Online Mechanism Design for Information Acquisition

Federico Cacciamani [1] **Matteo Castiglioni** [1] **Nicola Gatti** [1]

Abstract

We study the problem of designing mechanisms for *information acquisition* scenarios. This setting models strategic interactions between an uninformed *receiver* and a set of informed *senders*. In our model the senders receive information about the underlying state of nature and communicate their observation (either truthfully or not) to the receiver, which, based on this information, selects an action. Our goal is to design mechanisms maximizing the receiver's utility while incentivizing the senders to report truthfully their information. First, we provide an algorithm that efficiently computes an optimal *incentive compatible* (IC) mechanism. Then, we focus on the *online* problem in which the receiver sequentially interacts in an unknown game, with the objective of minimizing the *cumulative regret* w.r.t. the optimal IC mechanism, and the *cumulative violation* of the incentive compatibility constraints. We investigate two different online scenarios, *i.e.,* the *full* and *bandit feedback* settings. For the full feedback problem, we propose an algorithm that guarantees $\tilde{\mathcal{O}}(\sqrt{T})$ regret and violation, while for the bandit feedback setting we present an algorithm that attains $\tilde{\mathcal{O}}(T^{\alpha})$ regret and $\tilde{\mathcal{O}}(T^{1-\alpha/2})$ violation for any $\alpha \in [1/2, 1]$. Finally, we complement our results providing a tight lower bound.

1. Introduction

Information plays a crucial role in every decision-making process. The study of how to exploit information to shape the behavior of other self-interested agents has received a terrific attention in recent years under the Bayesian persuasion framework, with application to online advertising (Bro Miltersen & Sheffet, 2012), voting (Alonso & Câmara, 2016; Castiglioni et al., 2020a; Castiglioni & Gatti, 2021), traffic routing (Bhaskar et al., 2016; Castiglioni et al., 2021a), and security (Rabinovich et al., 2015; Xu et al., 2016). Bayesian persuasion (Kamenica & Gentzkow, 2011) studies how an informed agent (sender) can exploit an information advantage to influence the behavior of an uninformed agent (receiver). In particular, the sender can commit to an information-disclosure policy (signaling scheme) in order to persuade the receiver to play an action that is desirable for the sender. In this work we study the inverse problem faced by a receiver that wants to incentivize multiple senders to share their private information about a common state of nature. Surprisingly, while many real-world scenarios are affected by an increasing decentralization of information, the problem of strategically acquiring such information has received little attention from the scientific community. In particular, we pose the question:

> *Can a receiver with commitment power induce senders to reveal their information truthfully?*

In our model, there are multiple senders that receive noisy information (signals) about a common state of nature. Then, each sender strategically reports a signal to the receiver, which chooses an action that depends on the signals reported by the senders. This action is drawn from a randomized mechanism to which the receiver commits and depends on the signals reported by the senders. Finally, the senders and the receiver receive an utility that depends on the state of nature and the action played by the receiver. In this work, we study the problem of designing an (approximately) incentive compatible mechanism that maximizes the sender's utility. Moreover, we consider an *online* scenario in which the receiver has partial or no information about the parameters of the game. Our goal is to design algorithms that repeatedly interact with the senders guaranteeing sublinear regret and violations of the incentive compatibility constraints.

1.1. Original Contribution

We study the design of Incentive Compatible (IC) mechanisms for the information acquisition problem, focusing on the case in which the senders are *symmetric*, *i.e.*, they receive signals sampled by a common signaling scheme. First, we show that, under a mild assumption, the optimal mechanism can be computed efficiently. In doing so, we provide

[1]Politecnico di Milano, Milan, Italy. Correspondence to: Federico Cacciamani <federico.cacciamani@polimi.it>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

a class of succinctly representable *symmetric* mechanisms. This is a non-trivial task since a general mechanism should specify a distribution over actions for each possible tuple of signals, which are exponentially-many. Then, we study the online problem in which the receiver does not know the parameters of the game and need to learn them during a repeated interaction with the senders. The uncertainty on the game parameters comes in two flavors: *full* and *bandit* feedback. In the full feedback setting we assume that the receiver has partial knowledge of the game and, at the end of each iteration, can observe the state of nature that was sampled. In bandit feedback scenario, instead we drop the assumption on the receiver's knowledge of the game and study the problem in which the feedback received by the receiver is restricted to the utility values. We first present a negative result showing the impossibility of developing an algorithm that guarantees no violation of the IC constraints and sublinear regret on the receiver's utility with respect to the optimal IC mechanism. Thus, we relax our requirements to sublinear regret and sublinear IC violation. In the full feedback setting, we provide an efficient online algorithm that attains $\tilde{\mathcal{O}}(\sqrt{T})$ regret and constraint violation. We show that achieving the same guarantees is not possible under bandit feedback for which we provide a lower bound frontier on the trade-off between regret and constraint violation. This lower bound suggests that optimal algorithm can work in two phases: exploration and exploitation phase. Indeed, we show that our algorithm achieves $\tilde{\mathcal{O}}(T^{\alpha})$ regret and $\tilde{\mathcal{O}}(T^{1-\alpha/2})$ violation of the IC constraints for any $\alpha \in [1/2, 1]$, matching the lower bound frontier.

1.2. Related Works

The study of information acquisition has been mostly confined to economics, with particular focus on auctions (see, *e.g.,* (Bergemann & Välimäki, 2002; Bikhchandani & Obara, 2017; Persico, 2000)). Stegeman (1996) studies the problem of acquiring information from a set of buyers in an auction with costly communication, while Shi (2012) investigates a setting in which the buyers pay a cost to acquire information. Recently, some works addressed the computational problem of information acquisition. More in particular, a line of work Chen & Yu (2021); Oesterheld & Conitzer (2020); Papireddygari & Waggoner (2022); Li et al. (2022); Neyman et al. (2021) studies the problem of designing optimal scoring rules to incentivize an agent to acquire and report costly information. Crucially, these works differ from our work in many aspects. For instance, we study how to incentivize agents only by inducing favorable outcomes of the game (*i.e.,* actions), while previous works rely on payments. Moreover, to the best of our knowledge, this paper is the first work addressing the computational problem of information acquisition with multiple agents.

Our work is also related to the study of Bayesian persuasion in online settings. Some works study the learning problems in which the receivers' payoffs are unknown (see, e.g., (Castiglioni et al., 2021b; 2020b; 2023)). Other works study Bayesian persuasion in MDPs. Gan et al. (2022) show how to efficiently find a sender-optimal policy when the receiver is myopic. Wu et al. (2022) extend the work considering an unknown environment. The works closest to ours consider online problems in which the agents do not know the prior over the states of nature. Zu et al. (2021) studies a persuasion problem in which the sender and the receiver do not know the prior and interact online. Bernasconi et al. (2022) extend the analysis to sequential games.

2. Preliminaries

In this section we present the model of interaction between the different agents involved and introduce the main solution concepts on which we rely.

Game model. We study the case in which an agent R (called *receiver*) interacts with a set of agents[1] $\mathcal{N} = [n]$ (called *senders*). At the beginning of the game, a *state of nature* $\theta \in \Theta$ is sampled from a *prior*[2] $p \in \Delta(\Theta)$, where Θ is the set of possible states of nature. Neither the receiver, nor the senders, observe the state of nature. Instead, each sender $i \in \mathcal{N}$ observes a signal s_i from a specific set of signals S. The signal s_i is drawn independently for each $i \in \mathcal{N}$ from a probability distribution dependent on the state of nature, which is specified by the *signaling scheme* $\psi_{\text{S}} : \Theta \to \Delta(S)$ that is shared among all the senders. We write $\psi_{\text{S}}(s_i|\theta)$ to denote the probability with which, when the state of nature is θ, signal $s_i \in S$ is sampled. After observing the signal s_i, each sender $i \in \mathcal{N}$ communicates some signal s'_i (note that it might be the case that $s'_i \neq s_i$) to R, which, based on the *signal profile* $\boldsymbol{s}' = (s'_1, ..., s'_n)$ received, selects an action from a finite set $\mathcal{A}$. A mechanism $\boldsymbol{x}$ for the receiver defines a mapping $\boldsymbol{x} : \mathcal{S} \to \Delta(\mathcal{A})$ from signal profiles $\boldsymbol{s} \in \mathcal{S} := \times_{i \in \mathcal{N}} S$, to probability distributions defined over the set of actions. The set of possible mechanisms can be defined as the polytope $\mathcal{X}$ composed of vectors[3] indexed over pairs $(\boldsymbol{s}, a) \in \mathcal{S} \times \mathcal{A}$ such that[4]

$$\mathcal{X} := \left\{ \boldsymbol{x} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}||\mathcal{A}|} \mid \sum_{a \in \mathcal{A}} x[\boldsymbol{s}, a] = 1 \quad \forall \boldsymbol{s} \in \mathcal{S} \right\}.$$

[1] In this work, given $n \in \mathbb{N}_{>0}$, we denote as $[n] = \{1, ..., n\}$ the set of the first n natural numbers.

[2] In this work, for any finite set $\mathcal{Z}$, we denote as $\Delta(\mathcal{Z})$ the set of probability distributions defined over the elements of $\mathcal{Z}$.

[3] We denote vectors with bold symbols. For any finite set D, $\boldsymbol{v} \in \mathbb{R}^{|D|}$ denotes a $|D|$-dimensional vector indexed over D, where for each $d \in D$, $v[d]$ is the value of $\boldsymbol{v}$'s component corresponding to $d \in D$.

[4] We drop the parentheses and denote $x[(s, a)]$ as $x[s, a]$.

Given a signal profile $\boldsymbol{s} \in \mathcal{S}$, we denote as $\boldsymbol{s}_{-i}$ the signal profile obtained from $\boldsymbol{s}$ by removing the i-th element s_i. Similarly, we denote as $(s'_i, \boldsymbol{s}_{-i})$ the signal profile obtained from $\boldsymbol{s}$ by substituting signal s_i with signal s'_i. After the receiver's choice of an action $a \in \mathcal{A}$, she receives a payoff $u_{\text{R}}(a, \theta) \in [0, 1]$, while each sender $i \in \mathcal{N}$ receives an equal payoff $u_i(a, \theta) = u_{\text{S}}(a, \theta) \in [0, 1]$. Throughout this work, we will make the following assumption:

Assumption 1. The number of signals $|S|$ is a constant.

Assumption 1 is required to represent succinctly the set of possible mechanisms. Indeed, in general a mechanism must specify an action distribution for each signal profile in $\mathcal{S}$, which are exponentially many. Such mechanisms has an exponentially-large representation and cannot be represented efficiently. In the following Section, we show that thanks to the symmetry among the senders, symmetric mechanisms, *i.e.*, which recommend the same action distribution for symmetric signal profiles, are optimal. However, representing symmetric mechanisms requires space exponential in $|S|$. Our assumption is required to efficiently represent symmetric mechanisms. It is not clear whether there exists a class of optimal mechanisms that can be represented in polynomial space when $|S|$ is not constant.

Incentive compatibility. In this work, we are interested in characterizing the subset of receiver's mechanisms such that all the senders are incentivized to report truthfully the information they have (*i.e.,* the signal they observed). To this extent, we consider a finite set of *deviation functions* Φ for the senders, where each $\varphi \in \Phi$ defines a mapping $\varphi : S \to S$ that specifies the signal that the sender is going to report to R, following every possible signal observation. We are interested in characterizing the set of receiver's mechanisms – denoted as *Incentive Compatible* (IC) mechanisms – that are stable with respect to unilateral deviations of senders in reporting the information they have. More in detail, we require the expected utility of each sender reporting truthfully their information not to be smaller than the expected utility she would get following any deviation function $\varphi \in \Phi$. When R uses mechanism $\boldsymbol{x} \in \mathcal{X}$ and senders report truthfully the signals they observe, the expected utility that each agent $i \in \mathcal{N} \cup \{\text{R}\}$ gets is defined as:

$$U_i(\boldsymbol{x}) \coloneqq \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x[\boldsymbol{s}, a] \sum_{\theta \in \Theta} p(\theta) u_i(a, \theta) \psi_{\text{S}}(\boldsymbol{s}|\theta),$$

where, with an overload of the notation, we denoted $\psi_{\text{S}}(\boldsymbol{s}|\theta) = \prod_{j \in \mathcal{N}} \psi_{\text{S}}(s_j|\theta)$. Similarly, the expected utility of sender $i \in \mathcal{N}$, when she deviates following deviation function $\varphi \in \Phi$ and all the others behave honestly is the following:

$$U_i^{\varphi}(\boldsymbol{x}) \coloneqq \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x[(\varphi(s_i), \boldsymbol{s}_{-i}), a] \sum_{\theta \in \Theta} p(\theta) u_i(a, \theta) \psi_{\text{S}}(\boldsymbol{s}|\theta).$$

We are now ready to formally define the set of IC mechanisms.

Definition 2.1 (ε-IC mechanisms)**.** For $\varepsilon \geq 0$, the set of ε-IC mechanisms is defined as the polytope

$$\mathcal{X}_\varepsilon \coloneqq \{\boldsymbol{x} \in \mathcal{X} \mid U_i^{\varphi}(\boldsymbol{x}) - U_i(\boldsymbol{x}) \leq \varepsilon \ \forall i \in \mathcal{N}, \forall \varphi \in \Phi\}.$$

The set of IC mechanisms is the set $\mathcal{X}_0$.

***Ex-ante* and *ex-interim* deviations.** The set of deviations Φ can be suitably specified in order to capture different concepts of incentive compatibility. In this work, we are interested in two particular types of incentive compatibility, namely *ex-ante* incentive compatibility and *ex-interim* incentive compatibility. In the case of ex-ante incentive compatibility, each sender can choose to deviate *only before* observing the signal, thus she is not able to discriminate between the different signals actually sampled when choosing which signal to report. Hence, in this case, the set of possible deviations coincides with the set $\Phi^{ante} \coloneqq \{\varphi \mid \varphi : S \to S, \varphi(s) = \varphi(s') \ \forall s, s' \in S\}$. Instead, when considering the notion of ex-interim incentive compatibility, the senders first observe the signal sampled, and then decide which signal to report to R. Thus, the set of possible deviations for the ex-interim case can be formulated as $\Phi^{inter} \coloneqq \{\varphi \mid \varphi \in S \to S\}$. Crucially, while the number of ex-ante deviation functions is $|\Phi^{ante}| = |S|$, the number of possible ex-interim deviation functions is significantly larger, namely $|\Phi^{inter}| = |S|^{|S|}$. However, when the objective is to guarantee ex-interim incentive compatibility, similarly to (Greenwald et al., 2011), it is possible to show the following[5]:

Proposition 2.1. *For all $s \in S$ let $\overline{\Phi}(s)$ be the set of deviation functions such that*

$$\overline{\Phi}(s) \coloneqq \{\varphi \mid \varphi \in S \to S, \varphi(s') = s', \quad \forall s' \in S \setminus \{s\}\}.$$

Then, for all $\boldsymbol{x} \in \mathcal{X}$ and $\varepsilon \geq 0$, if $\boldsymbol{x}$ is ε-IC with respect to the set of deviation functions $\overline{\Phi} \coloneqq \cup_{s \in S} \overline{\Phi}(s)$, then it is also $|S|\varepsilon$-IC with respect to deviation functions Φ^{inter}.

Since the number of deviation functions in $\overline{\Phi} = \cup_{s \in S} \overline{\Phi}(s)$ is bounded by $|\overline{\Phi}| = |S|^2 - |S| + 1$, as a byproduct of Proposition 2.1, we have that it is possible to efficiently represent the needed deviation functions also when the incentive compatibility concept of interest is the ex-interim one.

Throughout the rest of this paper, we will analyze the problem of information acquisition for a general incentive compatibility concept and thus for a general set of deviation functions Φ. The results that we obtain can be easily adapted to capture both ex-ante and ex-interim incentive compatibility by instantiating the set Φ to Φ^{ante} and $\overline{\Phi}$, respectively.

[5] All the proofs are provided in the appendix.

3. Offline Information Acquisition

We are interested in cases in which the objective of the receiver is twofold: (i) on the one hand she is interested in incentivizing the senders to report truthfully their private information, (ii) while on the other hand she is interested in maximizing her expected utility when all the senders behave truthfully. Formally, such a twofold objective can be modeled in terms of a linear optimization problem where the feasibility set is restricted to IC mechanisms (i), and the objective is the maximization of the receiver's expected utility U_{R} (ii). Formally,

Definition 3.1 (Optimal ε-IC mechanism)**.** An optimal ε-IC mechanism $\boldsymbol{x}^\star$ is defined as:

$$\boldsymbol{x}^\star \in \operatorname*{argmax}_{\boldsymbol{x}\in\mathcal{X}_\varepsilon} U_{\mathrm{R}}\left(\boldsymbol{x}\right).$$

An optimal IC mechanism is an optimal 0-IC mechanism.

In this Section, we investigate the problem of computing an optimal IC mechanism when all the game parameters (*i.e.,* the utility functions, the prior and the signaling scheme) are known. The first problem that we investigate concerns the representation of the polytope $\mathcal{X}$. The trivial way of representing such space would require for each vector $\boldsymbol{x} \in \mathcal{X}$ a number of entries exponential in the number of agents n, since the number of possible signal profiles is $|\mathcal{S}| = |S|^n$. We tackle this problem by showing that, when the number of signals $|S|$ is a constant, in order to compute an optimal IC mechanism, it is possible to restrict our attention to a subset of $\mathcal{X}$ that can be represented succinctly.

3.1. Compact Formulation of the Set of Mechanisms

The rationale behind our construction exploits the fact that, since the senders are symmetric (*i.e.,* they share the same utility functions, signal spaces and signaling schemes), it is possible to safely aggregate different signal profiles, thus reducing the number of variables needed to describe the mechanisms' polytope. In order to characterize this notion of equivalence between signal profiles, we introduce the concept of *symmetric signal profiles*[6].

Definition 3.2 (Symmetric signal profiles)**.** Two signal profiles $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}$ are *symmetric* (in symbols $\boldsymbol{s} \bowtie \boldsymbol{s}'$) if

$$\sum_{i\in\mathcal{N}} \mathbb{1}\left[s_i = s\right] = \sum_{i\in\mathcal{N}} \mathbb{1}\left[s'_i = s\right] \quad \forall s \in S.$$

In other words, two signal profiles are symmetric if each signal $s \in S$ occurs the same number of times in both signal profiles. The $\bowtie$ relationship defines a partition of elements in $\mathcal{S}$. In particular, we can define a partition $\mathcal{C} = \{c_1, ..., c_K\}$ of $\mathcal{S}$ such that for each $c_i \in \mathcal{C}$, $\boldsymbol{s}, \boldsymbol{s}' \in c_i$ if and only if $\boldsymbol{s} \bowtie \boldsymbol{s}'$. Given a signal profile $\boldsymbol{s} \in \mathcal{S}$, we denote with $c(\boldsymbol{s}) \in \mathcal{C}$ the element in $\mathcal{C}$ such that $\boldsymbol{s} \in c(\boldsymbol{s})$. Moreover, the partition $\mathcal{C}$ can be used to define the set of *symmetric mechanisms*, which contains all the mechanisms that associate the same probability distribution over actions to signal profiles belonging to the same class[7]. Formally,

Definition 3.3 (Symmetric mechanisms)**.** The set of symmetric mechanisms is defined as the set $\mathcal{X}^\circ \subseteq \mathcal{X}$ such that:

$$\mathcal{X}^\circ \coloneqq \left\{\boldsymbol{x} \in \mathcal{X} \mid x[\boldsymbol{s}, a] = x[\boldsymbol{s}', a] \quad \forall \boldsymbol{s} \bowtie \boldsymbol{s}', \forall a \in \mathcal{A}\right\}.$$

The following Theorem provides a key result concerning the optimality of symmetric mechanisms.

Theorem 3.1. *For any $\varepsilon \geq 0$, there exists a symmetric mechanism $\boldsymbol{x}^\circ \in \mathcal{X}^\circ$ such that*

$$\boldsymbol{x}^\circ \in \operatorname*{argmax}_{\boldsymbol{x}\in\mathcal{X}_\varepsilon} U_{\mathrm{R}}(\boldsymbol{x}).$$

Theorem 3.1 states that, an optimal IC mechanism can be found by restricting our attention to symmetric mechanisms only. To this extent, we provide a compact formulation of the polytope of symmetric mechanisms as follows:

$$\Xi \coloneqq \left\{\boldsymbol{\xi} \in \mathbb{R}_{\geq 0}^{|\mathcal{C}||\mathcal{A}|} \mid \sum_{a\in\mathcal{A}} \xi[c, a] = 1 \quad \forall c \in \mathcal{C}\right\}.$$

We also define the set of *deterministic symmetric mechanisms* which is defined as the set Π of mechanisms that associate deterministically an action to each class $c \in \mathcal{C}$. Formally, $\Pi \coloneqq \Xi \cap \{0, 1\}^{|\mathcal{C}||\mathcal{A}|}$. Crucially, differently than $\mathcal{X}$, as we show in the following Lemma, the polytope Ξ can be represented efficiently.

Lemma 3.1. *The polytope Ξ admits a description of size polynomial in the dimension of the game.*

3.2. Efficient Computation of an Optimal IC Mechanism

To conclude this Section, we provide a compact formulation of the problem of computing an optimal IC mechanism, leveraging the definition of the polytope Ξ of symmetric mechanisms.

First, let us introduce the vectors $\boldsymbol{r}_i \in \mathbb{R}_{\geq 0}^{|\mathcal{C}||\mathcal{A}|}$, defined as follows:

$$r_i[c, a] \coloneqq \sum_{\boldsymbol{s}\in c} \sum_{\theta\in\Theta} p(\theta) u_i(a, \theta) \psi(\boldsymbol{s}|\theta) \quad \forall c \in \mathcal{C}, \forall a \in \mathcal{A}.$$

Notice that, since all the senders are symmetric (in particular since $\forall i, j \in \mathcal{N}$, $u_i = u_j = u_{\mathrm{S}}$), for each $i, j \in \mathcal{N}$ we have that $\boldsymbol{r}_i = \boldsymbol{r}_j$. As a consequence of this, in order to lighten the notation, we will use a single utility parameter

[6] In this work, $\mathbb{1}[\cdot]$ denotes the indicator function.

[7] Each element $c \in \mathcal{C}$ will be also called *class*.

$\boldsymbol{r}_{\text{S}} = \boldsymbol{r}_i$, $\forall i \in \mathcal{N}$, which is common to all the senders. Then, trivially, the expected utilities of each sender and of the receiver, when the mechanism used by R is $\boldsymbol{\xi}$ and all senders report truthfully the signal they observe, can be formulated, respectively, as the following linear functions in $\boldsymbol{\xi}$:

$$U_i(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top \boldsymbol{r}_{\text{S}} \quad \forall i \in \mathcal{N} \quad \text{and} \quad U_{\text{R}}(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top \boldsymbol{r}_{\text{R}}.$$

In a similar way, as the following Lemma shows, it is possible to express the expected utility $U_i^\varphi(\boldsymbol{\xi})$ of agent $i \in \mathcal{N}$, when she deviates according to deviation function $\varphi \in \Phi$ and all the other agents report truthfully the information they have, as a bilinear function of $\boldsymbol{\xi}$ and $\boldsymbol{r}_{\text{S}}$. Formally[8]:

Lemma 3.2. *For each $\varphi \in \Phi$ and $i \in \mathcal{N}$, let $\boldsymbol{A}_i^\varphi \in \mathbb{R}^{|\mathcal{C}||\mathcal{A}| \times |\mathcal{C}||\mathcal{A}|}$ be the matrix such that $\forall c, c' \in \mathcal{C}$, $\forall a, a' \in \mathcal{A}$,*

$$A_i^\varphi[(c,a),(c',a')] = \begin{cases} 0 & \text{if } a \neq a' \\ \frac{\sum_{s \in c'} \mathbb{1}[(\varphi(s_i), \boldsymbol{s}_{-i}) \in c]}{|c'|} & \text{otherwise.} \end{cases}$$

Then, it holds $U_i^\varphi(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top \boldsymbol{A}_i^\varphi \boldsymbol{r}_{\text{S}}$ for all $i \in \mathcal{N}$.

Intuitively, for any $c, c' \in \mathcal{C}$ and any action $a \in \mathcal{A}$, the parameter $A_i^\varphi[(c,a),(c',a)]$ encodes the probability with which, when sender i deviates according to φ, R receives a signal profile belonging to c, given that the true signal profile sampled was in c'. Again, it is possible to exploit the symmetry between the senders to show the following:

Lemma 3.3. *For any $i, j \in \mathcal{N}$ and any $\varphi \in \Phi$ it holds that $\boldsymbol{A}_i^\varphi = \boldsymbol{A}_j^\varphi$.*

As we did above, to ease the notation, given a deviation function $\varphi \in \Phi$, we define a single deviation matrix $\boldsymbol{A}_{\text{S}}^\varphi = \boldsymbol{A}_i^\varphi$ for all $i \in \mathcal{N}$. Let us introduce the linear optimization problem $\text{LP}(\varepsilon, \hat{\boldsymbol{r}}_{\text{R}}, \hat{\boldsymbol{r}}_{\text{S}})$, which is defined as follows:

$$\text{LP}(\varepsilon, \hat{\boldsymbol{r}}_{\text{R}}, \hat{\boldsymbol{r}}_{\text{S}}) := \begin{cases} \max_{\boldsymbol{\xi} \in \Xi} \ \boldsymbol{\xi}^\top \hat{\boldsymbol{r}}_{\text{R}} \\ \qquad \boldsymbol{\xi}^\top (\boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}} - \hat{\boldsymbol{r}}_{\text{S}}) \leq \varepsilon \ \ \forall \varphi \in \Phi. \end{cases}$$

The objective function of the LP consists in the maximization of the expected utility of the receiver computed with respect to the utility vector $\hat{\boldsymbol{r}}_{\text{R}}$, while the constraints restrict the feasibility set to ε-IC symmetric mechanisms. Crucially, given the structure of the game, the parameters needed to define $\text{LP}(\varepsilon, \hat{\boldsymbol{r}}_{\text{R}}, \hat{\boldsymbol{r}}_{\text{S}})$ can be determined efficiently, thus allowing us to efficiently compute an optimal ε-IC mechanism.

[8] We denote matrices with bold symbols. Given two finite sets D and E, $\boldsymbol{M} \in \mathbb{R}^{|D| \times |E|}$ denotes a $|D|$ by $|E|$-dimensional matrix indexed over D and E, where $M[d,e]$ is the value of M's component corresponding to the pair $d, e \in D \times E$.

Proposition 3.1. *For any deviation function $\varphi \in \Phi$, the coefficients of the matrix $\boldsymbol{A}_{\text{S}}^\varphi$ and those of the utility vectors $\boldsymbol{r}_{\text{R}}$ and $\boldsymbol{r}_{\text{S}}$ can be computed in time polynomial in the game size.*

Corollary 3.2. *For any $\varepsilon \geq 0$, an optimal ε-IC mechanism can be found in time polynomial in the game size by solving LP$(\varepsilon, \boldsymbol{r}_{\text{R}}, \boldsymbol{r}_{\text{S}})$.*

4. Online Information Acquisition

In this work, we study the information acquisition problem in a typical *online learning setting*, capturing the scenario in which a receiver R sequentially interacts with a set of senders in a game which structure might be either partially or completely unknown to R.

The sequential interaction unfolds as follows. At each round $t \in [T]$, a state of nature θ^t is sampled according to prior p. Then, each sender $i \in \mathcal{N}$ observes a signal $s_i^t \sim \psi_{\text{S}}(\cdot|\theta^t)$ and reports truthfully her observation to R. This latter selects a mechanism $\boldsymbol{\xi}^t \in \Xi$ and, according to such mechanism, she samples an action $a^t \sim \boldsymbol{\xi}^t[c(\boldsymbol{s}^t), \cdot]$. Finally, each sender receives a payoff $u_{\text{S}}^t = u_{\text{S}}(a^t, \theta^t)$, while the receiver receives a payoff $u_{\text{R}}^t = u_{\text{R}}(a^t, \theta^t)$.

As discussed in Section 2, the goal of the receiver is the maximization of her own expected utility, while incentivizing the senders to report truthfully the signal they observe. This second objective corresponds to requiring each mechanism $\boldsymbol{\xi}^t$, $t \in [T]$, to be IC, while the performances over T rounds for what concerns the first objective are measured by means of the cumulative regret R^T, which is defined as

$$R^T := \sum_{t \in [T]} \left[(\boldsymbol{\xi}^\star)^\top \boldsymbol{r}_{\text{R}} - (\boldsymbol{\xi}^t)^\top \boldsymbol{r}_{\text{R}} \right],$$

where $\boldsymbol{\xi}^\star$ is an optimal IC mechanism. The cumulative regret measures the loss in expected utility suffered by R for having adopted in the first T rounds mechanisms $\boldsymbol{\xi}^t$ instead of the optimal IC mechanism $\boldsymbol{\xi}^\star$. As customary, we require the cumulative regret to grow sublinearly in T, *i.e.,* $R^T = o(T)$.

While asking for each mechanism $\boldsymbol{\xi}^t$ to be IC seems a reasonable requirement, as we show in the following theorem, it is not possible to guarantee with high probability that each $\boldsymbol{\xi}^t$ is IC while attaining sublinear regret.

Theorem 4.1. *There exists a constant $\delta > 0$ such that no algorithm can guarantee to output a sequence of mechanisms $\boldsymbol{\xi}^1, ..., \boldsymbol{\xi}^T$ such that, with probability at least $1 - \delta$ all mechanisms are IC and $R^T = o(T)$.*

Motivated by such an impossibility result, we introduce the concept of cumulative IC violation V^T, defined as

$$V^T := \sum_{t \in [T]} \left[\max_{\varphi \in \Phi} (\boldsymbol{\xi}^t)^\top (\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}) \right].$$

Intuitively, the cumulative IC violation expresses how much the mechanisms $\boldsymbol{\xi}^t$ chosen by the receiver at each round t cumulatively violate the IC constraint of $\text{LP}(0, \boldsymbol{r}_{\text{R}}, \boldsymbol{r}_{\text{S}})$. From another perspective, V^T captures how much utility a sender reporting signals untruthfully would have gained with respect to what she gained by always behaving truthfully. Thus, V^T can also be interpreted as a measure of senders' regret for truthful behavior. Relaxing the incentive compatibility requirement, our new objective becomes the development of learning algorithms for the receiver that attain cumulative regret and cumulative IC violation that grow sublinearly in T, namely $R^T = o(T)$ and $V^T = o(T)$. Achieving $V^T = o(T)$ guarantees that, for each sender, truthful behavior converges to the optimal behavior as $T \to \infty$. We remark that, given the negative result of Theorem 4.1, it is impossible to guarantee that truthful behavior is an exact best-response to receiver's mechanisms $\boldsymbol{\xi}^1, ..., \boldsymbol{\xi}^T$.

In the following, we investigate two different scenarios, which relate to different degrees of knowledge that the receiver might have on the game structure and to different types of *feedback* that the receiver gets out of the sequential interaction. In particular, in the *full feedback setting* – which is discussed in Section 5 – we assume that the receiver knows both utility functions u_{S} and u_{R}, while she does not know the signaling scheme nor the prior p. Furthermore, in such setting, at the end of each round $t \in [T]$, the receiver observes, together with the signal profile $\boldsymbol{s}^t$ and the action a^t, the state of nature θ^t that has been sampled. The second scenario that we study is the *bandit feedback setting*, presented in Section 6. In this case, we assume no knowledge of the game structure on the receiver's side (*i.e.,* the receiver does not know the prior p, the signaling scheme ψ_{S}, and the utility functions u_{S} and u_{R}) and we restrict the feedback that R receives at each round $t \in [T]$ to the utility values u_{S}^t, u_{R}^t, the signal profile $\boldsymbol{s}^t$ and the action a^t.

5. Online Information Acquisition with Full Feedback

In the full feedback setting, the receiver can leverage her knowledge of the game parameters, as well as the observation of sampled signal profiles and states of nature, to obtain estimates of the products $p(\theta)\psi_{\text{S}}(\boldsymbol{s}|\theta)$ for each $\boldsymbol{s} \in \mathcal{S}$ and $\theta \in \Theta$. Then, we can easily recover estimators for the vectors $\boldsymbol{r}_i$, $i \in \{\text{R}, \text{S}\}$, with appropriate high-confidence bounds. As we show in this Section, such an high confidence region can then be exploited in order to define an online learning algorithm capable of guaranteeing $\tilde{\mathcal{O}}(\sqrt{T})$ cumulative regret and IC violations with high probability. In the first part of this section, we show how the information available to the receiver can be used in order to estimate the vectors $\boldsymbol{r}_{\text{R}}, \boldsymbol{r}_{\text{S}}$, while the second part of this section is devoted to the presentation and analysis of the algorithm for the full feedback setting.

5.1. Estimation of $\boldsymbol{r}_{\text{R}}$ and $\boldsymbol{r}_{\text{S}}$

Under the full feedback assumption, at each $t \in [T]$, the receiver observes both the signal profile $\boldsymbol{s}^t \in \mathcal{S}$ and the state of nature $\theta^t \in \Theta$. Thus, using the information collected up to round t, as well as the knowledge of the utility functions u_{R} and u_{S}, a possible estimator of the utility vectors is the following for each $i \in \{\text{R}, \text{S}\}$, for each $c \in \mathcal{C}$ and for each $a \in \mathcal{A}$:

$$\hat{r}_i^{t+1}[c, a] := \frac{1}{t} \sum_{\theta \in \Theta} u_i(a, \theta) \sum_{\tau \in [t]} \mathbb{1}\left[c(\boldsymbol{s}^\tau) = c, \theta^\tau = \theta\right]. \quad (1)$$

Intuitively, the term $\sum_{\tau \in [t]} \mathbb{1}\left[c(\boldsymbol{s}^\tau) = c, \theta^\tau = \theta\right]/t$ provides an unbiased estimator of the product $p(\theta) \sum_{\boldsymbol{s} \in c} \psi_{\text{S}}(\boldsymbol{s}|\theta)$, thus allowing us to conclude the following result:

Lemma 5.1. *For all $t \in [T]$ and $i \in \{\text{R}, \text{S}\}$, $\hat{\boldsymbol{r}}_i^t$ is an unbiased estimator of $\boldsymbol{r}_i$.*

As a second step, we are interested in deriving an high confidence region around $\boldsymbol{r}_{\text{R}}$ and $\boldsymbol{r}_{\text{S}}$. For $\delta \in (0, 1)$, let us introduce the event $\mathcal{E}_\delta^{\text{F}}$, defined as

$$\mathcal{E}_\delta^{\text{F}} := \left\{ ||\boldsymbol{r}_i - \hat{\boldsymbol{r}}_i^t||_\infty \leq \varepsilon_\delta^t \ \forall t \in [T], \forall i \in \{\text{R}, \text{S}\} \right\},$$

where

$$\varepsilon_\delta^t := \sqrt{\frac{\log\left(4T|\mathcal{C}||\mathcal{A}|/\delta\right)}{2(t-1)}}. \quad (2)$$

Then, using standard concentration arguments, it is possible to state the following Lemma.

Lemma 5.2. *For any $\delta \in (0, 1)$, the following holds*

$$\mathbb{P}\left(\mathcal{E}_\delta^{\text{F}}\right) \geq 1 - \delta.$$

5.2. Algorithm

Algorithm 1 Algorithm for the full feedback setting

Require: $T \in \mathbb{N}_{>0}$, $\delta \in (0, 1)$, u_{S}, u_{R}
  Initialize:
    $\hat{r}_i^1[c, a] \leftarrow 0 \quad \forall i \in \{\text{R}, \text{S}\}, \forall c \in \mathcal{C}, \forall a \in \mathcal{A}$
    $\nu^1 \leftarrow \infty, \ \varepsilon_\delta^1 \leftarrow \infty$
  for $t \in [T]$ **do**
    $\boldsymbol{\xi}^t \leftarrow$ optimal solution to $\text{LP}(\nu^t, \hat{\boldsymbol{r}}_{\text{R}}^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$
    Use $\boldsymbol{\xi}^t$ and observe θ^t, $\boldsymbol{s}^t$, a^t, u_{R}^t, u_{S}^t
    Update estimators as in Equation (1)
  end for

The procedure that we employ to solve the online problem in the full feedback setting is presented in Algorithm 1. The algorithm takes as input the number of rounds T,

the desired confidence level δ determining the probability with which the regret and IC violation bound hold, the utility functions u_{S} and u_{R}, and the prior p. At each round $t \in [T]$, the algorithm uses the information collected by R up to t and selects a mechanism $\boldsymbol{\xi}^t$ that is optimal for linear optimization problem $\text{LP}(\nu^t, \hat{\boldsymbol{r}}_{\text{R}}^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$, where $\nu^t = 2|\mathcal{C}|\varepsilon_\delta^t$. Notice that the feasibility parameter ν^t, which regulates the slackness of the linear program's incentive compatibility constraint, depends on the confidence bound ε_δ^t. Intuitively, this is needed to compensate the uncertainty arising from the use of the estimated utility parameters $\hat{\boldsymbol{r}}_{\text{R}}^t$, $\hat{\boldsymbol{r}}_{\text{S}}^t$ instead of the true parameters $\boldsymbol{r}_{\text{R}}$, $\boldsymbol{r}_{\text{S}}$. This choice of the slackness parameter ν^t is fundamental in order to guarantee, with high probability, sublinear cumulative regret and IC violations. Indeed, when event $\mathcal{E}_\delta^{\text{F}}$ holds, on the one hand it guarantees that the optimal IC mechanism $\boldsymbol{\xi}^\star$ is in the set of feasible mechanisms of $\text{LP}(\nu^t, \hat{\boldsymbol{r}}_{\text{R}}^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$ implying large receiver's utility, while, on the other hand, it ensures that $\boldsymbol{\xi}^t$ is $2\nu^t$-IC with respect to the real utility vectors $\boldsymbol{r}_{\text{R}}$, $\boldsymbol{r}_{\text{S}}$, implying a small violation of the IC constraints. These two properties can then be leveraged to obtain the following Theorem.

Theorem 5.1. *For any $\delta \in (0,1)$, with probability at least $1-\delta$, Algorithm 1 guarantees*

$$R^T = \mathcal{O}\left(|\mathcal{C}|\sqrt{8T\log(4T|\mathcal{C}||\mathcal{A}|/\delta)}\right),$$
$$V^T = \mathcal{O}\left(|\mathcal{C}|\sqrt{16T\log(4T|\mathcal{C}||\mathcal{A}|/\delta)}\right).$$

6. Online Information Acquisition with Bandit Feedback

Differently from the full feedback setting, where the receiver has access to complete observations of the signal profiles and states of nature, in the bandit feedback setting the receiver does not have access to this information and can only observe the utility $u_j(a^t, \theta^t)$ for $j \in \{\text{R}, \text{S}\}$, the received signal profiles $\boldsymbol{s}^t$, and the played actions a^t. This particular feedback introduces significant complications when trying to estimate the game parameters, thus requiring an adequate *exploration* of the game upfront. The algorithm presented in this section dedicates the first set of rounds to the refinement of the estimates of the game parameters, and subsequently leverages the information collected in this first phase, in order to achieve the desired performances. In the first part of this section we introduce the estimators that are used by our procedure, while the second part of the Section is devoted to the presentation of the algorithm. Finally, we provide a lower bound showing that the regret and IC violation bounds attained by our algorithm are tight, thus suggesting the need for an accurate exploration of the game in the first rounds of the interaction.

6.1. Estimation of $\boldsymbol{r}_{\text{R}}$ and $\boldsymbol{r}_{\text{S}}$

In order to appropriately define the estimators of the game parameters, we require the receiver to play at each round $t \in [T]$ a deterministic mechanism[9] $\boldsymbol{\pi}^t \in \Pi$. While in the full feedback setting the receiver could leverage the observations concerning the sampled state of nature to implicitly estimate the signaling scheme ψ_{S}, under bandit feedback the information available to R is not sufficient to acquire knowledge on ψ_{S}. Instead, the observations of the utility values u_{R}^t and u_{S}^t, the signal profile $\boldsymbol{s}^t$ and the action a^t, make it convenient to estimate directly the utility parameters $\boldsymbol{r}_{\text{R}}$ and $\boldsymbol{r}_{\text{S}}$. In particular, using the information collected by R up to time $t \in [T]$, it is possible to define, for all $j \in \{\text{R}, \text{S}\}$, $c \in \mathcal{C}$, and $a \in \mathcal{A}$, the following estimator of the utility vectors:

$$\hat{r}_j^{t+1}[c,a] := \frac{1}{N^t[c,a]} \sum_{\substack{\tau \in [t] \\ \pi^\tau[c,a]=1}} u_j(a^\tau, \theta^\tau)\mathbb{1}\left[s^\tau \in c\right], \quad (3)$$

where $N^t[c,a] := \sum_{\tau \in [t]} \pi^\tau[c,a]$ is a counter that keeps track of how many times R selected a deterministic mechanism that prescribed to play action a when receiving a signal profile belonging to class c.

Lemma 6.1. *For all $t \in [T]$, for all $i \in \{\text{R}, \text{S}\}$, $\hat{\boldsymbol{r}}_j^t$ is an unbiased estimator of $\boldsymbol{r}_i$.*

Given $\delta \in (0,1)$, let us define the event $\mathcal{E}_\delta^{\text{B}}$ such that

$$\mathcal{E}_\delta^{\text{B}} := \big\{|\hat{r}_i^t[c,a] - r_i[c,a]| \le \eta_\delta^t[c,a] \quad \forall t \in [T], \forall c \in \mathcal{C}, \forall a \in \mathcal{A} \forall i \in \{\text{R}, \text{S}\}\big\},$$

where $\boldsymbol{\eta}_\delta^t \in \mathbb{R}^{|\mathcal{C}|\times|\mathcal{A}|}$ is the vector such that, for any $\delta \in (0,1)$ and for all $t \in [T]$,

$$\eta_\delta^t[c,a] := \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2N^t[c,a]}}. \quad (4)$$

Also in this case, by means of standard concentration arguments, it is possible to derive a lower bound on the probability with which event $\mathcal{E}_\delta^{\text{B}}$ is verified, thus yielding an high confidence region for the utility vectors $\boldsymbol{r}_{\text{R}}$ and $\boldsymbol{r}_{\text{S}}$:

Lemma 6.2. *For any $\delta \in (0,1)$, the following holds:*

$$\mathbb{P}\left(\mathcal{E}_\delta^{\text{B}}\right) \ge 1 - \frac{\delta}{2}.$$

6.2. Algorithm

The algorithm for the bandit feedback setting is described in Algorithm 2. The algorithm, which is formulated as a

[9] Let us point out that, given a mechanism $\boldsymbol{\xi} \in \Xi$, a deterministic mechanism $\boldsymbol{\pi} \sim \boldsymbol{\xi}$ can be efficiently sampled from $\boldsymbol{\xi}$, simply by iterating over all elements $c \in \mathcal{C}$ and by sampling an action according to the probability distribution specified by $\xi[c,\cdot]$.

Algorithm 2 Algorithm for the bandit feedback setting

Require: $T, E \in \mathbb{N}_{>0}$, $\delta \in (0,1)$
  Initialize:
    $\hat{r}^1_j[c,a] \leftarrow 0 \quad \forall c \in \mathcal{C}, \forall a \in \mathcal{A}, \forall j \in \{\text{R}, \text{S}\}$
    $\eta^1_\delta[c,a] \leftarrow \infty, \quad N^0[c,a] \leftarrow 0 \quad \forall c \in \mathcal{C}, \forall a \in \mathcal{A}$
    $\nu \leftarrow 2|\mathcal{C}|\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}}$
  for $t \in [T]$ **do**
    if $t \leq |\mathcal{A}|E$ **then** ▷ *Exploration rounds*
      $a \leftarrow \arg\min_{a \in \mathcal{A}} \sum_{c \in \mathcal{C}} N^t[c,a]$
      $\xi^t[c,a'] \leftarrow \mathbb{1}[a' = a] \quad \forall c \in \mathcal{C}\, \forall a \in \mathcal{A}.$
    else ▷ *Optimization rounds*
      $\boldsymbol{\xi}^t \leftarrow$ optimal solution to LP $\left(\nu, \hat{\boldsymbol{r}}^t_{\text{R}} + \boldsymbol{\eta}^t_\delta, \hat{\boldsymbol{r}}^t_{\text{S}}\right)$
    end if
    ▷ *Interaction with the senders*
    $\boldsymbol{\pi}^t \leftarrow$ sample deterministic mechanism from $\boldsymbol{\xi}^t$
    Use $\boldsymbol{\pi}^t$ and observe $\boldsymbol{s}^t$, a^t, u^t_{R}, u^t_{S}
    Update Estimators
  end for

two-phases procedure, takes as input the time horizon T, the desired confidence level on the regret and IC violation bounds δ, and the number E of rounds that will be devoted to the exploration of each action. More in detail, in the first phase, which is called *exploration phase*, the objective of the receiver is to select mechanisms $\boldsymbol{\xi}^t$ that guarantee to obtain sufficiently accurate estimations of the game parameters. To this extent, the exploration phase is designed such that, for each action $a \in \mathcal{A}$, R plays for exactly E rounds a deterministic mechanism $\boldsymbol{\pi}$ that prescribes to play action a for each class $c \in \mathcal{C}$, *i.e.,* a $\boldsymbol{\pi}$ such that $\pi[c,a] = 1, \forall c \in \mathcal{C}$. Intuitively, by imposing that each couple $c, a \in \mathcal{C} \times \mathcal{A}$ is explored for an adequate number of rounds within the exploration phase, we are in fact guaranteeing an upper bound $\nu = 2|\mathcal{C}|\sqrt{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)/2E}$ on the confidence parameters $\boldsymbol{\eta}^t_\delta$ at subsequent rounds t. Such an upper bound will then prove to be crucial in order to obtain the desired performances. Furthermore, it is worth pointing out that such deterministic mechanisms do *not* contribute to the cumulative IC violation, since, trivially, selecting the same action for any possible class c constitutes an IC mechanism.

After the first $|\mathcal{A}|E$ rounds of exploration, the algorithm enters its second phase, which is called *optimization phase*. In this second set of rounds, the receiver aims at exploiting the information collected in order to select mechanisms that contribute to minimize the cumulative regret and IC violation. In particular, at each round $t > |\mathcal{A}|E$, the estimators $\hat{\boldsymbol{r}}^t_{\text{R}}$ and $\hat{\boldsymbol{r}}^t_{\text{S}}$, together with the confidence bounds $\boldsymbol{\eta}^t_\delta$ and ν, are used in the strategy selection procedure as input parameters to the linear optimization problem LP $\left(\nu, \hat{\boldsymbol{r}}^t_{\text{R}} + \boldsymbol{\eta}^t_\delta, \hat{\boldsymbol{r}}^t_{\text{S}}\right)$. More precisely, the linear program is instantiated so that the objective is exactly the optimization of the expected utility computed with respect to the upper confidence bound of the parameter $\boldsymbol{r}_{\text{R}}$ and the IC constraint is specified using the estimator $\hat{\boldsymbol{r}}^t_{\text{S}}$ as the utility parameter, with constraint slackness regulated by the parameter ν. These design choices have a twofold effect: (i) on the one hand they ensure that the optimal IC mechanism $\boldsymbol{\xi}^\star$ is in the feasibility set of the linear program, (ii) while on the other hand they provide guarantees on the approximate incentive compatibility of mechanisms $\boldsymbol{\xi}^t$ chosen at each round $t > |\mathcal{A}|E$. In particular, while (i), together with the choice of a UCB-like objective, yields optimal regret bounds, thanks to (ii) it is possible to recover optimal IC violation for Algorithm 2. The formal guarantees of the algorithm are stated in the following Theorem:

Theorem 6.1. *For any $\delta \in (0,1)$ and for any $E \in [T]$, with probability at least $1-\delta$, Algorithm 2 guarantees*

$$R^T = \mathcal{O}\left(|\mathcal{A}|E + |\mathcal{C}||\mathcal{A}|\sqrt{32T\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}\right),$$
$$V^T = \mathcal{O}\left(T|\mathcal{C}|\sqrt{8\log(8T|\mathcal{C}||\mathcal{A}|/\delta)/E}\right).$$

When the number of exploration rounds is set as $E = \lfloor T^\alpha \rfloor$ with $\alpha \in [1/2, 1]$, Theorem 6.1 implies that the cumulative regret of Algorithm 2 is $R^T = \tilde{\mathcal{O}}(T^\alpha)$, while the cumulative IC violation is $V^T = \tilde{\mathcal{O}}(T^{1-\alpha/2})$ Clearly, the choice of the number of exploration rounds introduces a fundamental trade-off between minimization of the cumulative regret and minimization of the cumulative IC violation. This trade-off is inherently related to the structure of the bandit feedback setting itself. Indeed, as we show in the remaining part of this section, it is not possible to improve over the regret and violation bounds that are attained by Algorithm 2.

6.3. Lower Bound

Before concluding the analysis of the bandit feedback setting, we derive a lower bound showing that the regret and IC violation bounds achieved by Algorithm 2 are tight.

Intuitively, the lower bound captures the need for exploration that characterizes the bandit feedback setting, fundamentally distinguishing it from the full feedback setting. Indeed, while, in the latter, the feedback received at each round t is completely informative and can be used to improve the receiver's knowledge of the game independently on the chosen mechanism $\boldsymbol{\xi}^t$, in the former setting the feedback is strongly related to the mechanisms $\boldsymbol{\xi}^t$ selected by R. This aspect makes it necessary to guarantee a suitable level of exploration, in fact – as already observed above – bringing up an actual trade-off between minimization of cumulative regret and minimization of cumulative IC deviation. The following Theorem formalizes the above ideas.

Theorem 6.2. *For any $\alpha \in [1/2, 1]$, there exists $\delta \in (0,1)$ such that no algorithm can achieve both $R^T = o(T^\alpha)$ and $V^T = o(T^{1-\alpha/2})$ with probability greater than $1-\delta$.*

As we can observe from Theorem 6.2, for $E = \lfloor T^\alpha \rfloor$ and $\alpha \in [1/2, 1]$, the regret and IC violation bounds attained by Algorithm 2 are tight.

Acknowledgements

This paper is supported by FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence).

References

Alonso, R. and Câmara, O. Persuading voters. *American Economic Review*, 106(11):3590–3605, 2016.

Bergemann, D. and Välimäki, J. Information acquisition and efficient mechanism design. *Econometrica*, 70(3): 1007–1033, 2002.

Bernasconi, M., Castiglioni, M., Marchesi, A., Gatti, N., and Trovò, F. Sequential information design: Learning to persuade in the dark. In *Advances in Neural Information Processing Systems*, 2022.

Bhaskar, U., Cheng, Y., Ko, Y. K., and Swamy, C. Hardness results for signaling in bayesian zero-sum and network routing games. In *EC*, pp. 479–496, 2016.

Bikhchandani, S. and Obara, I. Mechanism design with information acquisition. *Economic Theory*, 63(3):783–812, 2017.

Bro Miltersen, P. and Sheffet, O. Send mixed signals: earn more, work less. In *EC*, pp. 234–247, 2012.

Castiglioni, M. and Gatti, N. Persuading voters in district-based elections. In *AAAI*, pp. 5244–5251, 2021.

Castiglioni, M., Celli, A., and Gatti, N. Persuading voters: It's easy to whisper, it's hard to speak loud. In *AAAI*, pp. 1870–1877, 2020a.

Castiglioni, M., Celli, A., Marchesi, A., and Gatti, N. Online Bayesian persuasion. In *NeurIPS*, pp. 16188–16198, 2020b.

Castiglioni, M., Celli, A., Marchesi, A., and Gatti, N. Signaling in Bayesian network congestion games: the subtle power of symmetry. In *AAAI*, pp. 5252–5259, 2021a.

Castiglioni, M., Marchesi, A., Celli, A., and Gatti, N. Multi-receiver online bayesian persuasion. In *ICML*, volume 139, pp. 1314–1323, 2021b.

Castiglioni, M., Celli, A., Marchesi, A., and Gatti, N. Regret minimization in online bayesian persuasion: Handling adversarial receiver's types under full and partial feedback models. *Artificial Intelligence*, 314:103821, 2023.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, learning, and games*. Cambridge university press, 2006.

Chen, Y. and Yu, F.-Y. Optimal scoring rule design. *arXiv preprint arXiv:2107.07420*, 2021.

Gan, J., Majumdar, R., Radanovic, G., and Singla, A. Bayesian persuasion in sequential decision-making. In *AAAI*, 2022.

Greenwald, A., Jafari, A., and Marks, C. No-ϕ-regret: A connection between computational learning theory and game theory. In *Games, Norms and Reasons*, pp. 119–140. Springer, 2011.

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

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Li, Y., Hartline, J. D., Shan, L., and Wu, Y. Optimization of scoring rules. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 988–989, 2022.

Neyman, E., Noarov, G., and Weinberg, S. M. Binary scoring rules that incentivize precision. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 718–733, 2021.

Oesterheld, C. and Conitzer, V. Minimum-regret contracts for principal-expert problems. In *Web and Internet Economics: 16th International Conference, WINE 2020, Beijing, China, December 7–11, 2020, Proceedings 16*, pp. 430–443. Springer, 2020.

Papireddygari, M. and Waggoner, B. Contracts with information acquisition, via scoring rules. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 703–704, 2022.

Persico, N. Information acquisition in auctions. *Econometrica*, 68(1):135–148, 2000.

Rabinovich, Z., Jiang, A. X., Jain, M., and Xu, H. Information disclosure as a means to security. In *AAMAS*, pp. 645–653, 2015.

Shi, X. Optimal auctions with information acquisition. *Games and Economic Behavior*, 74(2):666–686, 2012.

Stegeman, M. Participation costs and efficient auctions. *Journal of Economic Theory*, 71(1):228–259, 1996.

Wu, J., Zhang, Z., Feng, Z., Wang, Z., Yang, Z., Jordan, M. I., and Xu, H. Sequential information design: Markov persuasion process and its efficient reinforcement learning. *arXiv preprint arXiv:2202.10678*, 2022.

Xu, H., Freeman, R., Conitzer, V., Dughmi, S., and Tambe, M. Signaling in Bayesian Stackelberg games. In *AAMAS*, pp. 150–158, 2016.

Zu, Y., Iyer, K., and Xu, H. Learning to persuade on the fly: Robustness against ignorance. In *EC*, pp. 927–928, 2021.

A. Proofs Omitted from Section 2

Proposition 2.1. *For all $s \in S$ let $\overline{\Phi}(s)$ be the set of deviation functions such that*

$$\overline{\Phi}(s) := \{\varphi \mid \varphi \in S \to S, \varphi(s') = s', \quad \forall s' \in S \setminus \{s\}\}.$$

Then, for all $\boldsymbol{x} \in \mathcal{X}$ and $\varepsilon \geq 0$, if $\boldsymbol{x}$ is ε-IC with respect to the set of deviation functions $\overline{\Phi} := \cup_{s \in S} \overline{\Phi}(s)$, then it is also $|S|\varepsilon$-IC with respect to deviation functions Φ^{inter}.

Proof. First, let us point out that given any mechanism $\boldsymbol{x} \in \mathcal{X}$, the utilities, $U_i(\boldsymbol{x})$ and $U_i^{\varphi}(\boldsymbol{x})$ of each deviation φ can be expressed as the following linear functions:

$$\begin{aligned} U_i(\boldsymbol{x}) &= \boldsymbol{x}^\top \boldsymbol{d}_i, \\ U_i^{\varphi}(\boldsymbol{x}) &= \boldsymbol{x}^\top \boldsymbol{M}_i^{\varphi} \boldsymbol{d}_i, \end{aligned} \qquad \forall i \in \mathcal{N},$$

where $\boldsymbol{d}_i \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ is the vector such that

$$d_i[\boldsymbol{s}, a] = \sum_{\theta \in \Theta} p(\theta) \psi_{\mathrm{s}}(\boldsymbol{s}|\theta) u_i(a, \theta) \quad \forall \boldsymbol{s} \in \mathcal{S}, \forall a \in \mathcal{A},$$

and $\boldsymbol{M}_i^{\varphi} \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}| \times |\mathcal{S}||\mathcal{A}|}$ is the matrix such that

$$M[(\boldsymbol{s}, a), (\boldsymbol{s}', a')] = \begin{cases} 1 & \text{if } \boldsymbol{s}_{-i} = \boldsymbol{s}'_{-i},\ \varphi(s_i) = s'_i,\ a = a' \\ 0 & \text{otherwise.} \end{cases}$$

For any $s, s' \in S$, let $\boldsymbol{M}^{s \to s'}$ be the matrix associated to deviation function $\varphi^{s \to s'} \in \overline{\Phi}(s)$ such that $\varphi^{s \to s'}(s) = s'$. Furthermore, let $\boldsymbol{M}^I$ be the matrix associated to the identity deviation function φ^I, *i.e.,* the deviation function such that $\varphi^I(s) = s$, for any $s \in S$. Then, for any $\varphi \in \Phi^{inter}$, we can write the following:

$$\boldsymbol{M}^{\varphi} = \sum_{s \in S} \left[\boldsymbol{M}^{s \to \varphi(s)}\right] + (1 - |S|)\boldsymbol{M}^I.$$

Now, let $\boldsymbol{x}$ be an ε-IC mechanism with respect to deviation functions $\overline{\Phi}$. It holds that for each $\varphi \in \Phi^{inter}$,

$$\begin{aligned} U_i^{\varphi}(\boldsymbol{x}) - U_i(\boldsymbol{x}) &= \boldsymbol{x}^\top \boldsymbol{M}_i^{\varphi} \boldsymbol{d}_i - \boldsymbol{x}^\top \boldsymbol{d}_i \\ &= \sum_{s \in S} \left[\boldsymbol{x}^\top \boldsymbol{M}^{s \to \varphi(s)} \boldsymbol{d}_i\right] + \boldsymbol{x}^\top \boldsymbol{M}^I \boldsymbol{d}_i - |S| \boldsymbol{x}^\top \boldsymbol{M}^I \boldsymbol{d}_i - \boldsymbol{x}^\top \boldsymbol{d}_i \\ &= \sum_{s \in S} \left[\boldsymbol{x}^\top \boldsymbol{M}^{s \to \varphi(s)} \boldsymbol{d}_i - \boldsymbol{x}^\top \boldsymbol{d}_i\right] \\ &\leq |S|\varepsilon, \end{aligned}$$

where the third equation follows from the fact that $\boldsymbol{x}^\top \boldsymbol{M}^I \boldsymbol{d}_i = \boldsymbol{x}^\top \boldsymbol{d}_i$, and the last inequality follows from the fact that $\boldsymbol{x}$ is ε-IC with respect to deviation functions $\overline{\Phi}$.

This concludes the proof. □

B. Proofs Omitted from Section 3

B.1. Proof of Theorem 3.1

Before proving Theorem 3.1, let us introduce some preliminary notation and results that will be needed in the proof.

Let Λ be the set of permutations of the elements in the set $[n]$. For each $\boldsymbol{\lambda} \in \Lambda$, let $f_{\boldsymbol{\lambda}} : \mathcal{S} \to \mathcal{S}$ be the permutation function that given a signal profile $\boldsymbol{s} \in \mathcal{S}$ returns a signal profile $\boldsymbol{s}' \in \mathcal{S}$ obtained changing the order of signals in $\boldsymbol{s}$ according to the permutation $\boldsymbol{\lambda}$. Formally,

Definition B.1. For any $\boldsymbol{\lambda} \in \Lambda$, the permutation function $f_{\boldsymbol{\lambda}} : \mathcal{S} \to \mathcal{S}$ is defined as the bijective function such that $\forall \boldsymbol{s} \in \mathcal{S}$, $f_{\boldsymbol{\lambda}}(\boldsymbol{s}) = \boldsymbol{s}' \in \mathcal{S}$, where

$$s'_i = s_{\lambda[i]} \quad \forall i \in \mathcal{N}.$$

Definition B.2. For any $\boldsymbol{\lambda} \in \Lambda$, the inverse permutation function $f_{\boldsymbol{\lambda}}^{-1} : \mathcal{S} \to \mathcal{S}$ is defined as the bijective function such that $\forall \boldsymbol{s} \in \mathcal{S}$, $f_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{s}) = \boldsymbol{s}' \in \mathcal{S}$, where

$$s'_{\lambda[i]} = s_i \quad \forall i \in \mathcal{N}.$$

For notational convenience, for any $i \in \mathcal{N}$ and $\varphi \in \Phi$, we introduce the vectors $\boldsymbol{d}_{\text{R}}, \boldsymbol{d}_i, \boldsymbol{d}_i^{\varphi} \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, where:

$$d_i[\boldsymbol{s}, a] = \sum_{\theta \in \Theta} p(\theta) u_i(a, \theta) \psi_{\text{S}}(\boldsymbol{s}|\theta) \quad \forall i \in \mathcal{N} \cup \{\text{R}\}$$

$$d_i^{\varphi}[\boldsymbol{s}, a] = \sum_{\substack{\boldsymbol{s}' \in \mathcal{S} \\ \boldsymbol{s}_{-i} = \boldsymbol{s}'_{-i} \\ \varphi(s'_i) = s_i}} d_i[\boldsymbol{s}', a] = \sum_{\theta \in \Theta} p(\theta) u_i(a, \theta) \sum_{\substack{\boldsymbol{s}' \in \mathcal{S} \\ \boldsymbol{s}_{-i} = \boldsymbol{s}'_{-i} \\ \varphi(s'_i) = s_i}} \psi_{\text{S}}(\boldsymbol{s}|\theta) \quad \forall i \in \mathcal{N}.$$

Then, exploiting the above formulations, we can rewrite the expected utilities as:

$$U_i(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x[\boldsymbol{s}, a] d_i[\boldsymbol{s}, a] \quad \forall i \in \mathcal{N} \cup \{\text{R}\}$$

$$U_i^{\varphi}(\boldsymbol{s}) = \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x[\boldsymbol{s}, a] d_i^{\varphi}[\boldsymbol{s}, a] \quad \forall i \in \mathcal{N}, \forall \varphi \in \Phi.$$

We can state the following auxiliary result:

Lemma B.1. *For any $\varepsilon \geq 0$, let $\boldsymbol{x}^\star \in \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}_\varepsilon} U_{\text{R}}(\boldsymbol{x})$. Define, for each permutation $\boldsymbol{\lambda} \in \Lambda$, the mechanism $\boldsymbol{x}_{\boldsymbol{\lambda}} \in \mathcal{X}$ as the mechanism such that, $\forall \boldsymbol{s} \in \mathcal{S}$, $\forall a \in \mathcal{A}$, $\boldsymbol{x}^\star[\boldsymbol{s}, a] = \boldsymbol{x}_{\boldsymbol{\lambda}}[f_{\boldsymbol{\lambda}}(\boldsymbol{s}), a]$. Then, the following holds:*

$$\boldsymbol{x}_{\boldsymbol{\lambda}} \in \operatorname*{argmax}_{\boldsymbol{x} \in \mathcal{X}_\varepsilon} U_{\text{R}}(\boldsymbol{x}).$$

Proof. In order to show that $\boldsymbol{x}_{\boldsymbol{\lambda}} \in \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}_\varepsilon} U_{\text{R}}(\boldsymbol{x})$, we first show that $U_{\text{R}}(\boldsymbol{x}^\star) = U_{\text{R}}(\boldsymbol{x}_{\boldsymbol{\lambda}})$ and then that $\boldsymbol{x}_{\boldsymbol{\lambda}} \in \mathcal{X}_\varepsilon$.

Objective function. Note that $\forall \theta \in \Theta$, $\forall \boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}$ such that $\boldsymbol{s} \bowtie \boldsymbol{s}'$, it holds that

$$\psi_{\text{S}}(\boldsymbol{s}|\theta) = \psi_{\text{S}}(\boldsymbol{s}'|\theta).$$

Hence, for any $i \in \mathcal{N} \cup \{\text{R}\}$ and $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}$ such that $\boldsymbol{s} \bowtie \boldsymbol{s}'$

$$d_i[\boldsymbol{s}, a] = \sum_{\theta \in \Theta} p(\theta) u_i(a, \theta) \psi_{\text{S}}(\boldsymbol{s}|\theta) = \sum_{\theta \in \Theta} p(\theta) u_i(a, \theta) \psi_{\text{S}}(\boldsymbol{s}'|\theta) = d_i[\boldsymbol{s}', a] \tag{5}$$

Furthermore, since the permuation function preserves the elements present in the signal profile and only changes their order, $\forall \boldsymbol{s} \in \mathcal{S}$, $\forall \boldsymbol{\lambda} \in \Lambda$, we have that $f_{\boldsymbol{\lambda}}(\boldsymbol{s}) \bowtie \boldsymbol{s}$. Thus, $\forall i \in \mathcal{N} \cup \{\text{R}\}$, we can write the following equations:

$$\begin{aligned}
U_i(\boldsymbol{x}_{\boldsymbol{\lambda}}) &= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x_{\boldsymbol{\lambda}}[\boldsymbol{s}, a] d_i[\boldsymbol{s}, a] \\
&= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x^\star[f_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{s}), a] d_i[\boldsymbol{s}, a] \\
&= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x^\star[\boldsymbol{s}, a] d_i[f_{\boldsymbol{\lambda}}(\boldsymbol{s}), a] \\
&= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x^\star[\boldsymbol{s}, a] d_i[\boldsymbol{s}, a] \\
&= U_i(\boldsymbol{x}^\star),
\end{aligned} \tag{6}$$

where the first equation follows from the definition of $\boldsymbol{x}_{\boldsymbol{\lambda}}$, the second and third equation follow from the fact that the function $f_{\boldsymbol{\lambda}}$ is bijective and the last equation follows from Equation (5). This concludes the first part of the proof.

Feasibility. Let $\mathcal{S}_{-j} = \times_{i \in \mathcal{N} \setminus \{j\}} \mathcal{S}_i$. Notice that $\forall \boldsymbol{s} \in \mathcal{S}, \forall \boldsymbol{\lambda} \in \Lambda$,

$$\left\{ \boldsymbol{s}'_{-i} \in \mathcal{S}_{-i} \mid \boldsymbol{s}'_{-i} = f_{\boldsymbol{\lambda}}(\boldsymbol{s})_{-i} \right\} = \left\{ \boldsymbol{s}'_{-\lambda[i]} \in \mathcal{S}_{-\lambda[i]} \mid \boldsymbol{s}'_{-\lambda[i]} = \boldsymbol{s}_{-\lambda[i]} \right\}.$$

Then, given any signal profile $\boldsymbol{s} \in \mathcal{S}$ and any permuation $\boldsymbol{\lambda} \in \Lambda$, we have that $\forall i \in \mathcal{N}$, $\forall \varphi \in \Phi$

$$\begin{aligned}
d_i^{\varphi}[f_{\boldsymbol{\lambda}}(\boldsymbol{s}), a] &= \sum_{\substack{\boldsymbol{s}' \in \mathcal{S} \\ f_{\boldsymbol{\lambda}}(\boldsymbol{s})_{-i} = \boldsymbol{s}'_{-i} \\ f_{\boldsymbol{\lambda}}(\boldsymbol{s})_i = \varphi(s'_i)}} d_i[\boldsymbol{s}', a] \\
&= \sum_{\substack{s'_{\lambda[i]} \in S \\ \varphi(s'_{\lambda[i]}) = s_{\lambda[i]}}} \sum_{\substack{\boldsymbol{s}'_{-\lambda[i]} \in \mathcal{S}_{-\lambda[i]} \\ \boldsymbol{s}'_{-\lambda[i]} = \boldsymbol{s}_{-\lambda[i]}}} d_i[(s'_{-\lambda[i]}, \boldsymbol{s}'_{-\lambda[i]}), a] \\
&= \sum_{\substack{\boldsymbol{s}' \in \mathcal{S} \\ \boldsymbol{s}_{-\lambda[i]} = \boldsymbol{s}'_{-\lambda[i]} \\ s_{\lambda[i]} = \varphi(s'_{\lambda[i]})}} d_i[\boldsymbol{s}', a] \\
&= d_{\lambda[i]}^{\varphi}[\boldsymbol{s}, a].
\end{aligned} \tag{7}$$

Then, it holds that

$$\begin{aligned}
U_i^{\varphi}(\boldsymbol{x}_{\boldsymbol{\lambda}}) &= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x_{\boldsymbol{\lambda}}[\boldsymbol{s}, a] d_i^{\varphi}[\boldsymbol{s}, a] \\
&= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x^{\star}[f_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{s}), a] d_i^{\varphi}[\boldsymbol{s}, a] \\
&= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x^{\star}[\boldsymbol{s}, a] d_i^{\varphi}[f_{\boldsymbol{\lambda}}(\boldsymbol{s}), a] \\
&= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} x^{\star}[\boldsymbol{s}, a] d_{\lambda[i]}^{\varphi}[\boldsymbol{s}, a] \\
&= U_{\lambda[i]}^{\varphi}(\boldsymbol{x}^{\star}),
\end{aligned} \tag{8}$$

where the first equation follows from the definition of $\boldsymbol{x}_{\boldsymbol{\lambda}}$, the second and third equations follow from the fact that the permutation function is bijective and the last equation follows from Equation (7).

Hence, we can conclude that, $\forall i \in \mathcal{N}$, $\forall \varphi \in \Phi$ and $\forall \boldsymbol{\lambda} \in \Lambda$,

$$U_i^{\varphi}(\boldsymbol{x}_{\boldsymbol{\lambda}}) - U_i(\boldsymbol{x}_{\boldsymbol{\lambda}}) = U_{\lambda[i]}^{\varphi}(\boldsymbol{x}^{\star}) - U_i(\boldsymbol{x}^{\star}) \leq \varepsilon$$

where the first equation follows from Equations (6) and (8), and the last inequality follows from the fact that $\boldsymbol{x}^{\star} \in \mathcal{X}_{\varepsilon}$.

This concludes the proof. □

We are now ready to prove Theorem 3.1

Theorem 3.1. *For any $\varepsilon \geq 0$, there exists a symmetric mechanism $\boldsymbol{x}^{\circ} \in \mathcal{X}^{\circ}$ such that*

$$\boldsymbol{x}^{\circ} \in \operatorname*{argmax}_{\boldsymbol{x} \in \mathcal{X}_{\varepsilon}} U_{\mathrm{R}}(\boldsymbol{x}).$$

Proof. Let $\boldsymbol{x}^\star \in \mathcal{X} \in \operatorname{argmax}_{\boldsymbol{x}\in\mathcal{X}_\varepsilon} U_{\mathrm{R}}(\boldsymbol{x})$. Define, for each permutation $\boldsymbol{\lambda} \in \Lambda$, the permutation strategy $\boldsymbol{x}_{\boldsymbol{\lambda}}$ such that $\forall \boldsymbol{s} \in \mathcal{S}$, $\forall a \in \mathcal{A}$, $x^\star[\boldsymbol{s}, a] = x_{\boldsymbol{\lambda}}[f_{\boldsymbol{\lambda}}(\boldsymbol{s}), a]$. The set of all possible permutation strategies is the following

$$\mathcal{X}^\Lambda = \{\boldsymbol{x}_{\boldsymbol{\lambda}} \mid \boldsymbol{\lambda} \in \Lambda\}.$$

Let $\overline{\boldsymbol{x}}$ be such that:

$$\overline{\boldsymbol{x}} = \frac{1}{|\Lambda|} \sum_{\boldsymbol{\lambda}\in\lambda} \boldsymbol{x}_{\boldsymbol{\lambda}}.$$

By convexity and by Lemma B.1, $\overline{\boldsymbol{x}} \in \operatorname{argmax}_{\boldsymbol{x}\in\mathcal{X}_\varepsilon} U_{\mathrm{R}}(\boldsymbol{x})$. In order to conclude the proof, we need to show that $\forall \boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}$ such that $\boldsymbol{s} \bowtie \boldsymbol{s}'$, it holds that

$$\overline{x}[\boldsymbol{s}, a] = \overline{x}[\boldsymbol{s}', a] \quad \forall a \in \mathcal{A}.$$

Fix any $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}$ such that $\boldsymbol{s} \bowtie \boldsymbol{s}'$. By definition of permutation, it holds that:

$$\{f_{\boldsymbol{\lambda}}(\boldsymbol{s}) \mid \boldsymbol{\lambda} \in \Lambda\} = \left\{f_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{s}) \mid \boldsymbol{\lambda} \in \Lambda\right\} = \left\{f_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{s}') \mid \boldsymbol{\lambda} \in \Lambda\right\} = \{f_{\boldsymbol{\lambda}}(\boldsymbol{s}') \mid \boldsymbol{\lambda} \in \Lambda\} = \{\boldsymbol{s}'' \in \mathcal{S} \mid \boldsymbol{s}'' \bowtie \boldsymbol{s}\}.$$

Hence, $\forall \boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}$ such that $\boldsymbol{s} \bowtie \boldsymbol{s}'$, and $\forall a \in \mathcal{A}$,

$$\begin{aligned}
\overline{x}[\boldsymbol{s}, a] &= \frac{1}{|\Lambda|} \sum_{\boldsymbol{\lambda}\in\Lambda} x_{\boldsymbol{\lambda}}[\boldsymbol{s}, a] \\
&= \frac{1}{|\Lambda|} \sum_{\boldsymbol{\lambda}\in\Lambda} x^\star[f_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{s}), a] \\
&= \frac{1}{|\Lambda|} \sum_{\boldsymbol{\lambda}\in\Lambda} x^\star[f_{\boldsymbol{\lambda}}^{-1}(\boldsymbol{s}'), a] \\
&= \frac{1}{|\Lambda|} \sum_{\boldsymbol{\lambda}\in\Lambda} x_{\boldsymbol{\lambda}}[\boldsymbol{s}', a] \\
&= \overline{x}[\boldsymbol{s}', a].
\end{aligned}$$

This concludes the proof. □

B.2. Additional Proofs

Lemma 3.1. *The polytope Ξ admits a description of size polynomial in the dimension of the game.*

Proof. To prove that the polytope Ξ admits a polynomially sized representation we provide an upper bound on the number of variables and constraints needed to describe Ξ.

Bound on the number of variables. By definition, we have that $\Xi \subset \mathbb{R}_{\geq 0}^{|\mathcal{C}||\mathcal{A}|}$. Hence, in order to bound the dimension of Ξ, we need to bound the number of elements in $\mathcal{C}$. In particular, we have that each element $c \in \mathcal{C}$ can be uniquely identified by a vector $\boldsymbol{c} \in \mathbb{N}^{|S|}$ in which the i-th entry $c[i]$ defines the number of occurrences of signal $s_i \in S$. Thus, we can write the following:

$$c[i] \leq n \quad \forall i \in [|S|].$$

As a consequence of this, it is possible to upper bound the maximum number of elements in $\mathcal{C}$ as

$$|\mathcal{C}| < n^{|S|},$$

hence, the number m_1 of variables is

$$m_1 = |\mathcal{C}| \cdot |\mathcal{A}| < |\mathcal{A}| n^{|S|}.$$

Bound on the number of constraints. The constraints that characterize the polytope Ξ are exactly the simplex constraints associated to each $c \in \mathcal{C}$. Formally:

$$\begin{aligned} \xi[c,a] &\geq 0 \quad \forall c \in \mathcal{C},\, \forall a \in \mathcal{A}, \\ \sum_{a \in \mathcal{A}} \xi[c,a] &= 1 \quad \forall c \in \mathcal{C}. \end{aligned}$$

Thus, the number m_2 of constraints is

$$m_2 = |\mathcal{C}| \cdot |\mathcal{A}| + |\mathcal{C}| < n^{|S|} \left(|\mathcal{A}| + 1\right).$$

The Lemma follows from the fact that, by Assumption 1, $|S|$ is an absolute constant. □

Lemma 3.2. *For each $\varphi \in \Phi$ and $i \in \mathcal{N}$, let $\boldsymbol{A}_i^{\varphi} \in \mathbb{R}^{|\mathcal{C}||\mathcal{A}| \times |\mathcal{C}||\mathcal{A}|}$ be the matrix such that $\forall c, c' \in \mathcal{C}$, $\forall a, a' \in \mathcal{A}$,*

$$A_i^{\varphi}[(c,a),(c',a')] = \begin{cases} 0 & \text{if } a \neq a' \\ \frac{\sum_{s \in c'} \mathbb{1}[(\varphi(s_i), \boldsymbol{s}_{-i}) \in c]}{|c'|} & \text{otherwise}. \end{cases}$$

Then, it holds $U_i^{\varphi}(\boldsymbol{\xi}) = \boldsymbol{\xi}^{\top} \boldsymbol{A}_i^{\varphi} \boldsymbol{r}_{\mathrm{S}}$ for all $i \in \mathcal{N}$.

Proof. For each $i \in \mathcal{N}$ and for each deviation function $\varphi \in \Phi$, from the definition of U_i^{φ} it follows that

$$\begin{aligned} U_i^{\varphi}(\boldsymbol{\xi}) &= \sum_{\substack{\boldsymbol{s} \in \mathcal{S} \\ a \in \mathcal{A}}} \xi[c((\varphi(s_i), \boldsymbol{s}_{-i})), a] \sum_{\theta \in \Theta} p(\theta) u_{\mathrm{S}}(a,\theta) \psi_{\mathrm{S}}(\boldsymbol{s}|\theta) \\ &= \sum_{\substack{c \in \mathcal{C} \\ a \in \mathcal{A}}} \sum_{\boldsymbol{s} \in c} \xi[c((\varphi(s_i), \boldsymbol{s}_{-i})), a] \sum_{\theta \in \Theta} p(\theta) u_{\mathrm{S}}(a,\theta) \psi_{\mathrm{S}}(\boldsymbol{s}|\theta) \\ &= \sum_{\substack{c \in \mathcal{C} \\ a \in \mathcal{A}}} \sum_{\theta \in \Theta} p(\theta) u_{\mathrm{S}}(a,\theta) \sum_{\boldsymbol{s} \in c} \xi[c((\varphi(s_i), \boldsymbol{s}_{-i})), a] \psi_{\mathrm{S}}(\boldsymbol{s}|\theta). \end{aligned}$$

Let us recall that $\forall \theta \in \Theta$ and $\forall \boldsymbol{s} \bowtie \boldsymbol{s}'$, $\psi_{\mathrm{S}}(\boldsymbol{s}|\theta) = \psi_{\mathrm{S}}(\boldsymbol{s}'|\theta)$. Hence, since $\forall c \in \mathcal{C}$, $\boldsymbol{s}, \boldsymbol{s}' \in c$ if and only if $\boldsymbol{s} \bowtie \boldsymbol{s}'$, with a slight abuse of notation, we can write $\psi_{\mathrm{S}}(\boldsymbol{s}|\theta) = \psi_{\mathrm{S}}(c(\boldsymbol{s})|\theta)$. Thus,

$$U_i^{\varphi}(\boldsymbol{\xi}) = \sum_{\substack{c \in \mathcal{C} \\ a \in \mathcal{A}}} \sum_{\theta \in \Theta} p(\theta) u_{\mathrm{S}}(a,\theta) \psi_{\mathrm{S}}(c|\theta) \sum_{\boldsymbol{s} \in c} \xi[c((\varphi(s_i), \boldsymbol{s}_{-i})), a]. \tag{9}$$

By definition of $\boldsymbol{r}_{\mathrm{S}}$, we have that $\forall c \in \mathcal{C}$, $\forall a \in \mathcal{A}$,

$$\begin{aligned} r_{\mathrm{S}}[c,a] &= \sum_{\boldsymbol{s} \in c} \sum_{\theta \in \Theta} p(\theta) u_{\mathrm{S}}(a,\theta) \psi_{\mathrm{S}}(\boldsymbol{s}|\theta) \\ &= \sum_{\theta \in \Theta} p(\theta) u_{\mathrm{S}}(a,\theta) \sum_{\boldsymbol{s} \in c} \psi_{\mathrm{S}}(\boldsymbol{s}|\theta) \\ &= |c| \sum_{\theta \in \Theta} p(\theta) u_{\mathrm{S}}(a,\theta) \psi_{\mathrm{S}}(c|\theta). \end{aligned}$$

Substituting into Equation (9), we get

$$U_i^{\varphi}(\boldsymbol{\xi}) = \sum_{\substack{c \in \mathcal{C} \\ a \in \mathcal{A}}} r_{\mathrm{S}}[c,a] \frac{\sum_{\boldsymbol{s} \in c} \xi[c((\varphi(s_i), \boldsymbol{s}_{-i})), a]}{|c|}. \tag{10}$$

Now, notice that the term $\frac{1}{|c|}\sum_{\boldsymbol{s}\in c}\xi[c((\varphi(s_i),\boldsymbol{s}_{-i})),a]$ can be rewritten as

$$\begin{aligned}\frac{1}{|c|}\sum_{\boldsymbol{s}\in c}\xi[c((\varphi(s_i),\boldsymbol{s}_{-i})),a] &= \sum_{c'\in\mathcal{C}}\xi[c',a]\sum_{\boldsymbol{s}\in c}\frac{\mathbb{1}[(\varphi(s_i),\boldsymbol{s}_{-i})\in c']}{|c|}\\ &= \sum_{c'\in\mathcal{C}}\xi[c',a]A_i^{\varphi}[(c',a),(c,a)]\\ &= \sum_{\substack{c'\in\mathcal{C}\\ a'\in\mathcal{A}}}\xi[c',a']A_i^{\varphi}[(c',a'),(c,a)],\end{aligned}\tag{11}$$

where the last equation follows from the fact that $A[(c',a'),(c,a)]=0,\forall a'\neq a$.

Hence, by plugging Equation (11) into Equation (10), it follows that

$$\begin{aligned}U_i^{\varphi}(\boldsymbol{\xi}) &= \sum_{\substack{c\in\mathcal{C}\\ a\in\mathcal{A}}} r_{\text{S}}[c,a]\sum_{\substack{c'\in\mathcal{C}\\ a'\in\mathcal{A}}}\xi[c',a']A_i^{\varphi}[(c',a'),(c,a)]\\ &= \boldsymbol{\xi}^{\top}\boldsymbol{A}_i^{\varphi}\boldsymbol{r}_{\text{S}},\end{aligned}$$

which gives the result. □

Lemma 3.3. *For any $i,j\in\mathcal{N}$ and any $\varphi\in\Phi$ it holds that $\boldsymbol{A}_i^{\varphi}=\boldsymbol{A}_j^{\varphi}$.*

Proof. Notice that each class $c\in\mathcal{C}$ contains a set of symmetric signal profiles. Hence, we can define the function $\lambda: c\to c$ such that $\forall\boldsymbol{s}\in c$, $\lambda(\boldsymbol{s})=\boldsymbol{s}'$, where $s_i=s'_j$, $s_j=s'_i$ and $s_k=s'_k$, $\forall k\neq i,j$ and obtain the following:

$$\begin{aligned}\frac{\sum_{s\in c}\mathbb{1}[(\varphi(s_i),\boldsymbol{s}_{-i})\in c']}{|c|} &= \frac{\sum_{s\in c}\mathbb{1}[(\varphi(\lambda(\boldsymbol{s})_j),\lambda(\boldsymbol{s})_{-j})\in c']}{|c|}\\ &= \frac{\sum_{s'\in c}\mathbb{1}[(\varphi(s'_j),s'_{-j})\in c']}{|c|},\end{aligned}$$

which gives the result. □

Proposition 3.1. *For any deviation function $\varphi\in\Phi$, the coefficients of the matrix $\boldsymbol{A}_{\text{S}}^{\varphi}$ and those of the utility vectors $\boldsymbol{r}_{\text{R}}$ and $\boldsymbol{r}_{\text{S}}$ can be computed in time polynomial in the game size.*

Proof. As a preliminary step towards the proof of this result, we first show how to efficiently determine the number of signal profiles belonging to a given class $c\in\mathcal{C}$, and then proceed to bounding the complexity of specifying the utility vectors $\boldsymbol{r}_{\text{R}}$, $\boldsymbol{r}_{\text{S}}$ and the deviation matrix $\boldsymbol{A}_{\text{S}}^{\varphi}$.

Let us remark that, similarly to the proof of Lemma 3.1, each element $c\in\mathcal{C}$ can be uniquely identified by a vector $\boldsymbol{c}\in\mathbb{N}^S$ in which the entry $c[s]$ defines the number of occurrences of signal $s\in S$. Hence, the number of different signal profiles in a given class c can be computed via the multinomial coefficient

$$|c| = \binom{n}{\boldsymbol{c}} = \frac{n!}{\prod_{s\in S}c[s_{\text{S}}]!}.\tag{12}$$

Computation of $\boldsymbol{r}_{\text{R}}$ and $\boldsymbol{r}_{\text{S}}$. Notice that, by definition, each $c\in\mathcal{C}$ contains signal profiles that are symmetric between them. Recalling the definition of symmetric signal profile, we can write the following:

$$\psi(\boldsymbol{s}|\theta)=\psi(\boldsymbol{s}'|\theta)\quad\forall\theta\in\Theta,\ \forall\boldsymbol{s}\bowtie\boldsymbol{s}'.$$

Thus, with a slight abuse of notation, since all signal profiles belonging to the same class share the same sampling probability, we can define $\psi(c|\theta):=\psi(\boldsymbol{s}|\theta)$, $\forall\theta\in\Theta$, $\forall\boldsymbol{s}\in c$. Hence, for each $j\in\{\text{R},\text{S}\}$, the utility vectors can be expressed as

$$r_j[c,a]=|c|\sum_{\theta\in\Theta}p(\theta)\psi(c|\theta)u_i(a,\theta)\quad\forall c\in\mathcal{C},\ \forall a\in\mathcal{A}.$$

Note that each entry of such vectors can now be computed efficiently without iterating on all signal profiles belonging to each class, simply by exploiting Equation (12). This gives the desired result.

Computation of $\boldsymbol{A}_{\text{S}}^{\varphi}$. To show that the coefficients of the matrix $\boldsymbol{A}_{\text{S}}^{\varphi}$ can be computed efficiently, we show that it is possible to determine in an efficient way the term $\sum_{\boldsymbol{s}\in c} \mathbb{1}[(\varphi(s_i), \boldsymbol{s}_{-i}) \in c']$ for any two $c, c' \in \mathcal{C}$.

Let $\boldsymbol{c}, \boldsymbol{c}' \in \mathbb{N}^{|S|}$ be the vectors describing classes c and c', respectively. Trivially, if $||\boldsymbol{c} - \boldsymbol{c}'||_1 > 2$, then, $\forall \varphi \in \Phi$, $\forall \boldsymbol{s} \in c$, $(\varphi(s_i), \boldsymbol{s}_{-i}) \notin c'$, since it is never the case that, by means of a single player deviation, a signal profile $\boldsymbol{s} \in c$ is transformed into a signal profile $\boldsymbol{s}' \in c'$. If $||\boldsymbol{c} - \boldsymbol{c}'||_1 \leq 2$, then we can be in one of two cases, namely $||\boldsymbol{c} - \boldsymbol{c}'||_1 = 2$ or $||\boldsymbol{c} - \boldsymbol{c}'||_1 = 0$. Let us first consider the case $||\boldsymbol{c} - \boldsymbol{c}'||_1 = 2$. Intuitively, $||\boldsymbol{c} - \boldsymbol{c}'||_1 = 2$ implies that there exist two signals s and s' such that $c[s] = c'[s] + 1$ and $c[s'] = c'[s'] - 1$. For notational convenience, let $\iota(\boldsymbol{c}, \boldsymbol{c}') = s$. Thus, we have $\sum_{\boldsymbol{s}\in c} \mathbb{1}[(\varphi(s_i), \boldsymbol{s}_{-i}) \in c'] \neq 0$ if and only if $\varphi(s) = s'$. In particular, when $\varphi(\iota(\boldsymbol{c}, \boldsymbol{c}')) = s'$, the number of tuples $\boldsymbol{s} \in c$ that can be transformed in tuples $\boldsymbol{s}' \in c'$ corresponds to the number of tuples that have element $\iota(\boldsymbol{c}, \boldsymbol{c}')$ as the i-th element, which can be computed as

$$\sum_{\boldsymbol{s}\in c} \mathbb{1}[(\varphi(s_i), \boldsymbol{s}_{-i}) \in c'] = \frac{(n-1)!}{(c[\iota(\boldsymbol{c}, \boldsymbol{c}')] - 1)! \prod_{s'' \neq \iota(\boldsymbol{c}, \boldsymbol{c}')} c[s'']!}.$$

Finally, when $||\boldsymbol{c} - \boldsymbol{c}'||_1 = 0$, *i.e.,* when $\boldsymbol{c} = \boldsymbol{c}'$, the number $\sum_{\boldsymbol{s}\in c} \mathbb{1}[(\varphi(s_i), \boldsymbol{s}_{-i}) \in c']$ is the number of tuples $\boldsymbol{s} \in c$ that have as i-th element a signal s_i such that $\varphi(s_i) = s_i$. Formally, this can be expressed as

$$\sum_{\boldsymbol{s}\in c} \mathbb{1}[(\varphi(s_i), \boldsymbol{s}_{-i}) \in c'] = \sum_{\substack{s\in S:\\ \varphi(s)=s,\\ c[s]>0}} \frac{(n-1)!}{(c[s] - 1)! \prod_{s'' \neq s} c[s'']!}.$$

In conclusion, we have that

$$\sum_{\boldsymbol{s}\in c} \mathbb{1}[(\varphi(s_i), \boldsymbol{s}_{-i}) \in c'] = \begin{cases} 0 & \text{if } ||\boldsymbol{c} - \boldsymbol{c}'||_1 > 2 \\ \frac{(n-1)!}{(c[\iota(\boldsymbol{c},\boldsymbol{c}')]-1)! \prod_{s'' \neq \iota(\boldsymbol{c},\boldsymbol{c}')} c[s'']!} & \text{if } ||\boldsymbol{c} - \boldsymbol{c}'||_1 = 2 \\ \sum_{\substack{s\in S:\\ \varphi(s)=s,\\ c[s]>0}} \frac{(n-1)!}{(c[s]-1)! \prod_{s'' \neq s} c[s'']!} & \text{otherwise.} \end{cases}$$

This concludes the proof. □

C. Proofs Omitted from Section 4

Theorem 4.1. *There exists a constant $\delta > 0$ such that no algorithm can guarantee to output a sequence of mechanisms $\boldsymbol{\xi}^1, ..., \boldsymbol{\xi}^T$ such that, with probability at least $1 - \delta$ all mechanisms are IC and $R^T = o(T)$.*

Proof. Let us define two instances X and Y of a game between a receiver R and a single sender S. The set of states of nature is composed by two states $\Theta = \{\theta_1, \theta_2\}$, the set of signals is $S = \{s_1, s_2\}$ and the set of actions is $\mathcal{A} = \{a, b\}$. Both instances share the same prior $p(\theta_1) = p(\theta_2) = 1/2$ and utility functions. In particular, given $\varepsilon > 0$, the utility functions are specified such that $u_{\text{R}}(a, \theta_1) = u_{\text{S}}(a, \theta_1) = 1$, $u_{\text{R}}(a, \theta_2) = u_{\text{S}}(a, \theta_2) = u_{\text{R}}(b, \theta_1) = u_{\text{S}}(b, \theta_1) = 0$, $u_{\text{R}}(b, \theta_2) = 1$ and $u_{\text{S}}(b, \theta_2) = 2\varepsilon$. Furthermore, the two instances differ for the signaling schemes, which are defined as

$$\text{X} = \begin{cases} \psi_{\text{S}}(s_1|\theta_1) = 1 - 4\varepsilon \\ \psi_{\text{S}}(s_2|\theta_1) = 4\varepsilon \\ \psi_{\text{S}}(s_1|\theta_2) = 0 \\ \psi_{\text{S}}(s_2|\theta_2) = 1, \end{cases} \qquad \text{Y} = \begin{cases} \psi_{\text{S}}(s_1|\theta_1) = 1 - 2\varepsilon \\ \psi_{\text{S}}(s_2|\theta_1) = 2\varepsilon \\ \psi_{\text{S}}(s_1|\theta_2) = 0 \\ \psi_{\text{S}}(s_2|\theta_2) = 1. \end{cases}$$

First, let us consider instance X. Fix a deviation function $\varphi \in \Phi$ such that $\varphi(s_1) = s_1$ and $\varphi(s_2) = s_1$. By simple calculations it is possible to show that for each mechanism $\boldsymbol{\xi} \in \Xi$, the expected utility of the sender when she deviates according to φ is the following

$$U_{\text{S}}^{\varphi}(\boldsymbol{\xi}) = \left(\frac{1}{2} - 2\varepsilon\right) \xi[s_1, a] + \varepsilon \xi^t[s_1, a] + \varepsilon.$$

Similarly, the expected utility of the sender when she communicates truthfully the signal observed is

$$U_{\text{S}}(\boldsymbol{\xi}) = \left(\frac{1}{2} - 2\varepsilon\right) \xi[s_1, a] - \varepsilon \xi[s_2, b] + 2\varepsilon.$$

Given $T \in \mathbb{N}_{>0}$ and $\delta \in (0,1)$, let us assume that, at each round $t \in [T]$ and with probability greater than $1-\delta$, the mechanism $\boldsymbol{\xi}^t$ selected by the the receiver in instance X is IC. Hence, we can write the following

$$\mathbb{P}_{\text{X}}\left(\sum_{t\in[T]}\left[U_{\text{S}}^{\varphi}(\boldsymbol{\xi}^t) - U_{\text{S}}(\boldsymbol{\xi}^t)\right] \leq 0\right) \geq 1-\delta,$$

which yields

$$\mathbb{P}_{\text{X}}\left(\sum_{t\in[T]}\left[\varepsilon\xi^t[s_1,a] + \varepsilon\xi^t[s_2,b] - \varepsilon\right] \leq 0\right) \geq 1-\delta,$$

where the subscript X indicates the probability measure associated to instance X. The, we can leverage the Pinsker's inequality to state the following:

$$\mathbb{P}_{\text{Y}}\left(\sum_{t\in[T]}\left[\varepsilon\xi^t[s_1,a] + \varepsilon\xi^t[s_2,b] - \varepsilon\right] \leq 0\right) \geq \mathbb{P}_{\text{X}}\left(\sum_{t\in[T]}\left[\varepsilon\xi^t[s_1,a] + \varepsilon\xi^t[s_2,b] - \varepsilon\right] \leq 0\right) - \sqrt{\frac{1}{2}\mathcal{K}(\text{X},\text{Y})}, \tag{13}$$

where $\mathcal{K}(\text{X},\text{Y})$ is the Kullback-Leibler divergence between instance X and instance Y.

It is easy to check that the two instances differ only for the probability distributions of the receiver's rewards that arise when the signal sampled is s_2. In particular, when the signal sampled is s_2 and the action selected is a, the utility of the receiver follows a Bernoulli distribution $\mathcal{B}\left(\frac{2\varepsilon}{1/2+2\varepsilon}\right)$ in instance X, and a Bernoulli distribution $\mathcal{B}\left(\frac{\varepsilon}{1/2+\varepsilon}\right)$ in instance Y. Similarly, when the action selected is b, the receiver's utility follows a Bernoulli distribution $\mathcal{B}\left(\frac{1/2}{1/2+2\varepsilon}\right)$ in instance X, and a Bernoulli distribution $\mathcal{B}\left(\frac{1/2}{1/2+\varepsilon}\right)$ in instance Y. Then, exploiting the fact that $\mathcal{K}\left(\mathcal{B}(p),\mathcal{B}(q)\right) = \mathcal{K}(\mathcal{B}(1-p),\mathcal{B}(1-q))$, we can apply the KL decomposition Theorem (Lattimore & Szepesvári, 2020) to state the following

$$\begin{aligned}
\mathcal{K}(\text{X},\text{Y}) &= \sum_{\substack{t\in[T]\\ s^t=s_2}}\left[\mathbb{E}_{\text{X}}\left[\xi^t[s_2,a]\right] + \mathbb{E}_{\text{X}}\left[\xi^t[s_2,b]\right]\right]\mathcal{K}\left(\mathcal{B}\left(\frac{2\varepsilon}{1/2+2\varepsilon}\right),\mathcal{B}\left(\frac{\varepsilon}{1/2+\varepsilon}\right)\right)\\
&\leq \sum_{\substack{t\in[T]\\ s^t=s_2}} 2\varepsilon\\
&\leq 2\varepsilon T.
\end{aligned}$$

Setting $\varepsilon = (0.01)^2/T$ and plugging into Equation (13), we get

$$\mathbb{P}_{\text{Y}}\left(\sum_{t\in[T]}\left[\varepsilon\xi^t[s_1,a] + \varepsilon\xi^t[s_2,b] - \varepsilon\right] \leq 0\right) \geq 0.99-\delta. \tag{14}$$

To conclude the proof we show that, with high probability, the receiver suffers linear regret in instance Y. It is easy to check that the optimal IC mechanism in instance Y is the mechanism $\boldsymbol{\xi}^\star$ such that $\xi^\star[s_1,a]=1$ and $\xi^\star[s_2,b]=1$, yielding an expected utility to the receiver $U_{\text{R}}(\boldsymbol{\xi}^\star) = 1-\varepsilon$. Hence, the cumulative regret can be expressed as

$$\begin{aligned}
R^T &= \sum_{t\in[T]}\left[1-\varepsilon-\left(\frac{1}{2}-\varepsilon\right)\left(\xi^t[s_1,a]+\xi^t[s_2,b]\right)-\varepsilon\right]\\
&= \sum_{t\in[T]}\left[1-2\varepsilon-\left(\frac{1}{2}-\varepsilon\right)\left(\xi^t[s_1,a]+\xi^t[s_2,b]\right)\right].
\end{aligned}$$

From Equation (14) it follows that with probability grater than $0.99-\delta$

$$\sum_{t\in[T]}\left[\xi^t[s_1,a]+\xi^t[s_2,b]\right] \leq T,$$

thus, with probability at least $0.99 - \delta$, it holds that

$$R^T \geq (1-2\varepsilon)\,T - \frac{1}{2}\,(1-2\varepsilon)\,T = \frac{1}{2}\,(1-2\varepsilon)\,T,$$

which gives the result. □

D. Proofs Omitted from Section 5

Lemma 5.1. *For all $t \in [T]$ and $i \in \{\text{R}, \text{S}\}$, $\hat{\boldsymbol{r}}_i^t$ is an unbiased estimator of $\boldsymbol{r}_i$.*

Proof. First, notice that for each $t \in [T]$, for all $c \in \mathcal{C}$ and for all $\theta \in \Theta$, it holds that

$$\mathbb{E}\left[\mathbb{1}\left[c(\boldsymbol{s}^t) = c, \theta^t = \theta\right]\right] = \mathbb{P}\left(c, \theta\right) = p(\theta)\mathbb{P}\left(c|\theta\right) = \sum_{\boldsymbol{s} \in c} p(\theta)\psi_{\text{S}}(\boldsymbol{s}|\theta),$$

where $\mathbb{P}\left(c, \theta\right)$ denotes the joint probability of sampling the state of nature θ and a signal profile belonging to class c, and $\mathbb{P}\left(c|\theta\right)$ is the probability of sampling a signal belonging to class c given that the state of nature sampled is θ. Then, for all $i \in \{\text{R}, \text{S}\}$, for all $t \in [T]$, for all $c \in \mathcal{C}$ and for all $a \in \mathcal{A}$, it holds that

$$\begin{aligned}
\mathbb{E}[\hat{r}_i^t[c, a]] &= \mathbb{E}\left[\frac{1}{t}\sum_{\theta \in \Theta} u_i(a, \theta) \sum_{\tau \in [t]} \mathbb{1}\left[c(\boldsymbol{s}^\tau) = c, \theta^\tau = \theta\right]\right] \\
&= \sum_{\theta \in \Theta} u_i(a, \theta) \frac{1}{t} \sum_{\tau \in [t]} \mathbb{E}\left[\mathbb{1}\left[c(\boldsymbol{s}^\tau) = c, \theta^\tau = \theta\right]\right] \\
&= \sum_{\theta \in \Theta} \sum_{\boldsymbol{s} \in c} u_i(a, \theta) p(\theta) \psi_{\text{S}}(\boldsymbol{s}|\theta) \\
&= r_i[c, a].
\end{aligned}$$

This concludes the proof. □

Lemma 5.2. *For any $\delta \in (0, 1)$, the following holds*

$$\mathbb{P}\left(\mathcal{E}_\delta^{\text{F}}\right) \geq 1 - \delta.$$

Proof. For each $\tau \in [T]$, for each $i \in \{\text{R}, \text{S}\}$, for each $c \in \mathcal{C}$ and for each $a \in \mathcal{A}$, let $\tilde{r}_i^\tau[c, a] := \sum_{\theta \in \Theta} u_i(a, \theta) \mathbb{1}\left[c(\boldsymbol{s}^\tau) = c, \theta^\tau = \theta\right]$. Notice that $\mathbb{E}\left[\tilde{\boldsymbol{r}}_i^\tau\right] = \boldsymbol{r}_i$, and $\hat{\boldsymbol{r}}_i^t = \frac{1}{t}\sum_{\tau \in [t]} \tilde{\boldsymbol{r}}_i^\tau$, for each $t \in [T]$. Then, since it holds that $0 \preceq \tilde{\boldsymbol{r}}_i^\tau \preceq 1$, where $\preceq$ denotes the element-wise $\leq$ operator, by Hoeffding's inequality we have that: for each $i \in \{\text{R}, \text{S}\}$, for each $t \in [T]$, for each $c \in \mathcal{C}$ and for each $a \in \mathcal{A}$ it holds that

$$\mathbb{P}\left(|r_i[c, a] - \hat{r}_i^t[c, a]| \geq \sqrt{\frac{\log(4T|\mathcal{C}||\mathcal{A}|/\delta)}{2(t-1)}}\right) \leq \frac{\delta}{2T|\mathcal{C}||\mathcal{A}|}.$$

The result follows by applying a union bound. □

Lemma D.1. *Let $\boldsymbol{r}, \boldsymbol{r}' \in \mathbb{R}^{|\mathcal{C}||\mathcal{A}|}$ be two vectors such that $||\boldsymbol{r} - \boldsymbol{r}'||_\infty \leq \varepsilon$. Then, for each $\boldsymbol{\xi} \in \Xi$ it holds that:*

$$|\boldsymbol{\xi}^\top \boldsymbol{r} - \boldsymbol{\xi}^\top \boldsymbol{r}'| \leq |\mathcal{C}|\varepsilon.$$

Proof. By Cauchy-Schwarz inequality it follows that:

$$|\boldsymbol{\xi}^\top(\boldsymbol{r} - \boldsymbol{r}')| \leq ||\boldsymbol{\xi}||_1 ||\boldsymbol{r} - \boldsymbol{r}'||_\infty.$$

The result follows since by assumption $||\boldsymbol{r} - \boldsymbol{r}'||_\infty \leq \varepsilon$ and since

$$||\boldsymbol{\xi}||_1 = \sum_{c \in \mathcal{C}} \sum_{a \in \mathcal{A}} \xi[c, a] = \sum_{c \in \mathcal{C}} 1 = |\mathcal{C}|.$$

□

Lemma D.2. *Let $\boldsymbol{r}, \boldsymbol{r}' \in \mathbb{R}^{|\mathcal{C}||\mathcal{A}|}$ be two vectors such that $||\boldsymbol{r} - \boldsymbol{r}'||_\infty \leq \varepsilon$. Then, for each $\varphi \in \Phi$ and $\boldsymbol{\xi} \in \Xi$ it holds that*

$$\left|\boldsymbol{\xi}^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r} - \boldsymbol{\xi}^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}'\right| \leq |\mathcal{C}|\varepsilon.$$

Proof. To prove this Lemma, we first show that for any $\boldsymbol{\xi} \in \Xi$ and $\varphi \in \Phi$, there exists a mechanism $\boldsymbol{\xi}' \in \Xi$ such that $\boldsymbol{\xi}^\top \boldsymbol{A}_{\text{S}}^\varphi = \left(\boldsymbol{\xi}'\right)^\top$ and then exploit Lemma D.1 to obtain the desired result.

Notice that, by letting $\boldsymbol{\xi}'$ be such that $\boldsymbol{\xi}^\top \boldsymbol{A}_{\text{S}}^\varphi = \left(\boldsymbol{\xi}'\right)^\top$, we obtain that $\forall c \in \mathcal{C}$ and $a \in \mathcal{A}$,

$$\xi'[c, a] = \frac{1}{|c|} \sum_{\boldsymbol{s} \in c} \xi[c(\psi(s_i), \boldsymbol{s}_{-i}), a].$$

Trivially, $\boldsymbol{\xi}' \in \mathbb{R}_{\geq 0}^{|\mathcal{C}||\mathcal{A}|}$. In order to show that $\boldsymbol{\xi}' \in \Xi$, we need to show that $\forall c \in \mathcal{C}$, $\sum_{a \in \mathcal{A}} \xi'[c, a] = 1$. To this extent, $\forall c \in \mathcal{C}$, we can write the following:

$$\begin{aligned}
\sum_{a \in \mathcal{A}} \xi'[c, a] &= \frac{1}{|c|} \sum_{a \in \mathcal{A}} \sum_{\boldsymbol{s} \in c} \xi[c(\psi(s_i), \boldsymbol{s}_{-i}), a] \\
&= \frac{1}{|c|} \sum_{\boldsymbol{s} \in c} \sum_{a \in \mathcal{A}} \xi[c(\psi(s_i), \boldsymbol{s}_{-i}), a] \\
&= \frac{1}{|c|} \sum_{\boldsymbol{s} \in c} 1 \\
&= 1,
\end{aligned}$$

which gives $\boldsymbol{\xi}' \in \Xi$. Thus,

$$\begin{aligned}
\left|\boldsymbol{\xi}^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r} - \boldsymbol{\xi}^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}'\right| &= \left|\left(\boldsymbol{\xi}'\right)^\top \boldsymbol{r} - \left(\boldsymbol{\xi}'\right)^\top \boldsymbol{r}'\right| \\
&\leq |\mathcal{C}|\varepsilon,
\end{aligned}$$

where the last inequality follows from Lemma D.1.

This concludes the proof. □

Lemma D.3. *Let $\boldsymbol{\xi}^\star$ be an optimal solution to LP$(0, \boldsymbol{r}_{\text{R}}, \boldsymbol{r}_{\text{S}})$. Then, if event $\mathcal{E}_\delta^{\text{F}}$ holds, $\boldsymbol{\xi}^\star$ is a feasible mechanism for LP$(2|\mathcal{C}|\varepsilon_\delta^t, \hat{\boldsymbol{r}}_{\text{R}}^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$ for all $t \in [T]$.*

Proof. Given that event $\mathcal{E}_\delta^{\text{F}}$ holds, we have that $\forall t \in [T]$, $||\boldsymbol{r}_{\text{S}} - \hat{\boldsymbol{r}}_{\text{S}}^t||_\infty \leq \varepsilon_\delta^t$. Thus, for all $t \in [T]$, by Lemma D.1, we have that

$$\left|(\boldsymbol{\xi}^\star)^\top \boldsymbol{r}_{\text{S}} - (\boldsymbol{\xi}^\star)^\top \hat{\boldsymbol{r}}_{\text{S}}^t\right| \leq |\mathcal{C}|\varepsilon_\delta^t.$$

Furthermore, by Lemma D.2, $\forall t \in [T]$ and $\forall \varphi \in \Phi$,

$$\left|\left(\boldsymbol{\xi}^\star\right)^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \left(\boldsymbol{\xi}^\star\right)^\top \boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t\right| \leq |\mathcal{C}|\varepsilon_\delta^t.$$

Hence, $\forall \varphi \in \Phi$,

$$\begin{aligned}
\left(\boldsymbol{\xi}^\star\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t - \hat{\boldsymbol{r}}_{\text{S}}^t\right) &\leq \left(\boldsymbol{\xi}^\star\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right) + 2|\mathcal{C}|\varepsilon_\delta^t \\
&\leq 2|\mathcal{C}|\varepsilon_\delta^t,
\end{aligned}$$

where the last inequality follows from the fact that, since $\boldsymbol{\xi}^\star$ is an optimal solution to LP$(0, \boldsymbol{r}_{\text{R}}, \boldsymbol{r}_{\text{S}})$, $\left(\boldsymbol{\xi}^\star\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right) \leq 0$ for all $\varphi \in \Phi$. This concludes the proof. □

Theorem 5.1. *For any $\delta \in (0, 1)$, with probability at least $1 - \delta$, Algorithm 1 guarantees*

$$\begin{aligned}
R^T &= \mathcal{O}\left(|\mathcal{C}|\sqrt{8T \log(4T|\mathcal{C}||\mathcal{A}|/\delta)}\right), \\
V^T &= \mathcal{O}\left(|\mathcal{C}|\sqrt{16T \log(4T|\mathcal{C}||\mathcal{A}|/\delta)}\right).
\end{aligned}$$

Proof. To prove this Theorem, we first derive the upper bound for the cumulative regret R^T and then show the upper bound for cumulative IC violation V^T. Throughout this proof, we assume that event $\mathcal{E}_\delta^{\text{F}}$ holds (let us remark that, by Lemma 5.2, $\mathbb{P}\left(\mathcal{E}_\delta^{\text{F}}\right) \geq 1-\delta$.

Regret. Recall the definition of cumulative regret:

$$R^T = \sum_{t\in[T]} \left[\left(\boldsymbol{\xi}^\star\right)^\top \boldsymbol{r}_{\text{R}} - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}} \right].$$

Since event $\mathcal{E}_\delta^{\text{F}}$ holds, we have that $\forall t \in [T]$, $||\boldsymbol{r}_{\text{R}} - \hat{\boldsymbol{r}}_{\text{R}}^t||_\infty \leq \varepsilon_\delta^t$. Then $\forall t \in [T]$, by Lemma D.1, we have that

$$\begin{aligned}
(\boldsymbol{\xi}^\star)^\top \boldsymbol{r}_{\text{R}} &\leq (\boldsymbol{\xi}^\star)^\top \hat{\boldsymbol{r}}_{\text{R}}^t + |\mathcal{C}|\varepsilon_\delta^t \\
(\boldsymbol{\xi}^t)^\top \boldsymbol{r}_{\text{R}} &\geq (\boldsymbol{\xi}^t)^\top \hat{\boldsymbol{r}}_{\text{R}}^t - |\mathcal{C}|\varepsilon_\delta^t
\end{aligned}$$

Furthermore, since, by Lemma D.3, $\forall t \in [T]$ $\boldsymbol{\xi}^\star$ is a feasible mechanism for LP$(2|\mathcal{C}|\varepsilon_\delta^t, \hat{\boldsymbol{r}}_{\text{R}}^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$, and since $\boldsymbol{\xi}^t$ is chosen as an optimal solution to the same linear optimization problem, we have that

$$\left(\boldsymbol{\xi}^\star\right)^\top \hat{\boldsymbol{r}}_{\text{R}}^t \leq \left(\boldsymbol{\xi}^t\right)^\top \hat{\boldsymbol{r}}_{\text{R}}^t.$$

By plugging into the definition of cumulative regret we obtain the following:

$$\begin{aligned}
R^T &= \sum_{t\in[T]} \left[\left(\boldsymbol{\xi}^\star\right)^\top \boldsymbol{r}_{\text{R}} - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}} \right] \\
&\leq \sum_{t\in[T]} \left[\left(\boldsymbol{\xi}^\star\right)^\top \hat{\boldsymbol{r}}_{\text{R}}^t - \left(\boldsymbol{\xi}^t\right)^\top \hat{\boldsymbol{r}}_{\text{R}}^t + 2|\mathcal{C}|\varepsilon_\delta^t \right] \\
&\leq \sum_{t\in[T]} 2|\mathcal{C}|\varepsilon_\delta^t \\
&\leq |\mathcal{C}|\sqrt{8T\log(4T|\mathcal{C}||\mathcal{A}|/\delta)},
\end{aligned}$$

where the last inequality follows from $\sum_{t\in[T]} \frac{1}{\sqrt{t}} \leq 2\sqrt{T}$.

This concludes the first part of the proof.

IC violation. Recall the definition of cumulative IC violations

$$V^T = \sum_{t\in[T]} \left[\max_{\varphi\in\Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right) \right].$$

Since event $\mathcal{E}_\delta^{\text{F}}$ holds, we have that $\forall t \in [T]$, $||\boldsymbol{r}_{\text{R}} - \hat{\boldsymbol{r}}_{\text{R}}^t||_\infty \leq \varepsilon_\delta^t$. Then $\forall t \in [T]$, by Lemma D.1, we have that

$$\left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{S}} \geq \left(\boldsymbol{\xi}^t\right)^\top \hat{\boldsymbol{r}}_{\text{S}}^t - |\mathcal{C}|\varepsilon_\delta^t.$$

Furthermore, by Lemma D.2, $\forall \varphi \in \Phi$

$$\left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} \leq \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t + |\mathcal{C}|\varepsilon_\delta^t.$$

By plugging into the definition of V^T, we get

$$\begin{aligned}
V^T &= \sum_{t\in[T]} \left[\max_{\varphi\in\Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right) \right] \\
&\leq \sum_{t\in[T]} \left[\max_{\varphi\in\Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t - \hat{\boldsymbol{r}}_{\text{S}}^t\right) + 2|\mathcal{C}|\varepsilon_\delta^t \right].
\end{aligned}$$

Notice that, since $\forall t \in [T]$ $\boldsymbol{\xi}^t$ is chosen as an optimal solution to LP$(2|\mathcal{C}|\varepsilon_\delta^t, \hat{\boldsymbol{r}}_{\text{R}}^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$,

$$\max_{\varphi \in \Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t - \hat{\boldsymbol{r}}_{\text{S}}^t\right) \leq 2|\mathcal{C}|\varepsilon_\delta^t.$$

Thus,

$$\begin{aligned} V^T &\leq \sum_{t \in [T]} \left[\max_{\varphi \in \Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t - \hat{\boldsymbol{r}}_{\text{S}}^t\right) + 2|\mathcal{C}|\varepsilon_\delta^t\right] \\ &\leq \sum_{t \in [T]} 4|\mathcal{C}|\varepsilon_\delta^t \\ &\leq |\mathcal{C}|\sqrt{16T\log(4T|\mathcal{C}||\mathcal{A}|/\delta)}, \end{aligned}$$

where the last inequality follows from $\sum_{t \in [T]} \frac{1}{\sqrt{t}} \leq 2\sqrt{T}$.

This concludes the proof. □

E. Proofs Omitted from Section 6

Lemma 6.1. *For all $t \in [T]$, for all $i \in \{\text{R}, \text{S}\}$, $\hat{\boldsymbol{r}}_j^t$ is an unbiased estimator of $\boldsymbol{r}_i$.*

Proof. Fix a round $t \in [T]$ and $j \in \{\text{R}, \text{S}\}$. Let us first notice that $\forall \tau \in [t]$ and for any $c \in \mathcal{C}$ that $\pi^t[c, a^\tau] = 1$

$$\begin{aligned} \mathbb{E}\left[u_j(a^\tau, \theta^\tau)\mathbb{1}\left[s^\tau \in c\right]\right] &= \sum_{\boldsymbol{s} \in c} \sum_{\theta \in \Theta} \mathbb{P}\left(\theta^\tau = \theta, \boldsymbol{s}^\tau = \boldsymbol{s}\right) u_j(a, \theta) \\ &= \sum_{\boldsymbol{s} \in c} \sum_{\theta \in \Theta} p(\theta)\psi_{\text{S}}(\boldsymbol{s}|\theta) u_j(a, \theta) \\ &= r_j[c, a]. \end{aligned}$$

Furthermore, recall that $N^t[c, a]$ is defined such that $N^t[c, a] = \sum_{\tau \in [t]} \pi^\tau[c, a]$. Thus, for any class $c \in \mathcal{C}$ and action $a \in \mathcal{A}$,

$$\begin{aligned} \mathbb{E}\left[\hat{r}^t[c, a]\right] &= \mathbb{E}\left[\frac{1}{N^t[c, a]} \sum_{\substack{\tau \in [t] \\ \pi^\tau[c,a]=1}} u_j(a^\tau, \theta^\tau)\mathbb{1}\left[s^\tau \in c\right]\right] \\ &= r_j[c, a]. \end{aligned}$$

This concludes the proof. □

Lemma 6.2. *For any $\delta \in (0, 1)$, the following holds:*

$$\mathbb{P}\left(\mathcal{E}_\delta^{\text{B}}\right) \geq 1 - \frac{\delta}{2}.$$

Proof. For $\delta \in (0, 1)$, given that for any $j \in \{\text{R}, \text{S}\}$ and $t \in [T]$, $\hat{\boldsymbol{r}}_j$ is an unbiased estimator of $\boldsymbol{r}_j^{\psi_{\text{S}}}$ (Lemma 6.1) we can use Hoeffding's inequality to state the following:

$$\mathbb{P}\left(|\hat{r}_j^t[c, a] - r_j[c, a]| \geq \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2N^t[c, a]}}\right) \leq \frac{\delta}{4T|\mathcal{C}||\mathcal{A}|} \quad \forall t \in [T], \forall c \in \mathcal{C}, \forall a \in \mathcal{A}, \forall j \in \{\text{R}, \text{S}\}.$$

Then, we can apply a union bound on $[T]$, $\mathcal{C}$, $\mathcal{A}$ and $\{\text{R}, \text{S}\}$ and obtain

$$\mathbb{P}\left(|\hat{r}_j^t[c, a] - r_j[c, a]| \leq \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2N^t[c, a]}} \quad \forall t \in [T], \forall c \in \mathcal{C}, \forall a \in \mathcal{A}, \forall j \in \{\text{R}, \text{S}\}\right) \geq 1 - \frac{\delta}{2},$$

which gives the result. □

Lemma E.1. *Let $\boldsymbol{\xi}^\star$ be an optimal solution to LP$(0, \boldsymbol{r}_{\text{R}}, \boldsymbol{r}_{\text{S}})$. Furthermore, define $\nu := 2|\mathcal{C}|\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}}$. Then, if event $\mathcal{E}_\delta^{\text{B}}$ holds, $\boldsymbol{\xi}^\star$ is a feasible mechanism for LP$(\nu, \hat{\boldsymbol{r}}_{\text{R}}^t + \boldsymbol{\eta}_\delta^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$ for all $t \in [|\mathcal{A}|E + 1, T]$.*

Proof. Fix a round $t \in [|\mathcal{A}|E + 1, T]$. First, let us notice that the initial exploration phase guarantees that all the counters $N^t[c, a]$ are lower bounded. Formally:

$$N^t[c, a] \geq E \quad \forall c \in \mathcal{C}, \forall a \in \mathcal{A}.$$

Hence, we have that

$$||\boldsymbol{\eta}_\delta^t||_\infty = \max_{\substack{c \in \mathcal{C} \\ a \in \mathcal{A}}} \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2N^t[c, a]}} \leq \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}}. \tag{15}$$

Furthermore, when event $\mathcal{E}_\delta^{\text{B}}$ is verified, we have that

$$||\hat{\boldsymbol{r}}_{\text{S}}^t - \boldsymbol{r}_{\text{S}}||_\infty \leq ||\boldsymbol{\eta}_\delta^t||_\infty \leq \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}}. \tag{16}$$

Then, by Lemma D.1, it holds that for each $\boldsymbol{\xi} \in \Xi$

$$\left|\boldsymbol{\xi}^\top \left(\hat{\boldsymbol{r}}_{\text{S}}^t - \boldsymbol{r}_{\text{S}}\right)\right| \leq |\mathcal{C}|\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}} \tag{17}$$

Additionally, by Lemma D.2, for all $\varphi \in \Phi$, it holds that

$$|(\boldsymbol{\xi}^\star)^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - (\boldsymbol{\xi}^\star)^\top \boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t| \leq |\mathcal{C}|\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}}.$$

Thus, we get that for each $\varphi \in \Phi$

$$(\boldsymbol{\xi}^\star)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t - \hat{\boldsymbol{r}}_{\text{S}}^t\right) \leq (\boldsymbol{\xi}^\star)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right) + 2|\mathcal{C}|\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}} \leq 2|\mathcal{C}|\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}},$$

where the last equation follows from the fact that $\boldsymbol{\xi}^\star$ is a feasible solution for LP$(0, \boldsymbol{r}_{\text{R}}, \boldsymbol{r}_{\text{S}})$. This concludes the proof. □

Lemma E.2. *For any $E \in [T/\mathcal{A}]$, the following holds:*

$$\sum_{t=1}^{|\mathcal{A}|E} \max_{\varphi \in \Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right) = 0$$

Proof. Note that, for each $\boldsymbol{\xi}^t$ in the exploration phase, by definition there exists an action $a' \in \mathcal{A}$ such that $\xi[c, a'] = 1$, for any $c \in \mathcal{C}$. Then,

$$\left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{S}} = \sum_{c \in \mathcal{C}} \sum_{\boldsymbol{s} \in c} \sum_{\theta \in \Theta} p(\theta)\psi_{\text{S}}(\boldsymbol{s}|\theta)u_{\text{S}}(a', \theta) = \sum_{\theta \in \Theta} p(\theta)u_{\text{S}}(a', \theta).$$

Furthermore, similarly to the proof of Lemma D.2, for each $\varphi \in \Phi$, we can define a mechanism $\boldsymbol{\xi}'$ such that $\left(\boldsymbol{\xi}'\right)^\top \boldsymbol{r}_{\text{S}} = \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}}$, where

$$\xi'[c, a] = \frac{1}{|c|} \sum_{\boldsymbol{s} \in c} \xi^t[c(\varphi(s_i), \boldsymbol{s}_{-i}), a] \quad \forall c \in \mathcal{C}\, \forall a \in \mathcal{A}.$$

Therefore, by simple calculations, we have that $\forall c \in \mathcal{C}$, $\xi'[c, a'] = 1$ and $\xi'[c, a] = 0$, $\forall a \neq a'$, which gives

$$\left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} = \sum_{c \in \mathcal{C}} \sum_{\boldsymbol{s} \in c} \sum_{\theta \in \Theta} p(\theta)\psi_{\text{S}}(\boldsymbol{s}|\theta)u_{\text{S}}(a', \theta) = \sum_{\theta \in \Theta} p(\theta)u_{\text{S}}(a', \theta).$$

Putting all together, we have

$$\sum_{t=1}^{|\mathcal{A}|E} \max_{\varphi \in \Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right) = \sum_{\theta \in \Theta} p(\theta)u_{\text{S}}(a', \theta) - p(\theta)u_{\text{S}}(a', \theta) = 0,$$

which gives the result. □

Lemma E.3. *For any $\delta \in (0,1)$ and $E \in [T]$ with probability at least $1 - \frac{\delta}{2}$, the following holds:*

$$\sum_{t=|\mathcal{A}|E+1}^{T} \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{\eta}_\delta^t \leq |\mathcal{C}|\mathcal{A}|\sqrt{2T\log(8T|\mathcal{C}||\mathcal{A}|/\delta)} + |\mathcal{C}||\mathcal{A}|\sqrt{2T\log(2/\delta)}$$

Proof. As a first step towards proving this Lemma, we leverage the Azuma-Hoeffding inequality (Cesa-Bianchi & Lugosi, 2006) to show that

$$\sum_{t=|\mathcal{A}|E+1}^{T} \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{\eta}_\delta^t \leq \sum_{t=|\mathcal{A}|E+1}^{T} \left[\left(\boldsymbol{\pi}^t\right)^\top \boldsymbol{\eta}_\delta^t\right] + |\mathcal{C}||\mathcal{A}|\sqrt{2T\log(2/\delta)}$$

In particular, for all $t \in [|\mathcal{A}|E+1, T]$, we can define the random variable $X^t \coloneqq \sum_{\tau=|\mathcal{A}|E+1}^{t} \left[\left(\boldsymbol{\pi}^\tau\right)^\top \boldsymbol{\eta}_\delta^\tau - \left(\boldsymbol{\xi}^\tau\right)^\top \boldsymbol{\eta}_\delta^\tau\right]$. Note that, since deterministic mechanisms $\boldsymbol{\pi}^t$ are sampled according from $\boldsymbol{\xi}^t$, we have that, for every $t \in [|\mathcal{A}|E+1, T]$,

$$\mathbb{E}\left[\boldsymbol{\pi}^t | \mathcal{F}^{t-1}\right] = \boldsymbol{\xi}^t.$$

where $\mathcal{F}^{t-1}$ is the filtration generated up to time $t-1$ during the online interaction between the receiver's algorithm and the senders. Thus, for each $t \in [|\mathcal{A}|E+1, T]$, X^t is a martingale. Hence, for $\delta \in (0,1)$, from Azuma-Hoeffding inequality we can conclude that, with probability at least $1 - \frac{\delta}{2}$, it holds:

$$\sum_{t=|\mathcal{A}|E+1}^{T} \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{\eta}_\delta^t \leq \sum_{t=|\mathcal{A}|E+1}^{T} \left[\left(\boldsymbol{\pi}^t\right)^\top \boldsymbol{\eta}_\delta^t\right] + |\mathcal{C}||\mathcal{A}|\sqrt{2T\log(2/\delta)},$$

which gives the desired result.

Then, we conclude the proof by deriving an upper bound to $\sum_{t=|\mathcal{A}|E+1}^{T} \left(\boldsymbol{\pi}^t\right)^\top \boldsymbol{\eta}_\delta^t$. We can state the following chain of inequalities:

$$\begin{aligned}
\sum_{t=|\mathcal{A}|E+1}^{T} \left(\boldsymbol{\pi}^t\right)^\top \boldsymbol{\eta}_\delta^t &= \sum_{t=|\mathcal{A}|E+1}^{T} \sum_{\substack{c\in\mathcal{C}\\ a\in\mathcal{A}}} \pi^t[c,a]\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2N^t[c,a]}} \\
&= \sum_{\substack{c\in\mathcal{C}\\ a\in\mathcal{A}}} \sum_{t=|\mathcal{A}|E+1}^{T} \pi^t[c,a]\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2N^t[c,a]}} \\
&= \sum_{\substack{c\in\mathcal{C}\\ a\in\mathcal{A}}} \sum_{\substack{t\in[|\mathcal{A}|E+1,T]:\\ \pi^t[c,a]=1}} \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2N^t[c,a]}} \\
&= \sum_{\substack{c\in\mathcal{C}\\ a\in\mathcal{A}}} \sum_{t=N^{|\mathcal{A}|E+1}[c,a]}^{N^T[c,a]} \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2t}} \qquad (18) \\
&\leq \sum_{\substack{c\in\mathcal{C}\\ a\in\mathcal{A}}} \sqrt{2\log(8T|\mathcal{C}||\mathcal{A}|/\delta)N^T[c,a]} \qquad (19) \\
&\leq |\mathcal{C}|\mathcal{A}|\sqrt{2T\log(8T|\mathcal{C}||\mathcal{A}|/\delta)},
\end{aligned}$$

where Equation (18) follows from the definition of $N^t[c,a]$ and Equation (19) follows from the fact that $\sum_{t=1}^{T} \frac{1}{\sqrt{t}} \leq 2\sqrt{T}$. Thus, for any $\delta \in (0,1)$, we can conclude that, with probability at least $1 - \frac{\delta}{2}$,

$$\begin{aligned}
\sum_{t=|\mathcal{A}|E+1}^{T} \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{\eta}_\delta^t &\leq \sum_{t=|\mathcal{A}|E+1}^{T} \left[\left(\boldsymbol{\pi}^t\right)^\top \boldsymbol{\eta}_\delta^t\right] + |\mathcal{C}||\mathcal{A}|\sqrt{2T\log(2/\delta)} \\
&\leq |\mathcal{C}|\mathcal{A}|\sqrt{2T\log(8T|\mathcal{C}||\mathcal{A}|/\delta)} + |\mathcal{C}||\mathcal{A}|\sqrt{2T\log(2/\delta)}.
\end{aligned}$$

This concludes the proof. □

Theorem 6.1. *For any $\delta \in (0,1)$ and for any $E \in [T]$, with probability at least $1-\delta$, Algorithm 2 guarantees*

$$R^T = \mathcal{O}\left(|\mathcal{A}|E + |\mathcal{C}||\mathcal{A}|\sqrt{32T\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}\right),$$
$$V^T = \mathcal{O}\left(T|\mathcal{C}|\sqrt{8\log(8T|\mathcal{C}||\mathcal{A}|/\delta)/E}\right).$$

Proof. To prove this Theorem, we first derive the upper bound for the cumulative regret R^T and then show the upper bound for cumulative IC violation V^T. For $\delta \in (0,1)$, let $\overline{\mathcal{E}}_\delta$ be the event defined such that

$$\overline{\mathcal{E}}_\delta := \left\{ \sum_{t=|\mathcal{A}|E+1}^{T} \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{\eta}_\delta^t \leq |\mathcal{C}|\mathcal{A}|\sqrt{2T\log(8T|\mathcal{C}||\mathcal{A}|/\delta)} + |\mathcal{C}||\mathcal{A}|\sqrt{2T\log(2/\delta)} \right\}.$$

By Lemma E.3 we have that event $\overline{\mathcal{E}}_\delta$ is verified with probability at least $1-\frac{\delta}{2}$. Furthermore, from Lemma 6.2, we know that event $\mathcal{E}_\delta^{\text{B}}$ holds with probability greater than $1-\frac{\delta}{2}$. Throughout this proof, we assume that both events $\overline{\mathcal{E}}_\delta$ and $\mathcal{E}_\delta^{\text{B}}$ hold. Hence, by a simple union bound, it is possible to show that

$$\mathbb{P}\left(\overline{\mathcal{E}}_\delta, \mathcal{E}_\delta^{\text{B}}\right) \geq 1-\delta.$$

Regret. Notice that, when event $\mathcal{E}_\delta^{\text{B}}$ is verified, we can write the following:

$$\boldsymbol{r}_{\text{R}} \preceq \hat{\boldsymbol{r}}_{\text{R}}^t + \boldsymbol{\eta}_\delta^t \preceq \boldsymbol{r}_{\text{R}} + 2\boldsymbol{\eta}_\delta^t \quad \forall t \in [T],$$

where $\preceq$ denotes the elemnt-wise $\leq$ relationship.

Hence, given that $\Xi \subset \mathbb{R}_{\geq 0}^{|\mathcal{C}||\mathcal{A}|}$, the following inequalities follow:

$$\begin{aligned}
R^T &= \sum_{t\in[T]} \left[\left(\boldsymbol{\xi}^\star\right)^\top \boldsymbol{r}_{\text{R}} - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}}\right] \\
&= \sum_{t=1}^{|\mathcal{A}|E} \left[\left(\boldsymbol{\xi}^\star\right)^\top \boldsymbol{r}_{\text{R}} - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}}\right] + \sum_{t=|\mathcal{A}|E+1}^{T} \left[\left(\boldsymbol{\xi}^\star\right)^\top \boldsymbol{r}_{\text{R}} - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}}\right] \\
&\leq |\mathcal{A}|E + \sum_{t=|\mathcal{A}|E+1}^{T} \left[\left(\boldsymbol{\xi}^\star\right)^\top \boldsymbol{r}_{\text{R}} - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}}\right] \\
&\leq |\mathcal{A}|E + \sum_{t=|\mathcal{A}|E+1}^{T} \left[\left(\boldsymbol{\xi}^\star\right)^\top \left(\hat{\boldsymbol{r}}_{\text{R}}^t + \boldsymbol{\eta}_\delta^t\right) - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}}\right] && (20) \\
&\leq |\mathcal{A}|E + \sum_{t=|\mathcal{A}|E+1}^{T} \left[\left(\boldsymbol{\xi}^t\right)^\top \left(\hat{\boldsymbol{r}}_{\text{R}}^t + \boldsymbol{\eta}_\delta^t\right) - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}}\right] && (21) \\
&\leq |\mathcal{A}|E + \sum_{t=|\mathcal{A}|E+1}^{T} \left[\left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{r}_{\text{R}} + 2\boldsymbol{\eta}_\delta^t\right) - \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{r}_{\text{R}}\right] \\
&= |\mathcal{A}|E + 2\sum_{t=|\mathcal{A}|E+1}^{T} \left(\boldsymbol{\xi}^t\right)^\top \boldsymbol{\eta}_\delta^t \\
&\leq |\mathcal{A}|E + 2|\mathcal{C}|\mathcal{A}|\left(\sqrt{2T\log(8T|\mathcal{C}||\mathcal{A}|/\delta)} + \sqrt{2T\log(2/\delta)}\right) && (22) \\
&\leq |\mathcal{A}|E + |\mathcal{C}|\mathcal{A}|\left(\sqrt{32T\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}\right),
\end{aligned}$$

where Equation (20) follows from the fact that event $\mathcal{E}_\delta^{\text{B}}$ holds, Equation (21) is a consequence of the fact that $\boldsymbol{\xi}^\star$ and $\boldsymbol{\xi}^t$ are, respectively a feasible (by Lemma E.1) and an optimal (by definition) mechanism for $\text{LP}(\nu, \hat{\boldsymbol{r}}_{\text{R}}^t + \boldsymbol{\eta}_\delta^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$, and Equation (22) follows from the assumption that event $\overline{\mathcal{E}}_\delta$ holds.

This concludes the first part of the proof.

IC violation. Let us notice that the initial exploration phase guarantees that at each round $t \in [|\mathcal{A}|E+1, T]$, all the counters $N^t[c,a]$ are lower bounded. Formally:

$$N^t[c,a] \geq E \quad \forall t \in [|\mathcal{A}|E+1, T], \forall c \in \mathcal{C}, \forall a \in \mathcal{A}.$$

Thus, $\boldsymbol{\eta}_\delta^t \preceq \boldsymbol{\eta}_\delta^{|\mathcal{A}|E}$, $\forall t \in [|\mathcal{A}|E+1, T]$. Additionally, since event $\mathcal{E}_\delta^{\text{B}}$ holds, we have that

$$\hat{\boldsymbol{r}}_{\text{S}}^t - \boldsymbol{\eta}_\delta^{|\mathcal{A}|E} \preceq \hat{\boldsymbol{r}}_{\text{S}}^t - \boldsymbol{\eta}_\delta^t \preceq \boldsymbol{r}_{\text{S}} \preceq \hat{\boldsymbol{r}}_{\text{S}}^t + \boldsymbol{\eta}_\delta^t \preceq \hat{\boldsymbol{r}}_{\text{S}}^t + \boldsymbol{\eta}_\delta^{|\mathcal{A}|E},$$

where $\preceq$ denotes the element-wise $\leq$ relationship. Furthermore, notice that the design of the exploration phase guarantees that for each $c \in \mathcal{C}$ and $a \in \mathcal{A}$,

$$\eta_\delta^{|\mathcal{A}|E}[c,a] = \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}}.$$

In the following, with a slight abuse of notation, we denote $\eta_\delta^{|\mathcal{A}|E} = \sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}}$. Therefore, by Lemma D.1 it holds that for all $t \in [|\mathcal{A}|E+1, T]$

$$|(\boldsymbol{\xi}^t)^\top \boldsymbol{r}_{\text{S}} - (\boldsymbol{\xi}^t)^\top \hat{\boldsymbol{r}}_{\text{S}}^t| \leq |\mathcal{C}|\eta_\delta^{|\mathcal{A}|E},$$

and by Lemma D.2, for all $t \in [|\mathcal{A}|E+1, T]$ and $\varphi \in \Phi$, it holds that

$$|(\boldsymbol{\xi}^t)^\top A_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - (\boldsymbol{\xi}^t)^\top A_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t| \leq |\mathcal{C}|\eta_\delta^{|\mathcal{A}|E}.$$

This allows us to obtain the following chain of inequalities:

$$\begin{aligned}
V^T &= \sum_{t\in[T]} \left[\max_{\varphi\in\Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right)\right] \\
&= \sum_{t=1}^{|\mathcal{A}|E} \left[\max_{\varphi\in\Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right)\right] + \sum_{t=|\mathcal{A}|E+1}^{T} \left[\max_{\varphi\in\Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right)\right] \\
&= \sum_{t=|\mathcal{A}|E+1}^{T} \left[\max_{\varphi\in\Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \boldsymbol{r}_{\text{S}} - \boldsymbol{r}_{\text{S}}\right)\right] && (23) \\
&\leq \sum_{t=|\mathcal{A}|E+1}^{T} \left[\max_{\varphi\in\Phi} \left(\boldsymbol{\xi}^t\right)^\top \left(\boldsymbol{A}_{\text{S}}^\varphi \hat{\boldsymbol{r}}_{\text{S}}^t - \hat{\boldsymbol{r}}_{\text{S}}^t\right) + 2|\mathcal{C}|\eta_\delta^{|\mathcal{A}|E}\right] \\
&\leq \sum_{t=|\mathcal{A}|E+1}^{T} 4|\mathcal{C}|\eta_\delta^{|\mathcal{A}|E} && (24) \\
&= 4T|\mathcal{C}|\sqrt{\frac{\log(8T|\mathcal{C}||\mathcal{A}|/\delta)}{2E}},
\end{aligned}$$

where Equation (23) follows from Lemma E.2, Equation (24) follows from the fact that $\boldsymbol{\xi}^t$ is chosen as an optimal solution to $\text{LP}(\nu, \hat{\boldsymbol{r}}_{\text{R}}^t + \boldsymbol{\eta}_\delta^t, \hat{\boldsymbol{r}}_{\text{S}}^t)$ with $\nu = 2|\mathcal{C}|\eta_\delta^{|\mathcal{A}|E}$. This concludes the proof. □

Theorem 6.2. *For any $\alpha \in [1/2, 1]$, there exists $\delta \in (0,1)$ such that no algorithm can achieve both $R^T = o(T^\alpha)$ and $V^T = o(T^{1-\alpha/2})$ with probability greater than $1-\delta$.*

Proof. Let us introduce two instances of a game between a receiver R and a single sender S. The set of states of nature is composed by three distinct states $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and the set of signals is $S = \{s_1, s_2, s_3\}$. Both instances share the same signaling scheme ψ such that, $\psi(s_1|\theta) = \psi(s_2|\theta) = 1/2$ and $\psi(s_3|\theta) = 0$, for $\theta \in \{\theta_1, \theta_2\}$, while $\psi(s_3|\theta_3) = 1$ and $\psi(s_1|\theta_3) = \psi(s_2|\theta_3) = 0$. The set of actions available to the receiver is composed of two actions, namely $\mathcal{A} = \{a, b\}$. The utility function is specified such that

$$u_{\text{R}}(a,\theta) = \begin{cases} 1 & \text{if } \theta \in \{\theta_1, \theta_2\} \\ 0 & \text{if } \theta = \theta_3, \end{cases} \qquad u_{\text{R}}(b,\theta) = \begin{cases} 0 & \text{if } \theta \in \{\theta_1, \theta_2\} \\ 1 & \text{if } \theta = \theta_3, \end{cases}$$

for the receiver, while

$$u_{\text{S}}(a,\theta) = \begin{cases} 1/2 & \text{if } \theta \in \{\theta_1, \theta_2\} \\ 0 & \text{if } \theta = \theta_3, \end{cases} \qquad u_{\text{S}}(b,\theta) = \begin{cases} 0 & \text{if } \theta \in \{\theta_1, \theta_2\} \\ 1 & \text{if } \theta = \theta_3, \end{cases}$$

for the sender.

The only difference between the two instances lies in the prior. In particular, given $\varepsilon > 0$,

$$\text{X} = \begin{cases} p(\theta_1) = \frac{1}{3} - \varepsilon \\ p(\theta_2) = \frac{1}{3} + \varepsilon \\ p(\theta_3) = \frac{1}{3}, \end{cases} \qquad \text{Y} = \begin{cases} p(\theta_1) = \frac{1}{3} + \varepsilon \\ p(\theta_2) = \frac{1}{3} - \varepsilon \\ p(\theta_3) = \frac{1}{3}, \end{cases}$$

It is easy to verify that in instance X the optimal IC mechanism $\boldsymbol{\xi}^\star$ for R prescribes to select action a when the signal reported are s_1 and s_2, while it selects action b when the signal reported is s_3. Thus, straightforwardly, we can express the cumulative regret R_{X}^T in instance X as

$$R_{\text{X}}^T = \frac{1}{3} \sum_{t \in [T]} \left[\xi^t[s_1, b] + \xi^t[s_2, b] + \xi^t[s_3, a] \right]. \tag{25}$$

Let $\mathbb{P}_i$ be the probability measure associated to instance $i \in \{\text{X}, \text{Y}\}$. For $\delta \in (0,1)$, let us assume that, under probability measure $\mathbb{P}_{\text{X}}$, there exists an algorithm that guarantees $R_{\text{X}}^T \leq \frac{K}{3}$ with probability at least $1 - \delta$, *i.e.,* such that

$$\mathbb{P}_{\text{X}} \left(\sum_{t \in [T]} \left[\xi^t[s_1, b] + \xi^t[s_2, b] + \xi^t[s_3, a] \right] \leq K \right) \geq 1 - \delta.$$

Then, from the Pinsker's inequality we know that

$$\begin{aligned} &\mathbb{P}_{\text{Y}} \left(\sum_{t \in [T]} \left[\xi^t[s_1, b] + \xi^t[s_2, b] + \xi^t[s_3, a] \right] \leq K \right) \\ &\geq \mathbb{P}_{\text{X}} \left(\sum_{t \in [T]} \left[\xi^t[s_1, b] + \xi^t[s_2, b] + \xi^t[s_3, a] \right] \leq K \right) - \sqrt{\frac{1}{2} \mathcal{K}(\text{X}, \text{Y})} \\ &\geq 1 - \delta - \sqrt{\frac{1}{2} \mathcal{K}(\text{X}, \text{Y})}, \end{aligned} \tag{26}$$

where $\mathcal{K}(\text{Y}, \text{X})$ is the Kullback-Leibler divergence between instance Y and instance X.

Let us notice that the two instances are identical, except for what concerns the sender's reward distribution obtained when the signal state sampled is s_1 or s_2 and the action played is b. More in detail, we have that in instance X such a reward distribution is a Bernoulli random variable with parameter $\frac{1}{2} + \frac{3}{2}\varepsilon$, while in instance Y it is a Bernoulli with parameter $\frac{1}{2} - \frac{3}{2}\varepsilon$. Hence, by applying the divergence decomposition theorem (Lattimore & Szepesvári, 2020) we obtain the following:

$$\begin{aligned} \mathcal{K}(\text{X}, \text{Y}) &= \left(\sum_{\substack{t \in [T], \\ s^t = s_1}} \left[\mathbb{E}_{\text{X}} \left[\xi^t[s_1, b] \right] \right] + \sum_{\substack{t \in [T], \\ s^t = s_2}} \left[\mathbb{E}_{\text{X}} \left[\xi^t[s_2, b] \right] \right] \right) \mathcal{K}\left(\mathcal{B}\left(\frac{1}{2} + \frac{3}{2}\varepsilon \right), \mathcal{B}\left(\frac{1}{2} - \frac{3}{2}\varepsilon \right) \right) \\ &\leq \sum_{t \in [T]} \mathbb{E}_{\text{X}} \left[\xi^t[s_1, b] + \xi^t[s_2, b] + \xi^t[s_3, a] \right] \mathcal{K}\left(\mathcal{B}\left(\frac{1}{2} + \frac{3}{2}\varepsilon \right), \mathcal{B}\left(\frac{1}{2} - \frac{3}{2}\varepsilon \right) \right) \\ &\leq \mathbb{E}_{\text{X}} \left[\sum_{t \in [T]} \left[\xi^t[s_1, b] + \xi^t[s_2, b] + \xi^t[s_3, a] \right] \right] \mathcal{K}\left(\mathcal{B}\left(\frac{1}{2} + \frac{3}{2}\varepsilon \right), \mathcal{B}\left(\frac{1}{2} - \frac{3}{2}\varepsilon \right) \right) \\ &\leq \mathbb{E}_{\text{X}} \left[\sum_{t \in [T]} \left[\xi^t[s_1, b] + \xi^t[s_2, b] + \xi^t[s_3, a] \right] \right] 36\varepsilon^2. \end{aligned} \tag{27}$$

Additionally, from reverse Markov inequality it follows that

$$\mathbb{E}_{\text{X}}\left[\sum_{t\in[T]}\left[\xi^t[s_1,b]+\xi^t[s_2,b]+\xi^t[s_3,a]\right]\right] \leq (T-K)\,\mathbb{P}_{\text{X}}\left(\sum_{t\in[T]}\left[\xi^t[s_1,b]+\xi^t[s_2,b]+\xi^t[s_3,a]\right]\geq K\right)+K$$
$$\leq (T-K)\,\delta+K. \tag{28}$$

Thus, plugging inequalities (27) and (28) into Equation (26), we obtain

$$\mathbb{P}_{\text{Y}}\left(\sum_{t\in[T]}\left[\xi^t[s_1,b]+\xi^t[s_2,b]+\xi^t[s_3,a]\right]\leq K\right)\geq 1-\delta-3\varepsilon\sqrt{2\,(T-K)\,\delta+2K}. \tag{29}$$

Now let us focus on the formulation of the definition of the cumulative IC deviation V_{Y}^T for instance Y. Let $\varphi\in\Phi$ be such that $\varphi(s_i)=s_3$, for any $s_i\in S$. Trivially, by considering the IC deviation computed with respect to φ at every round $t\in[T]$, we can provide a lower bound to V_{Y}^T. Formally,

$$V_{\text{Y}}^T\geq\sum_{t=1}^{T}U_{\text{Y,S}}^{\varphi}(\boldsymbol{\xi}^t)-U_{\text{Y,S}}(\boldsymbol{\xi}^t).$$

By simple calculations it is possible to show that

$$U_{\text{Y,S}}(\boldsymbol{\xi}^t)=\frac{2}{3}+\frac{\varepsilon}{2}\xi^t[s_1,b]+\frac{\varepsilon}{2}\xi^t[s_2,b]-\frac{1}{3}\xi^t[s_3,a],$$

and

$$U_{\text{Y,S}}^{\varphi}(\boldsymbol{\xi}^t)=\frac{2}{3}+\varepsilon-\varepsilon\xi[s_3,a]-\frac{1}{3}\xi[s_3,a].$$

Thus, we can write the following:

$$\begin{aligned}
V_{\text{Y}}^T &\geq\sum_{t=1}^{T}\left[U_{\text{Y,S}}^{\varphi}(\boldsymbol{\xi}^t)-U_{\text{Y,S}}(\boldsymbol{\xi}^t)\right]\\
&=\varepsilon\sum_{t=1}^{T}\left[1-\frac{1}{2}\xi^t[s_1,b]-\frac{1}{2}\xi^t[s_2,b]-\xi^t[s_3,a]\right]\\
&\geq\varepsilon\sum_{t=1}^{T}\left[1-\xi^t[s_1,b]-\xi^t[s_2,b]-\xi^t[s_3,a]\right]\\
&=\varepsilon T-\varepsilon\sum_{t=1}^{T}\left[\xi^t[s_1,b]+\xi^t[s_2,b]+\xi^t[s_3,a]\right],
\end{aligned}$$

which, together with Equation (29), gives us that, with probability at least $1-\delta-3\varepsilon\sqrt{2\,(T-K)\,\delta+2K}$,

$$V_{\text{Y}}^T\geq\varepsilon T-\varepsilon K.$$

To recover the desired result, set $\varepsilon=\frac{T^{-\alpha/2}}{100}$ and $K=\frac{T^{\alpha}}{9}$. Then, assuming without loss of generality $\delta\leq\frac{T^{\alpha-1}}{9}$, we get

$$\mathbb{P}_{\text{Y}}\left(V_{\text{Y}}^T\geq\frac{T^{1-\alpha/2}}{100}\right)\geq 0.98-\delta.$$

This concludes the proof. □