoulos et al. 2019), covering constraints in sensor networks, capacity constraints in online routing (Agrawal and Devanur 2014), etc. However, this type of constraint is not reasonable in our setting, as it would permit the sender to make poor recommendations in some rounds, as long as it can be compensated by good recommendations in other rounds. In contrast, the *cumulative* constraint formulation focuses on bounding the sum of the positive parts of the constraints (which require some quantity to be nonpositive). This formulation is equivalent to our formulation if the cumulative constraint can be made zero. However, most previous work allows for some constraint violation and seeks to bound the order of the violations. In the presence of such violations, our formulation is stronger.

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

2. Model
Consider a persuasion setting with a single long-run *sender* persuading a stream of homogeneous *receivers*

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

At each time $t \in [T]$, the sender observes the realized state ω_t and shares with the arriving receiver an *action recommendation* $a_t \in A$ (chosen according to a *signaling algorithm*, as described below), where A is a finite set of actions available to the receivers. The receiver then chooses an action $\hat{a}_t \in A$ (not necessarily equal to a_t). This results in the receiver obtaining a utility $u(\omega_t, \hat{a}_t)$ and the sender obtaining a utility $v(\omega_t, \hat{a}_t)$. Without loss of generality, we assume that $v(\omega, a) \in [0,1]$ for all $\omega \in \Omega$ and $a \in A$. Further, to avoid trivialities, we assume $|\Omega| \geq 2$ and $|A| \geq 2$. We refer to the tuple $\mathcal{I} = (\Omega, A, u, v, \mathcal{B}_0)$ with $u : \Omega \times A \to \mathbb{R}$ and $v : \Omega \times A \to [0,1]$ as an *instance* of our problem.

Before we proceed, we make a few remarks on the persuasion instance. First, the preceding description does not specify a model of the receivers' actions $\hat{a}_t$. As we discuss in Section 2.2, this issue is intertwined with the *persuasiveness constraints* that we impose on the sender's signaling algorithm, and, hence, we postpone the discussion until then. Second, and relatedly, although we have assumed that that sender shares information in the form of action recommendations, under the persuasiveness constraints that we consider, it can be shown that this is without loss of generality. Third, although our definition of an instance assumes that the receivers are homogeneous, it can be extended to allow for heterogeneity of receivers' utility; our results continue to apply in the setting where the receivers' types are observable to the sender. Finally, we assume that the sender knows the receivers' utility. This is justified in the context of our applications of interest—namely, online platforms, where, given the scale, the platform may have good estimates about user utility from extensive user-level data.

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

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

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

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

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

Thus, to evaluate the performance of a signaling algorithm, we need a model of the receivers' response subsequent to receiving the action recommendations. Rather than directly specifying such a response model, we instead model conditions on the signaling algorithm $\mathfrak{a}$, which result in *obedient* responses from the receivers—that is, which lead each receiver to choose the action recommended: $\hat{a}_t = a_t$. Any such condition on the signaling algorithm $\mathfrak{a}$ implies a model of receivers' response, and the converse can be assumed without loss of generality by invoking incentive compatibility and the revelation principle. Henceforth, we refer to such a condition as a *persuasiveness criterion*.

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

2.1. Benchmark: Known State Distribution

Consider the setting where the sender and the receivers commonly know the state distribution $\mu^* = \mu \in \Delta(\Omega)$. In this setting, each receiver responds by choosing the action that maximizes her expected utility under the posterior belief about the state, given the action recommendation. In particular, the sender's problem decouples across time periods, and standard results (Kamenica and Gentzkow 2011, Bergemann and Morris

2019, Dughmi and Xu 2021) imply that the sender's problem at each period can be formulated as a linear program.

To elaborate, fix a time $t \in [T]$ and history h_t, and consider the persuasion problem between the sender and the arriving receiver. Recall that $\sigma(h_t, \omega, a)$ denotes the probability with which the sender recommends action $a_t = a$ if the realized state is $\omega_t = \omega$. We refer to $\sigma[h_t] \triangleq (\sigma(h_t, \omega, a) : \omega \in \Omega, a \in A)$ as the *signaling mechanism* at time t and drop the dependence on h_t if the context is clear. Finally, let $\mathcal{S} = \{\sigma : \sigma(\omega, \cdot) \in \Delta(A) \text{ for each } \omega \in \Omega\}$ denote the set of all signaling mechanisms.

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

$$\mathsf{Pers}(\mu) \triangleq \left\{\sigma \in \mathcal{S} : \sum_{\omega \in \Omega} \mu(\omega)\sigma(\omega, a)(u(\omega, a) - u(\omega, a')) \geq 0, \text{ for all } a, a' \in A\right\}. \quad (1)$$

We note that the set $\mathsf{Pers}(\mu)$ is a convex polytope for all $\mu \in \Delta(\Omega)$. Furthermore, the set $\mathsf{Pers}(\mu)$ is nonempty because it always contains the "full-information mechanism," which recommends the receiver's optimal action at each state.

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

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

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

$$\mathsf{OPT}_{\mathcal{I}}(\mu) \triangleq \max_{\sigma} V(\mu, \sigma), \text{ subject to } \sigma \in \mathsf{Pers}(\mu). \quad (2)$$

Finally, letting σ^* denote an optimal signaling mechanism to the preceding optimization problem, the algorithm $\mathfrak{a}$

that sets $\sigma^{\mathfrak{a}}(h_t, \omega_t, a) = \sigma^*(\omega_t, a)$ after any history h_t optimizes the sender's total expected utility when the state distribution is known, with total expected utility given by $T \cdot \mathsf{OPT}_{\mathcal{I}}(\mu)$.

2.2. Persuasiveness Criterion: Unknown Distribution

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

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

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

The naive criterion can be motivated through a particular behavioral model of the receivers involving social learning. Specifically, consider a platform setting, where each receiver (i.e., a user) arrives with an uninformative Haldane prior (Haldane 1948, Villegas 1977, Jaynes 2003) over the state distribution μ^* and observes all the past state realizations. The latter holds if we assume there is social learning among the receivers, where each receiver leaves a feedback that is read by all subsequent receivers. Then, at each time t, the corresponding receiver's belief about the state would be exactly the empirical distribution γ_t, and, thus, the receiver would optimally accept the recommendation made by the platform if it uses a signaling algorithm satisfying the naive criterion.

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

about the state would be, in general, different from the empirical distribution.

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

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

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

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

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

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

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

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

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

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

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

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

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

2.3. Sender's Learning Problem

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

$$\mathsf{Reg}_{\mathcal{I}}(\mathfrak{a}, T, \mu^*) \triangleq T \cdot \mathsf{OPT}_{\mathcal{I}}(\mu^*) - V_{\mathcal{I}}(\mathfrak{a}, T)$$
$$= T \cdot \mathsf{OPT}_{\mathcal{I}}(\mu^*) - \sum_{t \in [T]} v(\omega_t, a_t). \quad (3)$$

We are now ready to formalize the sender's learning problem. Begin by noticing that one must require the signaling algorithm $\mathfrak{a}$ to be β-robustly persuasive for some small β in order for the second equality above to hold—that is, for the receivers' responses to match the recommendations. At the same time, 0-robustly persuasiveness is an excessive requirement, with no hope of resulting in a sublinear regret. (In Appendix A2 of the e-companion, we present an example instance where any 0-robustly persuasive algorithm necessarily obtains a linear regret.) Thus, the central problem is to design, for any given instance $\mathcal{I}$, an algorithm $\mathfrak{a}$ that is β-robustly persuasive for small (vanishing) β and simultaneously achieves sublinear regret with high probability.

3. The Robustness Against Ignorance Algorithm

Having described the learning problem faced by the sender, in this section, we present a signaling algorithm that we call the Robustness Against Ignorance ($\mathfrak{Rai}$) algorithm. Here, we show that the $\mathfrak{Rai}$ algorithm is β-robustly persuasive with $\beta = o(1)$, relegating the regret analysis to Section 5.

Before describing our proposed algorithm, we briefly motivate our design approach. Observe that if the state distribution μ^* is known, then the sender's problem is given by the Linear Program (2), and, thus, the optimal

signaling mechanism can be efficiently computed. Thus, a natural learning approach is to solve at each time t the estimated version of the LP (2), where the unknown distribution μ^* is replaced by the empirical distribution γ_t, and use the corresponding optimal signaling mechanism for that time period. However, this alone is not sufficient to obtain an algorithm that is β-robustly persuasive, which requires the signaling mechanisms to be persuasive for all distributions in some small neighborhood of μ^*. To elaborate, simply solving the estimated LP may yield solutions that are only ϵ-feasible for distributions close to the empirical distribution—that is, some of the persuasiveness constraints for such nearby distributions may get violated. In fact, optimizing the estimated LP may result in a mechanism that is not persuasive for *any* other distribution close to the empirical distribution. Thus, an immediate challenge is in determining how to use the empirical distribution estimate to find well-performing signaling mechanisms that are persuasive (with high probability) for *all* distributions in a small neighborhood around the unknown state distribution. Part of this challenge is to carefully choose the corresponding neighborhoods without significantly sacrificing the performance of the mechanism.

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

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

because it contains the full-information signaling mechanism. Furthermore, we let $B_1(\mu, \epsilon) \triangleq \{\mu' \in \Delta(\Omega) : \|\mu' - \mu\|_1 \leq \epsilon\}$ denote the (closed) ℓ_1-ball of radius $\epsilon > 0$ at $\mu \in \Delta(\Omega)$.

Algorithm 1 (The Robustness Against Ignorance ($\mathfrak{Rai}$) Algorithm)

> **Input:** Instance $\mathcal{I}$, Time horizon T
> **Parameters:** $\gamma_0 \in \mathcal{B}_0$, $\{\epsilon_t > 0 : t \in [T]\}$
> **Output:** $a_t \in A$ for each $t \in [T]$
> **for** $t = 0$ to $T - 1$ **do**
>   Choose any $\sigma[h_t] \in \arg\max_\sigma\{V(\gamma_t, \sigma) : \sigma \in \mathsf{Pers}(\mathcal{B}_t)\}$;
>   Recommend $a_t = a \in A$ with probability $\sigma(\omega_t, a; h_t)$;
>   Update $\gamma_{t+1}(\omega) \leftarrow \frac{1}{t+1}\sum_{\tau=0}^{t} \mathbf{I}\{\omega_\tau = \omega\}$ for each $\omega \in \Omega$;
>   Set $\mathcal{B}_{t+1} \leftarrow B_1(\gamma_{t+1}, \epsilon_{t+1})$;
> **end**

From the intuitive description, it follows that the sets $\mathcal{B}_t = B_1(\gamma_t, \epsilon_t)$ naturally play the role of the covering sets $\mathcal{C}_t$ in the definition of β-robustly persuasiveness. Specifically, the parameters $\{\epsilon_t : t \in [T]\}$ control the degree of persuasiveness of the algorithm: larger values of ϵ_t imply that the algorithm is β-robustly persuasive for smaller values of β. (In particular, if all ϵ_t are larger than two, the algorithm reduces to the full-information algorithm $\mathfrak{Full}$ and is 0-robustly persuasive.) Unsurprisingly, larger values of ϵ_t also lead to larger regret, and, hence, the sender must choose ϵ_t to optimally trade off the persuasiveness of the algorithm against its regret.

Our first main result characterizes $\mathfrak{Rai}$'s persuasiveness for a particular choice of parameter values, which we show in Section 5 to be regret-optimal.

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

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

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

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

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

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

Proposition 1. *The $\mathfrak{Rai}$ algorithm requires solving at each time a linear program with size polynomial in $|\Omega|$ and $|A|$.*

Proof. To see the efficiency of the $\mathfrak{Rai}$ algorithm, note that at each time t, the algorithm has to solve the optimization problem $\max_\sigma\{V(\gamma_t, \sigma) : \sigma \in \mathsf{Pers}(\mathcal{B}_t)\}$. Because $\mathcal{B}_t = B_1(\gamma_t, \epsilon_t)$ is an ℓ_1-ball of radius ϵ_t, it is a convex polyhedron with at most $|\Omega| \cdot (|\Omega| - 1)$ vertices. (These vertices are all of the form $\gamma_t + \frac{\epsilon_t}{2}(e_\omega - e_{\omega'})$, where e_ω is the belief that puts all its weight on ω.) By the linearity of the obedience constraints and the convexity of $\mathcal{B}_t$, it follows that $\mathsf{Pers}(\mathcal{B}_t)$ is obtained by imposing the obedience constraints at distributions corresponding to each of these vertices. Because there are $O(|\Omega| + |A|^2)$ obedience constraints for each distribution, we obtain that the optimization problem is a polynomially sized linear program and, hence, can be solved efficiently. □

Having addressed the persuasiveness and the computational efficiency of the $\mathfrak{Rai}$ algorithm, we devote the rest of the paper to analyzing its regret. To do this, we first take a digression to define (and bound) the cost of robust persuasion in static persuasion problems. Armed with this result, we then characterize the algorithm's regret in Section 5.

4. Digression: Cost of Robust Persuasion

In this section, we consider the static persuasion problem with known state distribution (discussed in Section 2.1) and study the loss in the sender's expected utility from requiring the signaling mechanism to be persuasive for all distributions in a neighborhood around the state distribution. To measure this loss, we first define the notion of the *cost of robust persuasion*, a quantity that depends on the neighborhood, and provide upper and lower bounds under some minor regularity conditions.

Fix a persuasion instance $\mathcal{I}$. In the static setting with known state distribution μ, the sender's optimal expected

payoff is given by $\mathsf{OPT}_\mathcal{I}(\mu) = \sup_{\sigma \in \mathsf{Pers}(\mu)} V(\mu, \sigma)$. Next, for any set of distributions $\mathcal{B} \subseteq \mathcal{B}_0$, the set of signaling mechanisms that are simultaneously persuasive for all distributions in $\mathcal{B}$ is given by $\mathsf{Pers}(\mathcal{B}) = \cap_{\mu' \in \mathcal{B}} \mathsf{Pers}(\mu')$. Hence, the sender's optimal expected utility among all such mechanisms is given by $\sup_{\sigma \in \mathsf{Pers}(\mathcal{B})} V(\mu, \sigma)$. Thus, we define the *cost of robust persuasion* as

$$\mathsf{Gap}(\mu, \mathcal{B}) \triangleq \sup_{\sigma \in \mathsf{Pers}(\mu)} V(\mu, \sigma) - \sup_{\sigma \in \mathsf{Pers}(\mathcal{B})} V(\mu, \sigma). \quad (4)$$

Thus, $\mathsf{Gap}(\mu, \mathcal{B})$ captures the difference in the sender's expected utility (under μ) between using the optimal persuasive signaling mechanism for the distribution μ and using the optimal signaling mechanism that is persuasive for all distributions $\mu' \in \mathcal{B}$.

For general persuasion instances, one can show that the cost of robust persuasion can be severe: in Appendix A1 of the e-companion, we present a persuasion instance and a distribution μ such that for any $\epsilon > 0$, the cost of being robustly persuasive for the set $\mathsf{B}_1(\mu, \epsilon)$ of distributions satisfies $\mathsf{Gap}(\mu, \mathsf{B}_1(\mu, \epsilon)) = \frac{1}{2}$. The instance we present there is pathological, with an action that is optimal for the receiver at a single unique distribution. To obtain meaningful insights on the cost of robust persuasion, we seek to exclude such instances by imposing some regularity condition on the instances.

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

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

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

We consider the following regularity conditions on the persuasion instances:

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

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

2. *There exists a $p_0 > 0$ such that for all $\mu \in \mathcal{B}_0$, we have $\min_\omega \mu(\omega) \geq p_0 > 0$.*

The first condition requires that each such set $\mathcal{P}_a$ has a nonempty relative interior; this excludes the pathological instances like that in the e-companion (Appendix A1), for which there exists an action a with $\mathcal{P}_a$ a singleton. We note that this condition is analogous to the Slater condition in convex optimization, imposing

nonempty interior on the feasibility region to obtain strong duality. The second condition is technical and is made primarily to ensure the potency of the first condition: without it, the sets $\{\mathcal{P}_a\}_{a \in A}$ may satisfy the first condition in $\Delta(\Omega)$, while failing to satisfy it relative to the subset $\Delta(\{\omega : \mu(\omega) > 0\})$ for some $\mu \in \mathcal{B}_0$. Taken together, these regularity conditions serve to avoid pathologies, and, henceforth, we restrict our attention only to those instances satisfying these regularity conditions.

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

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

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

Next, we provide a (worst-case) lower bound on Gap. We accomplish this by carefully constructing a persuasion instance $\mathcal{I}_0$, where being robustly persuasive leads to a substantial loss to the sender. The instance $\mathcal{I}_0$ has three states, $\Omega = \{\omega_0, \omega_1, \omega_2\}$, and five actions, $A = \{a_0, a_1, a_2, a_3, a_4\}$, for the receiver. At a high level, the receiver's preference can be illustrated as in Figure 1(a), which depicts the receiver's optimal action for any belief in the simplex. The regions $\mathcal{P}_i$ in the figure correspond to the set of beliefs that induce action $a_i \in A$ as the receiver's best response. The instance is crafted so that the sets $\mathcal{P}_1$ and $\mathcal{P}_2$ that induce actions a_1 and a_2, respectively, are symmetric and extremely narrow with the width controlled by an ℓ_1-ball of radius D contained within, as depicted in Figure 1(b). (Because $|\Omega| = 3$, the ℓ_1-ball here is a hexagon.) For completeness, the receiver's utility is listed explicitly in Table 1. The sender seeks to persuade the receiver into choosing one of actions a_1 and a_2 (regardless of the state); all other actions are strictly worse for the sender. Formally, we set $v(\omega, a) = 1$ if $a \in \{a_1, a_2\}$ and 0 otherwise, for all ω.

Figure 1. (Color online) The Persuasion Instance $\mathcal{I}_0$
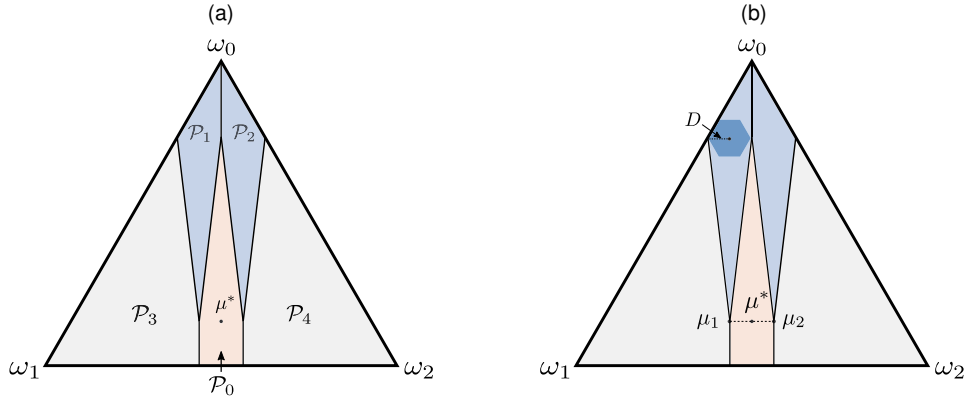
Notes. (a) Receiver's preferences. (b) Prior μ^*.

The sender's initial knowledge regarding the state distribution is captured by the set $\mathcal{B}_0 = \{\mu \in \Delta(\Omega) : \min_\omega \mu \geq p_0\}$, while the distribution of interest is $\mu^* = (p_0, \frac{1-p_0}{2}, \frac{1-p_0}{2})$, corresponding to the midpoint of the tips of the sets $\mathcal{P}_i$, as shown in Figure 1(b). We focus on the setting where the instance parameters D and p_0 satisfy $Dp_0 < 1/64$. The following proposition shows that in the instance $\mathcal{I}_0$, it is costly to require the signaling mechanism to be robustly persuasive for a set of distributions around μ^*. The result also implies that the bound on $\mathsf{Gap}(\cdot)$ obtained in Proposition 2 is almost tight, except for a factor of $1/p_0$.

Proposition 3. *For the instance $\mathcal{I}_0$, we have $\mathsf{OPT}(\mu^*) = 1$. Furthermore, for all $\epsilon \in (0, D)$, we have*

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

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

We defer the rigorous algebraic proof of the lower bound to Appendix C of the e-companion and present a brief sketch using a geometric argument here. In the instance $\mathcal{I}_0$, the distribution μ^* can be written as a convex combination $\mu^* = (\mu_1 + \mu_2)/2$, where μ_1 and μ_2 are the tips of regions $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively (see Figure 1(b)). Thus, by the splitting lemma (Aumann et al. 1995), it follows that the optimal signaling mechanism sends signals that induce posterior beliefs μ_1 and μ_2, leading to the receiver's choice of a_1 and a_2, respectively. Because the sender can always persuade the receiver to choose one of her preferred actions, we

obtain $\mathsf{OPT}(\mu^*) = 1$. On the other hand, for a signaling mechanism to be robustly persuasive for all distributions ϵ-close to the distribution μ^* for sufficiently small ϵ, the posteriors for the sender's preferred actions a_1, a_2 induced by the signaling mechanism have to be shifted up significantly in the narrow region. Such a large discrepancy ultimately forces the sender to suffer a substantial loss in the expected payoff.

5. Regret Analysis

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

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

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

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

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

Table 1. Receiver's Utility in Instance $\mathcal{I}_0$, with $u(\omega, a_0)$ Normalized to Zero for All $\omega \in \Omega$

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

The central step in the proof is the following decomposition of the regret, established in Lemma EC2 in Appendix D1 of the e-companion:

$$\text{Reg}_{\mathcal{I}}(\mathfrak{Rai}, \mu^*, T) \leq \sum_{t \in [T]} \text{Gap}(\mu^*, \mathsf{B}_1(\mu^*, \|\mu^* - \gamma_t\|_1))$$

$$+ \sum_{t \in [T]} \text{Gap}(\gamma_t, \mathsf{B}_1(\gamma_t, \epsilon_t))$$

$$+ \sum_{t \in [T]} \|\mu^* - \gamma_t\|_1$$

$$+ \sum_{t \in [T]} (\mathbf{E}_{\mu^*}[v(\omega_t, a_t)|h_t] - v(\omega_t, a_t)).$$

Observe that on the event $\{\mu^* \in \cap_{t \in [T]} \mathcal{B}_t\}$, we have $\|\mu^* - \gamma_t\|_1 \leq \epsilon_t$. Thus, on this event, the first two terms on the right-hand side of the preceding inequality capture the cost of persuading robustly for all distributions in an ℓ_1-ball of radius ϵ_t around the distribution μ^* and its estimate γ_t. Moreover, the third term represents the estimation error between μ^* and γ_t. Together with Proposition 2, we thus obtain that the first three terms are of order $\sum_{t \in [T]} \epsilon_t = O(\sqrt{T \log T})$. Finally, the last term, which captures the randomness in the sender's payoff, is also of the same order due to a simple application of the Azuma-Hoeffding inequality. The details are provided in Appendix D1 of the e-companion.

5.1. Lower Bound

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

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

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

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

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

$$+ \sum_{t \in [T]} V(\mu^*, \sigma^{\mathfrak{a}}[h_t]) - \sum_{t \in [T]} v(\omega_t, a_t).$$

Let $\mathcal{E}_T(\mu)$ be the event under which the signaling mechanism $\sigma^{\mathfrak{a}}[h_t]$ chosen by the algorithm $\mathfrak{a}$ after any history

$h_t \in \mathcal{E}_T(\mu)$ is persuasive for the distribution μ. Hence, on the event $\mathcal{E}_T(\mu^*) \cap \mathcal{E}_T(\overline{\mu}_1) \cap \mathcal{E}_T(\overline{\mu}_2)$, the signaling mechanism $\sigma^{\mathfrak{a}}[h_t]$ is persuasive for all three distributions $\mu^*, \overline{\mu}_1 = \mu^* + \frac{\epsilon}{2}(e_1 - e_2)$ and $\overline{\mu}_2 = \mu^* + \frac{\epsilon}{2}(e_2 - e_1)$. From Proposition 3, we have that on this event, the first term, which is the sender's expected loss, is no less than $T \cdot \text{Gap}(\mu^*, \{\mu^*, \overline{\mu}_1, \overline{\mu}_2\})$. We lower-bound the second term using the Azuma-Hoeffding inequality. The remaining step is to show that the probability of the event $\mathcal{E}_T(\mu^*) \cap \mathcal{E}_T(\overline{\mu}_1) \cap \mathcal{E}_T(\overline{\mu}_2)$ does not vanish as T goes to infinity, which follows from robust persuasiveness of the algorithm $\mathfrak{a}$ and careful choice of ϵ. The details are provided in Appendix D2 of the e-companion.

6. Conclusion

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

Although social learning is a strong motivation for our robust persuasiveness criterion, there are other motivations as well. For instance, a platform concerned about its long-run reputation may want to design a recommendation algorithm that guarantees verifiably good-quality recommendations, not just with respect to currently available state realization data, but also with respect to any additional data obtained in the future. An algorithm satisfying our robust persuasiveness criterion enables such a platform to meet its goals.

Although in our analysis, we have assumed that the receiver's utility is fixed across time periods, our model and the analysis can be easily extended to accommodate heterogeneous receivers, as long as the sender observes the receiver's type prior to making the recommendation and the cost of robustness Gap can be uniformly bounded across different receiver types. More interesting is the setting where the sender must persuade a receiver with an unknown type. In such a setting, assuming that the sender cannot elicit the receiver's type prior to making the recommendation, the sender makes a menu of action recommendations (one for each receiver type). It can be shown that the complete information problem in this setting corresponds to public persuasion of a group of receivers with no externality, which is known to be a computationally hard linear program with exponentially many

constraints (Dughmi and Xu 2017). Consequently, our algorithm ceases to be computationally efficient. Nevertheless, our results imply that the algorithm continues to maintain the $O(\sqrt{T \log T})$ regret bound.

Our characterization of the cost of robust persuasion may be of independent interest. For instance, one can derive the sample complexity bounds for the static persuasion problem when the sender only has access to the samples from the underlying distribution. To obtain a signaling mechanism that is persuasive with probability at least $1 - \beta$ and is ϵ-optimal, our characterization yields a sample complexity of $\Theta\left(\frac{|\Omega| + \log(1/\beta)}{p_0^4 D^2 \epsilon^2}\right)$. Note that for large enough ϵ, one can simply use the full-information mechanism with no need for any samples.

Our analysis highlights two main technical contributions. One is the characterization of the cost of robust persuasion for the underlying linear program and using this characterization to perform a tight regret analysis for the online learning problem. The former result heavily uses the specifics of the persuasion problem (for instance, the use of the splitting lemma to construct a feasible robust solution), whereas the latter result is more agnostic to the setting. Given this, we believe our approach can be extended to other online linear programming settings, as long as one can obtain a characterization of the corresponding cost of robustness. Note that even in our persuasion setting, we had to impose the regularity conditions to obtain the linear bounds on the cost of robustness, without which the cost could be $O(1)$ and the regret would be linear. Whether these regularity conditions can be generalized to other linear settings is an interesting question for further investigation.

Acknowledgments

References

Agrawal S, Devanur NR (2014) Bandits with concave rewards and convex knapsacks. *Proc. Fifteenth ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 989–1006.

Amani S, Alizadeh M, Thrampoulidis C (2019) Linear stochastic bandits under safety constraints. *Adv. Neural Inform. Processing Systems* 32:9256–9266.

Antelmi A, Malandrino D, Scarano V (2019) Characterizing the behavioral evolution of Twitter users and the truth behind the 90-9-1 rule. *Companion Proc. 2019 World Wide Web Conf.* (Association for Computing Machinery, New York), 1035–1038.

Aumann RJ, Maschler M, Stearns RE (1995) *Repeated Games with Incomplete Information* (MIT Press, Cambridge, MA).

Balcan MF, Blum A, Haghtalab N, Procaccia AD (2015) Commitment without regrets: Online learning in Stackelberg security games. *Proc. Sixteenth ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 61–78.

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

Bergemann D, Morris S (2019) Information design: A unified perspective. *J. Econom. Lit.* 57(1):44–95.

Besson L, Kaufmann E (2018) What doubling tricks can and can't do for multi-armed bandits. Preprint, submitted March 19, https://arxiv.org/abs/1803.06971.

Camara MK, Hartline JD, Johnsen A (2020) Mechanisms for a no-regret agent: Beyond the common prior. *2020 IEEE 61st Annual Sympos. Foundations of Computer Science FOCS*, volume 1 (IEEE, Piscataway, NJ), 259–270.

Candogan O (2020) Information design in operations. *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*, INFORMS TutORials in Operations Research (INFORMS, Catonsville, MD), 176–201.

Cao X, Liu KR (2018) Online convex optimization with time-varying constraints and bandit feedback. *IEEE Trans. Automatic Control* 64(7):2665–2680.

Cao X, Zhang J, Poor HV (2019) On the time-varying distributions of online stochastic optimization. *2019 Amer. Control Conf. ACC* (IEEE, Piscataway, NJ), 1494–1500.

Castiglioni M, Celli A, Marchesi A, Gatti N (2020) Online Bayesian persuasion. *Adv. Neural Inform. Processing Systems* 33:16188–16198.

Chawla S, Hartline JD, Malec D, Sivan B (2013) Prior-independent mechanisms for scheduling. *Proc. Forty Fifth Annu. ACM Sympos. Theory Comput.* (Association for Computing Machinery, New York), 51–60.

Chen Y, Liu Y, Podimata C (2020) Learning strategy-aware linear classifiers. Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, eds. *Advances in Neural Information Processing Systems*, vol. 33 (Curran Associates, Inc., Red Hook, NY), 15265–15276.

Dhangwatnotai P, Roughgarden T, Yan Q (2015) Revenue maximization with a single sample. *Games Econom. Behav.* 91:318–333.

Dong J, Roth A, Schutzman Z, Waggoner B, Wu ZS (2018) Strategic classification from revealed preferences. *Proc. 2018 ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 55–70.

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

Dughmi S, Xu H (2017) Algorithmic persuasion with no externalities. *Proc. 2017 ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 351–368.

Dughmi S, Xu H (2021) Algorithmic Bayesian persuasion. *SIAM J. Comput.* 50(3):STOC16-68–STOC16-97.

Dworczak P, Pavan A (2020) Preparing for the worst but hoping for the best: Robust (Bayesian) persuasion. CEPR Discussion Paper No. DP15017, Centre for Economic Policy Research, London.

Elliott M, Galeotti A, Koh A, Li W (2022) Matching and information design in marketplaces. Preprint, submitted December 13, https://dx.doi.org/10.2139/ssrn.4283968.

Foucart S, Rauhut H (2013) *A Mathematical Introduction to Compressive Sensing* (Birkhäuser, New York).

Guo H, Liu X, Wei H, Ying L (2024) Online convex optimization with hard constraints: Toward the best of two worlds and beyond. *Proc. 36th Internat. Conf. Adv. Neural Inform. Processing Systems.* (Curran Associates Inc., Red Hook, NY), 36426–36439.

Gur Y, Macnamara G, Morgenstern I, Saban D (2023) Information disclosure and promotion policy design for platforms. *Management Sci.* 69(10):5883–5903.

Hahn N, Hoefer M, Smorodinsky R (2020) Prophet inequalities for Bayesian persuasion. Bessiere C, ed. *Proc. Twenty-Ninth Internat. Joint Conf. Artificial Intelligence (IJCAI-20)* (International Joint Conferences on Artificial Intelligence Organization), 175–181.

Hahn N, Hoefer M, Smorodinsky R (2022) The secretary recommendation problem. *Games Econom. Behav.* 134:199–228.

Haldane JBS (1948) The precision of observed values of small frequencies. *Biometrika* 35(3/4):297–300.

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

Jaynes ET (2003) *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, UK).

Kamenica E, Gentzkow M (2011) Bayesian persuasion. *Amer. Econom. Rev.* 101(6):2590–2615.

Khezeli K, Bitar E (2019) Safe linear stochastic bandits. Preprint, submitted November 21, https://arxiv.org/abs/1911.09501.

Kim Y, Lee D (2023) Online convex optimization with stochastic constraints: Zero constraint violation and bandit feedback. Preprint, submitted January 26, https://arxiv.org/abs/2301.11267.

Kleinberg R, Leighton T (2003) The value of knowing a demand curve: Bounds on regret for online posted-price auctions. *Proc. 44th Annu. IEEE Sympos. Foundations Comput. Sci. FOCS'03* (IEEE Computer Society, Piscataway, NJ), 594–605.

Kosterina S (2022) Persuasion with unknown beliefs. *Theoret. Econom.* 17:1075–1107.

Kremer I, Mansour Y, Perry M (2014) Implementing the "wisdom of the crowd". *J. Polit. Econom.* 122(5):988–1012.

Lattimore T, Szepesvári C (2020) *Bandit Algorithms* (Cambridge University Press, Cambridge, UK).

Liakopoulos N, Destounis A, Paschos G, Spyropoulos T, Mertikopoulos P (2019) Cautious regret minimization: Online optimization with long-term budget constraints. Kamalika C, Ruslan S, eds. *Internat. Conf. Machine Learn.*, vol. 97 (PMLR, New York), 3944–3952.

Mahdavi M, Jin R, Yang T (2011) Trading regret for efficiency: Online convex optimization with long term constraints. Preprint, submitted November 25, https://arxiv.org/abs/1111.6082.

Mahdavi M, Yang T, Jin R (2013) Stochastic convex optimization with multiple objectives. Burges CJ, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, eds. *Advances in Neural Information Processing Systems*, vol. 26 (Curran Associates, Inc., Red Hook, NY).

Mansour Y, Slivkins A, Syrgkanis V (2015) Bayesian incentive-compatible bandit exploration. *Proc. Sixteenth ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 565–582.

Mansour Y, Slivkins A, Syrgkanis V, Wu ZS (2016) Bayesian exploration: Incentivizing exploration in Bayesian games. *Proc. 2016 ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 661.

Moradipari A, Thrampoulidis C, Alizadeh M (2020) Stage-wise conservative linear bandits. *Adv. Neural Inform. Processing Systems* 33:11191–11201.

Moradipari A, Amani S, Alizadeh M, Thrampoulidis C (2021) Safe linear Thompson sampling with side information. *IEEE Trans. Signal Process.* 69:3755–3767.

Neely MJ, Yu H (2017) Online convex optimization with time-varying constraints. Preprint, submitted February 15, https://arxiv.org/abs/1702.04783.

Pacchiano A, Ghavamzadeh M, Bartlett P, Jiang H (2021) Stochastic bandits with linear constraints. Arindam B, Kenji F, eds. *Internat. Conf. Artificial Intelligence Statist.*, vol. 30 (PMLR, New York), 2827–2835.

Romanyuk G, Smolin A (2019) Cream skimming and information design in matching markets. *Amer. Econom. J. Microeconom.* 11(2): 250–276.

Usmanova I, Krause A, Kamgarpour M (2019) Safe convex learning under uncertain constraints. Kamalika C, Masashi S, eds. *22nd Internat. Conf. Artificial Intelligence Statist.*, vol. 89 (PMLR, New York), 2106–2114.

van Mierlo T (2014) The 1% rule in four digital health social networks: An observational study. *J. Med. Internet Res.* 16(2):e2966.

Villegas C (1977) On the representation of ignorance. *J. Amer. Statist. Assoc.* 72(359):651–654.

Yang P, Iyer K, Frazier P (2019) Information design in spatial resource competition. Preprint, submitted September 29, https://arxiv.org/abs/1909.12723.

Yi X, Li X, Yang T, Xie L, Chai T, Johansson K (2021) Regret and cumulative constraint violation analysis for online convex optimization with long term constraints. Marina M, Tong Z, eds. *Internat. Conf. Machine Learn.*, vol. 139 (PMLR, New York), 11998–12008.

Yi X, Li X, Yang T, Xie L, Chai T, Johansson KH (2023) Regret and cumulative constraint violation analysis for distributed online constrained convex optimization. *IEEE Trans. Automatic Control* 68(5):2875–2890.

Yu H, Neely MJ, Wei X (2017) Online convex optimization with stochastic constraints. Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. *Adv. Neural Inform. Processing Systems*, vol. 30 (Curran Associates, Inc., Red Hook, NY).

Yuan J, Lamperski AG (2018) Online convex optimization for cumulative constraints. Preprint, submitted February 19, https://arxiv.org/abs/1802.06472.

You Zu is a PhD candidate in the Department of Industrial and Systems Engineering at the University of Minnesota. Her research focuses on information design.

Krishnamurthy Iyer is an associate professor in the Department of Industrial and Systems Engineering at the University of Minnesota. His research interests include game theory, mechanism design, and stochastic modeling, with applications in markets and service systems.

Haifeng Xu is an assistant professor in the Department of Computer Science at the University of Chicago. He studies the economics of data and machine learning, including designing learning algorithms for multiagent decision making and designing markets for data and ML algorithms. His work has been awarded an AI2050 Early Career Fellow, an IJCAI Early Career Spotlight, a Google Faculty Research Award, etc.