

# Learning to Persuade on the Fly: Robustness Against Ignorance

Online Appendix

You Zu, Krishnamurthy Iyer and Haifeng Xu

## Appendix A: Examples of Persuasion Instances

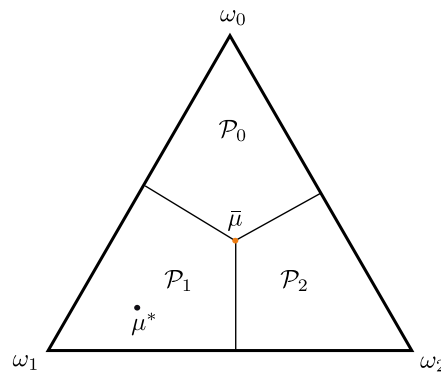
In this section, we provide examples of instances that illuminate various aspects of our theoretical results.

### A.1. Failure of the Regularity Condition

We begin with an example of an instance  $\mathcal{I}_1$  in which the regularity condition does not hold, and in which any  $\beta$ -robustly persuasive algorithm incurs a linear regret. We establish this by proving that in this instance, the cost of robust persuasion  $\text{Gap}(\mu^*, B_1(\mu^*, \epsilon))$  is a constant independent of  $\epsilon$  for all  $\epsilon > 0$ .

In the persuasion instance  $\mathcal{I}_1$ , the state space is given by  $\Omega = \{\omega_0, \omega_1, \omega_2\}$  and the receiver has four actions  $A = \{a_0, a_1, a_2, a_3\}$ . The receiver's utility is given by  $u(\omega_i, a_j) = \mathbf{I}\{i = j\} + \frac{1}{3}\mathbf{I}\{j = 3\}$  for  $i \in \{0, 1, 2\}$  and  $j \in \{0, 1, 2, 3\}$ . The sender's payoff is given by  $v(\omega_i, a_j) = \mathbf{I}\{j = 3\}$ ; in other words, the sender strictly prefers the receiver choosing action  $a_3$  over any other action in all states. The sender's initial knowledge regarding the underlying state distribution is captured by  $\mathcal{B}_0 = B_1(\mu^*, \epsilon_0)$  for some  $\epsilon_0 > 0$ , where  $\mu^* = (\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$ .

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



**Figure EC.1** The persuasion instance  $\mathcal{I}_1$ .

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

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

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

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

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

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

## A.2. Linear regret for 0-robustly persuasive mechanisms

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

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

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

Now, since all the distributions in  $\mathcal{B}_0$  are absolutely continuous with respect to each other, any algorithm  $\mathbf{a}$  that is 0-robustly persuasive must select at each time  $t \in [T]$  a signaling mechanism  $\sigma_t$  in the set  $\text{Pers}(\mathcal{B}_0)$ . However, it is straightforward to verify that among all mechanisms that are persuasive for all distributions

in  $\mathcal{B}_0$ , the one that maximizes sender's payoff is given by  $\hat{\sigma}(\omega_0, a_0) = 1 - \hat{\sigma}(\omega_0, a_1) = 1, \hat{\sigma}(\omega_1, a_0) = 1 - \hat{\sigma}(\omega_1, a_1) = \frac{1}{3}$ . For the distribution  $\mu^* = (\frac{1}{2}, \frac{1}{2}) \in \mathcal{P}_0 \cap \mathcal{B}_0$ , it follows that the sender's payoff under  $\hat{\sigma}$  is  $V(\mu^*, \hat{\sigma}) = \frac{2}{3}$ .

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

## Appendix B: Proofs from Section 3

This section provides the proof of Theorem 1, along with a helper lemma establishing the concentration of the empirical distribution around the (unknown) distribution.

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

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

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

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

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

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

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

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

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

The following lemma provides a bound on the  $\ell_1$ -norm of the deviation of the empirical distribution from its mean.

**LEMMA EC.1.** *For each  $t \in [T]$ , and for any  $\mu \in \Delta(\Omega)$ , we have for all  $0 < \Phi_t \leq 2\sqrt{t}$ ,*

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

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

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

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

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

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

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

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

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

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

2.  $\square$

## Appendix C: Proofs from Section 4

This section provides the proofs of the propositions in Section 4. Throughout, we use the same notation as in the main text.

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

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

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

Let  $\eta_a \in \mathcal{P}_a$  be such that  $B_1(\eta_a, D) \subseteq \mathcal{P}_a$ . For  $\delta = \frac{2\epsilon}{p_0 D} \in [0, 1]$ , define  $\xi_a = (1 - \delta)\mu_a + \delta\eta_a \in \mathcal{P}_a$  for each  $a \in A_+$  and let  $\xi = \sum_{a \in A_+} w_a \xi_a$ . Furthermore, since  $\mu_a \in \mathcal{P}_a$  and  $B_1(\eta_a, D) \subseteq \mathcal{P}_a$ , the convexity of the set  $\mathcal{P}_a$  implies that  $B_1(\xi_a, \delta D) \subseteq \mathcal{P}_a$ .

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

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

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

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

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

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

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

We now show that the signaling mechanism  $\hat{\sigma}$  is persuasive for all distributions in  $B_1(\mu, \epsilon)$ , in the sense that for all signals  $s \in A_+$  it is optimal for the receiver to play  $s$ , and for all signals  $s = (\omega, a_\omega) \in S$ , it is optimal for the receiver to play  $a_\omega$ . To see this, for any  $\gamma \in B_1(\mu, \epsilon)$ , let  $\gamma(\cdot|s)$  denote the receiver's posterior under signaling mechanism  $\hat{\sigma}$  upon receiving the signal  $s \in A_+ \cup S$ . For  $s = (\omega, a_\omega) \in S$ , we have  $\gamma(\cdot|s) = e_\omega$ , where  $e_\omega$  is the belief that puts all its weight on  $\omega \in \Omega$ . Thus, upon receiving the signal  $s = (\omega, a_\omega)$  it is optimal for the receiver with the distribution  $\gamma$  to take action  $a_\omega$ . Thus, it only remains to show that signals  $s = a \in A_+$  are persuasive.

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

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

Then, using triangle inequality and some algebra, we obtain

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

where in the final inequality, we have used  $\min_{\omega} \mu(\omega) \geq p_0$  to get  $\sup_{\omega \in \Omega} \frac{\xi_a(\omega)}{\mu(\omega)} \leq \frac{1}{p_0}$ . Since  $\mu(\cdot|a) = \xi_a$ , this implies that  $\gamma(\cdot|a) \in B_1\left(\xi_a, \frac{2\epsilon}{p_0}\right) = B_1(\xi_a, \delta D) \subseteq \mathcal{P}_a$ . Thus, the signal  $a \in A_+$  is persuasive for the distribution  $\gamma \in B_1(\mu, \epsilon)$ . Taken together, we obtain that the signaling mechanism  $\hat{\sigma}$  is persuasive for all  $\gamma \in B_1(\mu, \epsilon)$ .

The persuasiveness of  $\hat{\sigma}$  for all  $\gamma \in B_1(\mu, \epsilon)$  implies that

$$\begin{aligned}\sup_{\sigma' \in \text{Pers}(B_1(\mu, \epsilon))} V(\mu, \sigma') &\geq V(\mu, \hat{\sigma}) \\ &= \sum_{\omega \in \Omega} \sum_{a \in A_+} \mu(\omega)\hat{\sigma}(\omega, a)v(\omega, a) + \sum_{\omega \in \Omega} \sum_{s \in S} \mu(\omega)\hat{\sigma}(\omega, s)v(\omega, a_\omega) \\ &\geq \sum_{\omega \in \Omega} \sum_{a \in A_+} \bar{\rho} w_a \xi_a(\omega) v(\omega, a) \\ &= \bar{\rho} \sum_{\omega \in \Omega} \sum_{a \in A_+} w_a ((1 - \delta)\mu_a(\omega) + \delta\eta_a(\omega)) v(\omega, a) \\ &\geq \bar{\rho}(1 - \delta) \sum_{\omega \in \Omega} \sum_{a \in A_+} w_a \mu_a(\omega) v(\omega, a) \\ &= \bar{\rho}(1 - \delta) \text{OPT}(\mu).\end{aligned}$$

Thus, we obtain

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

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

*Proof of Proposition 3.* It is straightforward to verify that the following signaling mechanism  $\sigma^* \in \text{Pers}(\mu^*)$  optimizes the sender's expected utility among all mechanisms in  $\text{Pers}(\mu^*)$ :

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

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

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

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

where

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

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

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

## Appendix D: Proofs from Section 5

### D.1. Proof of Theorem 2

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

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

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

Now, from Proposition 2, we have

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

Thus, we obtain

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

Finally, in Lemma EC.5, we show that on the event  $\{\mu^* \in \mathcal{B}_t\}$ , we have  $\text{Gap}(\gamma_t, \mathcal{B}_t) \leq \left( \frac{16}{p_0^2 D} \right) \epsilon_t$ . Thus, on the event  $\{\mu^* \in \cap_{t \in [T]} \mathcal{B}_t\}$ , we obtain

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

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

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

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

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

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

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

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

**LEMMA EC.2.** *The  $\mathfrak{A}i$  algorithm satisfies*

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

*Proof.* From the definition (3) of regret, we have

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

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

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

where in the final equality, we have used the fact that  $\text{OPT}(\gamma_t) - V(\gamma_t, \sigma[h_t]) = \text{Gap}(\gamma_t, \mathcal{B}_t)$ . Substituting the preceding expression into (EC.2) yields

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

Now, in Lemma EC.3, we prove  $\text{OPT}(\mu^*) - \text{OPT}(\gamma_t) \leq \text{Gap}(\mu^*, \mathbf{B}_1(\mu^*, \|\mu^* - \gamma_t\|_1)) + \frac{1}{2} \cdot \|\mu^* - \gamma_t\|_1$ . Furthermore, in Lemma EC.4, we show that  $V(\gamma_t, \sigma[h_t]) - V(\mu^*, \sigma[h_t]) \leq \frac{1}{2} \|\mu^* - \gamma_t\|_1$ . Putting it all together yields the lemma statement.  $\square$

**LEMMA EC.3.** *For any  $\mu_1, \mu_2 \in \Delta(\Omega)$ , we have*

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

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

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

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

Here, the inequality follows from the definition of  $\text{Gap}(\cdot)$ , and from Lemma EC.4.  $\square$

**LEMMA EC.4.** *For any  $\mu_1, \mu_2 \in \Delta(\Omega)$  and any signaling mechanism  $\sigma$ , we have*

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

*Proof.* Fix  $\mu_1, \mu_2 \in \Delta(\Omega)$ . For any signaling mechanism  $\sigma$  that is persuasive under  $\mu_1$ , we have for any  $x \in \mathbb{R}$ ,

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

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

LEMMA EC.5. For  $t \in [T]$ , on the event  $\{\mu^* \in \mathcal{B}_t\}$ , we have

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

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

$$\begin{aligned} \gamma_t(\omega) &\geq \mu(\omega) - \|\gamma_t - \mu^*\|_1 \\ &\geq p_0 - \epsilon_t. \end{aligned}$$

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

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

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

## D.2. Proof of Theorem 3

We conclude this section with the proof of the lower bound in Theorem 3.

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

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

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

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

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

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

$$T \cdot \text{OPT}(\mu^*) - \sum_{t \in [T]} V(\mu^*, \sigma^a[h_t]) \geq T \cdot \text{Gap}(\mu^*, \{\mu^*, \bar{\mu}_1, \bar{\mu}_2\}) \geq \frac{\epsilon T}{8Dp_0},$$

where the first inequality follows from the definition of Gap in (4), and the second inequality follows from Proposition 3.

Now, we have

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

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

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

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

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

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

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

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

Finally, by the Azuma-Hoeffding inequality, we obtain

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

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

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

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

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

for  $Dp_0 < 1/32$ .  $\square$

## References

- Agrawal S, Devanur NR (2014) Bandits with concave rewards and convex knapsacks. *Proceeding of the fifteenth ACM conference on Economics and computation*, 989–1006.
- Amani S, Alizadeh M, Thrampoulidis C (2019) Linear stochastic bandits under safety constraints. *Advances in Neural Information Processing Systems* 32.
- Antelmi A, Malandrino D, Scarano V (2019) Characterizing the behavioral evolution of twitter users and the truth behind the 90-9-1 rule. *Companion Proceedings of The 2019 World Wide Web Conference*, 1035–1038.
- Anunrojwong J, Iyer K, Lingenbrink D (2020) Persuading risk-conscious agents: A geometric approach. *Available at SSRN 3386273* .
- Aumann RJ, Maschler M, Stearns RE (1995) *Repeated games with incomplete information* (MIT press).
- Balcan MF, Blum A, Haghtalab N, Procaccia AD (2015) Commitment without regrets: Online learning in stackelberg security games. *Proceedings of the sixteenth ACM conference on economics and computation*, 61–78.
- Bergemann D, Morris S (2016) Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics* 11(2):487–522.
- Bergemann D, Morris S (2019) Information design: A unified perspective. *Journal of Economic Literature* 57(1):44–95.
- Besson L, Kaufmann E (2018) What doubling tricks can and can't do for multi-armed bandits. *arXiv preprint arXiv:1803.06971* .
- Boucheron S, Lugosi G, Massart P (2013) *Concentration inequalities: A nonasymptotic theory of independence* (Oxford university press).
- Camara MK, Hartline JD, Johnsen A (2020) Mechanisms for a no-regret agent: Beyond the common prior. *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, volume 1, 259–270 (Los Alamitos, CA, USA: IEEE Computer Society), URL <http://dx.doi.org/10.1109/FOCS46700.2020.00033>.
- Candogan O (2020) Information design in operations. *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*, 176–201 (INFORMS).
- Cao X, Liu KR (2018) Online convex optimization with time-varying constraints and bandit feedback. *IEEE Transactions on automatic control* 64(7):2665–2680.

- Cao X, Zhang J, Poor HV (2019) On the time-varying distributions of online stochastic optimization. *2019 American Control Conference (ACC)*, 1494–1500 (IEEE).
- Castiglioni M, Celli A, Marchesi A, Gatti N (2020) Online bayesian persuasion. *Advances in Neural Information Processing Systems* 33.
- Chawla S, Hartline JD, Malec D, Sivan B (2013) Prior-independent mechanisms for scheduling. *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, 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*, volume 33, 15265–15276 (Curran Associates, Inc.), URL <https://proceedings.neurips.cc/paper/2020/file/ae87a54e183c075c494c4d397d126a66-Paper.pdf>.
- Dhangwatnotai P, Roughgarden T, Yan Q (2015) Revenue maximization with a single sample. *Games and Economic Behavior* 91:318–333.
- Dong J, Roth A, Schutzman Z, Waggoner B, Wu ZS (2018) Strategic classification from revealed preferences. *Proceedings of the 2018 ACM Conference on Economics and Computation*, 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. *Proceedings of the 2017 ACM Conference on Economics and Computation*, 351–368.
- Dughmi S, Xu H (2021) Algorithmic bayesian persuasion. *SIAM Journal on Computing* 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*.
- Elliott M, Galeotti A, Koh A, Li W (2022) Matching and information design in marketplaces. *Available at SSRN*.
- Foucart S, Rauhut H (2013) *A Mathematical Introduction to Compressive Sensing* (Springer), URL <http://dx.doi.org/10.1007/978-0-8176-4948-7>.
- Guo H, Liu X, Wei H, Ying L (2022) Online convex optimization with hard constraints: Towards the best of two worlds and beyond. *Advances in Neural Information Processing Systems*.

- Gur Y, Macnamara G, Morgenstern I, Saban D (2023) Information disclosure and promotion policy design for platforms. *Management Science* .
- Hahn N, Hoefler M, Smorodinsky R (2019) The secretary recommendation problem. *arXiv preprint arXiv:1907.04252* .
- Hahn N, Hoefler M, Smorodinsky R (2020) Prophet inequalities for bayesian persuasion. *Proc. 29th Int. Joint Conf. Artif. Intell.(IJCAI)*, 175–181.
- Haldane JBS (1948) The precision of observed values of small frequencies. *Biometrika* 35(3/4):297–300, ISSN 00063444, URL <http://www.jstor.org/stable/2332350>.
- Hu J, Weng X (2020) Robust persuasion of a privately informed receiver. *Economic Theory* 1–45.
- Jaynes ET (2003) *Probability theory: The logic of science* (Cambridge university press).
- Kamenica E, Gentzkow M (2011) Bayesian persuasion. *American Economic Review* 101(6):2590–2615.
- Khezeli K, Bitar E (2020) Safe linear stochastic bandits. *ArXiv abs/1911.09501*.
- Kim Y, Lee D (2023) Online convex optimization with stochastic constraints: Zero constraint violation and bandit feedback. *arXiv preprint arXiv:2301.11267* .
- Kleinberg R, Leighton T (2003) The value of knowing a demand curve: Bounds on regret for online posted-price auctions. *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 594, FOCS '03 (USA: IEEE Computer Society), ISBN 0769520405.
- Kosterina S (2018) Persuasion with unknown beliefs. *Work. Pap., Princeton Univ., Princeton, NJ* .
- Kremer I, Mansour Y, Perry M (2014) Implementing the “wisdom of the crowd”. *Journal of Political Economy* 122(5):988–1012.
- Lattimore T, Szepesvári C (2020) *Bandit Algorithms* (Cambridge University Press), ISBN 9781108486828, URL <https://books.google.com/books?id=bbjpdwAAQBAJ>.
- Liakopoulos N, Destounis A, Paschos G, Spyropoulos T, Mertikopoulos P (2019) Cautious regret minimization: Online optimization with long-term budget constraints. *International Conference on Machine Learning*, 3944–3952 (PMLR).
- Mahdavi M, Jin R, Yang T (2011) Trading regret for efficiency: Online convex optimization with long term constraints. *CoRR abs/1111.6082*, URL <http://arxiv.org/abs/1111.6082>.

- Mahdavi M, Yang T, Jin R (2013) Stochastic convex optimization with multiple objectives. *Advances in neural information processing systems* 26.
- Mansour Y, Slivkins A, Syrgkanis V (2015) Bayesian incentive-compatible bandit exploration. *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 565–582.
- Mansour Y, Slivkins A, Syrgkanis V, Wu ZS (2016) Bayesian exploration: Incentivizing exploration in bayesian games. *Proceedings of the 2016 ACM Conference on Economics and Computation*, 661–661.
- Moradipari A, Amani S, Alizadeh M, Thrampoulidis C (2021) Safe linear thompson sampling with side information. *IEEE Transactions on Signal Processing* 69:3755–3767.
- Moradipari A, Thrampoulidis C, Alizadeh M (2020) Stage-wise conservative linear bandits. *Advances in neural information processing systems* 33:11191–11201.
- Neely MJ, Yu H (2017) Online convex optimization with time-varying constraints. *arXiv preprint arXiv:1702.04783* .
- Pacchiano A, Ghavamzadeh M, Bartlett P, Jiang H (2021) Stochastic bandits with linear constraints. *International Conference on Artificial Intelligence and Statistics*, 2827–2835 (PMLR).
- Romanyuk G, Smolin A (2019) Cream skimming and information design in matching markets. *American Economic Journal: Microeconomics* 11(2):250–76.
- Usmanova I, Krause A, Kamgarpour M (2019) Safe convex learning under uncertain constraints. *The 22nd International Conference on Artificial Intelligence and Statistics*, 2106–2114 (PMLR).
- Van Mierlo T, et al. (2014) The 1% rule in four digital health social networks: an observational study. *Journal of medical Internet research* 16(2):e2966.
- Villegas C (1977) On the representation of ignorance. *Journal of the American Statistical Association* 72(359):651–654, ISSN 01621459, URL <http://www.jstor.org/stable/2286233>.
- Yang P, Iyer K, Frazier P (2019) Information design in spatial resource competition. *arXiv preprint arXiv: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. *International Conference on Machine Learning*, 11998–12008 (PMLR).
- Yi X, Li X, Yang T, Xie L, Chai T, Karl H (2022) Regret and cumulative constraint violation analysis for distributed online constrained convex optimization. *IEEE Transactions on Automatic Control* .

Yu H, Neely MJ, Wei X (2017) Online convex optimization with stochastic constraints.

Yuan J, Lamperski AG (2018) Online convex optimization for cumulative constraints. *CoRR* abs/1802.06472, URL

<http://arxiv.org/abs/1802.06472>.