Multi-Sender Persuasion: A Computational Perspective

Safwan Hossain [*1] **Tonghan Wang** [*1] **Tao Lin** [*1] **Yiling Chen** [1] **David C. Parkes** [1] **Haifeng Xu** [2]

Abstract

We consider *multiple senders* with informational advantage signaling to convince a single self-interested actor to take certain actions. Generalizing the seminal *Bayesian Persuasion* framework, such settings are ubiquitous in computational economics, multi-agent learning, and machine learning with multiple objectives. The core solution concept here is the Nash equilibrium of senders' signaling policies. Theoretically, we prove that finding an equilibrium in general is PPAD-Hard; in fact, even computing a sender's best response is NP-Hard. Given these intrinsic difficulties, we turn to finding local Nash equilibria. We propose a novel differentiable neural network to approximate this game's non-linear and discontinuous utilities. Complementing this with the extra-gradient algorithm, we discover local equilibria that Pareto dominates full-revelation equilibria and those found by existing neural networks. Broadly, our theoretical and empirical contributions are of interest to a large class of economic problems.

1. Introduction

Bayesian Persuasion (BP) (Kamenica & Gentzkow, 2011) has emerged as a seminal concept in economics and decision theory. At its heart, it is a principal-agent problem that models an informed sender strategically revealing some information to affect the decisions of a self-interested receiver. Both parties are assumed to be Bayesian and have distinct utilities that depend on some realized *state of nature*, and the action taken by the receiver. The sender privately observes the state and can commit to selectively disclosing this information through a randomized *signaling policy*. The receiver updates their posterior belief based on the realized signal and best responds with an optimal action for this belief. The sender's goal is to maximize their utility by designing a signaling policy that nudges the receiver toward decisions preferred by the sender. This information design problem has found widespread applicability in a myriad of domains including recommendation systems (Mansour et al., 2015; 2016), auctions and advertising (Bro Miltersen & Sheffet, 2012; Emek et al., 2014; Badanidiyuru et al., 2018), social networks (Candogan & Drakopoulos, 2020; Acemoglu et al., 2021), and reinforcement learning (Castiglioni et al., 2020; Wu et al., 2022).

The standard BP model is however significantly constrained by a strong assumption: the presence of only one sender. In the applications mentioned above and indeed more broadly in settings like multi-agent learning (Balduzzi et al., 2018) and machine learning with multiple objectives (Pfau & Vinyals, 2016; Jaderberg et al., 2017), it is natural to have multiple parties who wish to influence the receiver toward their respective, possibly conflicting goals. As a demonstrative example, consider two ride-sharing firms, Uber and Lyft, and a dual-registered driver. While the driver is unaware of real-time demand patterns, both firms have access to and can strategically signal this to the driver and influence them toward certain pick-ups. The platforms' goals however are not aligned, with each wishing to direct the driver to their respective optimal pick-ups. The driver is also self-interested and may prefer pick-ups that are on the way home. Our work aims to study this tension induced by multiple informed parties attempting to influence a self-interested receiver's decision-making, within the BP paradigm. Crucially, while the sender-receiver relation still outlines a sequential game, the interaction *among the multiple senders* in our setting forms a simultaneous game, with the resulting Nash Equilibrium (NE) being of core interest.

While this setup has been modeled in economic literature Gentzkow & Kamenica (2017b); Ravindran & Cui (2022), the multi-sender persuasion problem has not been formally studied from a computational perspective and presents distinct challenges. In standard single-sender BP, the optimal signaling policy for the sender can be computed efficiently by a linear program (Dughmi & Xu, 2016), which no longer holds in the multi-sender case where we need to compute a sender's best-responding signaling policy given others' policies. We give a non-convex optimization program for the best response problem (Proposition 2) and through an

[*] Equal contribution [1] Harvard University [2] University of Chicago. Correspondence to: Safwan Hossain, Tonghan Wang, Tao Lin <{shossain, twang1, tlin}@g.harvard.edu>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

involved reduction, prove that computing best response is in-fact NP-Hard in multi-sender persuasion games (Theorem 3). For the equilibrium computation, we significantly generalize a specific characterization from prior works to show that a trivial equilibrium can be found easily under certain conditions, but it might offer poor utility to the senders (Theorem 4). We then prove that finding an equilibrium in general settings is PPAD-hard (Theorem 5). These computational hardness results are our main theoretical contribution.

The intrinsic difficulty of finding (global) equilibrium in multi-sender persuasion motivates us to propose a deep-learning approach to finding ϵ-*local* equilibria (no unilateral deviation in a limited range is beneficial). This spiritually straddles two bodies of work - the emergent area of *differentiable economics* that builds a parameterized representation for optimal economic design (Dütting et al., 2023), and the rich literature on learning in games (Bowling, 2004; Balduzzi et al., 2018; Azizian et al., 2020; Fiez et al., 2020; Bai et al., 2021; Bichler et al., 2021; Haghtalab et al., 2022; Goktas et al., 2023). Mirroring the obstacles encountered in theoretical analysis, the non-differentiable and indeed discontinuous nature of the utility functions (Proposition 1) also pose hurdles to identifying even local equilibria. To address this, we propose a novel end-to-end differentiable network architecture that is expressive enough to model the abrupt changes in utilities. Once trained, these networks can complement algorithms like extra-gradient (Korpelevich, 1976; Jelassi et al., 2020) to locate ϵ-local NE. The quality of the approximated utility landscape confirms the superior expressive capacity of our networks. Further, we demonstrate that this improvement helps to discover ϵ-local NE that Pareto dominates the full-revelation equilibria (Theorem 4) and the ϵ-local NE found in both synthetic and real-world scenarios by existing continuous and discontinuous (Wang et al., 2023) networks. Our novel techniques may be of independent interest for learning in general games with discontinuous and non-linear utilities.

1.1. Additional Related Work

The study of Bayesian persuasion and its various iterations has been extensively explored in the literature, as evidenced by the comprehensive surveys of Dughmi (2017); Kamenica (2019); Bergemann & Morris (2019). Among these, the most closely aligned with our work are the investigations involving multiple senders. The model of Gentzkow & Kamenica (2017b) explores a scenario where senders can arbitrarily correlate their signals, whereas Li & Norman (2021) consider sequential senders who choose signaling policies after observing those of previous senders. These two models differ from ours wherein the senders send signals to the receiver *independently* and *simultaneously* conditioning on the realized state. Further, they do not provide significant computational insights. Ding et al. (2023) study multi-sender information design in a special Pandora Box setup which differs significantly from ours and is thus not comparable.

Ravindran & Cui (2022) also study Bayesian persuasion games featuring multiple independent and simultaneous senders but assume that the senders have zero-sum utilities. They show that with a sufficiently large signaling space, the only Nash equilibrium is full revelation, wherein the state of nature is fully revealed to the receiver. However, many multi-sender persuasion games do not conform to a zero-sum utility framework. Consequently, two important structural questions arise from their work: (1) whether such a full-revelation equilibrium exists under a limited signaling space, and (2) whether multi-sender persuasion games with general utility structures give rise to other types of equilibria that extend beyond full revelation. We provides affirmative answers to these queries, as delineated in Theorem 4.

Bayesian persuasion is subsumed within the broader principal-agent model (Gan et al., 2024), a concept that addresses a multitude of economics problems, including contract design (Zhu et al., 2023) and Stackelberg games (Myerson, 1982). In economic theory, the notion of incorporating multiple principals, analogous to senders in our context, has been proposed to model a range of important settings (Waterman & Meier, 1998; Hu et al., 2023). However, similar to the existing work on multi-sender persuasion, these contributions typically retain a conceptual focus from an economic perspective. Our work diverges by taking a computational lens. We introduce rigorous hardness guarantees for the best-response computation and equilibrium determination and propose a novel deep learning approach for identifying ϵ-local equilibria that may hold wider applicability.

2. Model

Preliminaries There are n senders $\{1, \ldots, n\}$ and a receiver. Let Ω be a finite set of possible states, with $\omega \in \Omega$ denoting an arbitrary one. All senders and the receiver share a common prior distribution μ_0 over the states Ω. We use $\mu \in \Delta(\Omega)$ to denote a distribution over states. Receiver takes some action $a \in \mathcal{A}$ whose utility depends on the realized state ω, and is given by $v : \Omega \times \mathcal{A} \to \mathbb{R}$. The receiver's utility can also be represented as an $|\Omega| \times |\mathcal{A}|$ matrix V, with $V[i, j]$ denoting the utility the receiver has for action j at state i. The utility function of the jth sender $u_j : \Omega \times \mathcal{A} \to \mathbb{R}$ also depends on the realized state and the receiver's action. While the receiver only knows the prior, senders privately observe the state realization $\omega \sim \mu_0$ and can use this informational advantage to alter the receiver's belief and persuade it to take certain actions.

Persuasion We model the interaction between senders and the receiver using the seminal Bayesian Persuasion

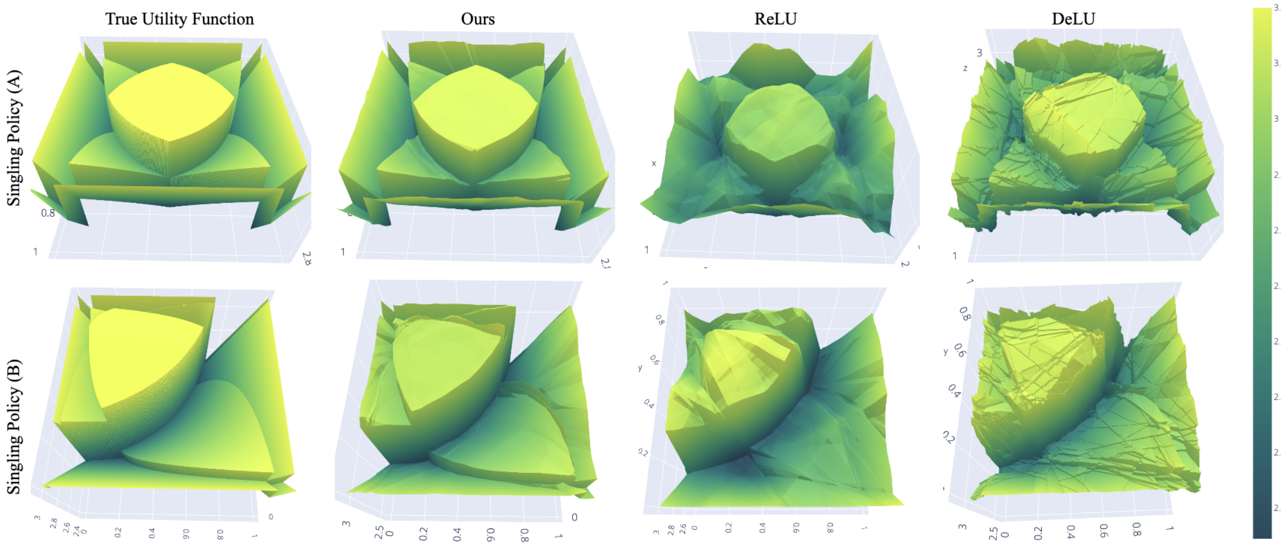

Figure 1. Discontinuous utility functions in a multi-sender persuasion game with 2 senders, 2 signals, 2 actions, and 2 states. In each subplot: the x-axis represents the probability of `Sender1` transmitting `Signal1` at `State1`, the y-axis shows the probability of `Sender2` emitting `Signal1` at `State1`, and the z-axis quantifies `Sender2`'s utility. Signaling strategies of both senders at `State2` are set to $(0.5, 0.5)$ in the top row and to $(0.2, 0.8)$ and $(0.8, 0.2)$ in the bottom row. In each column, we show the groundtruth ex-ante utility, and the approximation results achieved by our method, ReLU, and DeLU (Wang et al., 2023) networks, respectively.

(BP) framework. Senders can leverage their private observation of ω by strategically signaling the receiver. Formally, letting $\mathcal{S}$ be a finite signal space, each sender j has an independent *signaling policy* $\pi_j(s_j|\omega)$ which specifies the probability of sending signal $s_j \in \mathcal{S}$ when the realized state is ω. From the receiver's perspective, it observes a joint signal $\boldsymbol{s} = (s_1, \ldots, s_n)$ sampled from the joint conditional distribution $\boldsymbol{\pi}(\boldsymbol{s}|\omega) = \prod_{j=1}^{n} \pi_j(s_j|\omega)$. While many works on Bayesian persuasion assume the signal space $|\mathcal{S}| \geq |\mathcal{A}|$ (Kamenica & Gentzkow, 2011; Dughmi & Xu, 2016), we study the multi-sender problem in full generality (allowing $|\mathcal{S}| < |\mathcal{A}|$), since in many settings, action space can be arbitrarily large or even continuous (common in economic literature), but signaling/communication space may be limited. Consistent with the classical BP model, we assume that senders announce and commit to their signaling policies before observing state realizations. The receiver is considered Bayesian rational, and upon signal realization, it updates its belief about the state and takes a resulting optimal action according to its utility. We denote the interaction between senders and the receiver as a *multi-sender persuasion game* and summarize it as follows:

- All senders simultaneously announce their signaling policies $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$.
- State $\omega \sim \mu_0$ is observed by senders but not the receiver.
- Each sender j simultaneously draws a signal $s_j \sim \pi_j(\cdot|\omega)$ to send to the receiver. For $\boldsymbol{s} = (s_1, \ldots, s_n)$, $\boldsymbol{\pi}(\boldsymbol{s}|\omega) = \prod_{j=1}^{n} \pi_j(s_j|\omega)$ is the joint signal probability.
- After observing joint signal $\boldsymbol{s}$, the receiver forms posterior belief $\mu_{\boldsymbol{s}}$ about the state ($\mu_{\boldsymbol{s}}(\omega) = \frac{\mu_0(\omega)\boldsymbol{\pi}(\boldsymbol{s}|\omega)}{\boldsymbol{\pi}(\boldsymbol{s})}$ for every $\omega \in \Omega$) and takes an optimal action
$$a^*(\mu_{\boldsymbol{s}}) = \arg\max_{a \in \mathcal{A}} \mathbb{E}_{\omega \sim \mu_{\boldsymbol{s}}} v(\omega, a).$$
- Each sender j obtains utility $u_j(\omega, a^*(\mu_{\boldsymbol{s}}))$.

The senders attempt to use signaling to maximize their ex-ante utility, described below.

Definition 1. *The ex-ante utility for sender j under joint signaling policy $\boldsymbol{\pi} = (\pi_j, \boldsymbol{\pi}_{-j})$ is $\overline{u}_j(\boldsymbol{\pi}) = \sum_{\omega \in \Omega} \mu_0(\omega) \sum_{\boldsymbol{s} \in \mathcal{S}^n} \boldsymbol{\pi}(\boldsymbol{s}|\omega) u_j(\omega, a^*(\mu_{\boldsymbol{s}}))$, where $\mu_{\boldsymbol{s}}$ is the posterior distribution induced by joint signal $\boldsymbol{s}$ and policy $\boldsymbol{\pi}$, and $a^*(\mu_{\boldsymbol{s}})$ is the receiver's optimal (utility-maximizing) action at belief $\mu_{\boldsymbol{s}}$.*

The relationship between senders and the receiver forms a multi-leader-single-follower game since senders reveal their policies first and the receiver subsequently best responds to joint signal realizations generated by these policies. While the senders and the receiver have a sequential relationship, the senders choose their signaling policies simultaneously. Thus, we consider Nash equilibria among the senders:

Definition 2. *A Nash equilibrium (NE) for the multi-sender persuasion game is a profile of signaling policies $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ such that for any sender j and deviating policy π'_j, $\overline{u}_j(\boldsymbol{\pi}) \geq \overline{u}_j(\pi'_j, \boldsymbol{\pi}_{-j})$.*

The equilibrium defined above is in fact a subgame perfect equilibrium of the extensive-form game among the senders and the receiver. We use the term "Nash equilibrium" to emphasize the simultaneity of the senders' interaction.

3. Theoretical Results

We now look to theoretically understand the equilibrium properties of the multi-sender persuasion game. We first consider the canonical best response problem and show that solving it, even approximately, is NP-Hard. We then extend and generalize a known equilibrium characterization that relies on revealing maximal information to the receiver. This equilibrium is generally not ideal for senders and is possible only under certain conditions. Furthermore, in the general case, we show that equilibrium computation is PPAD Hard. Cumulatively, our strong intractability results together suggest that developing provably efficient algorithms for finding global equilibria would be extremely challenging in our setting.

3.1. Best Response

We first consider the *best response* problem for an sender; namely, fixing other senders' signaling schemes $\boldsymbol{\pi}_{-i}$, what is the optimal signaling scheme π_i that maximizes the ex-ante utility $\overline{u}_i(\pi_i, \boldsymbol{\pi}_{-i})$ of sender i? The best-response problem is essential to verifying whether a given joint signaling scheme $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ is a Nash equilibrium. Further, standard equilibrium solving techniques often rely on simulating best response dynamics.

In normal-form games, fixing others' strategies $\boldsymbol{x}_{-i}$, the utility $u_i(x_i, \boldsymbol{x}_{-i})$ of a player is linear in x_i, so the best response problem can be solved by a linear program efficiently. In persuasion, a sender's signaling policy changes the induced posteriors, which changes the optimal action the receiver takes since the receiver maximizes expected utility. Correspondingly, a sender's utility function $\overline{u}_i(\pi_i, \boldsymbol{\pi}_{-i})$ is piece-wise linear with discontinuities corresponding to signaling schemes wherein the mapping from signal realization to optimal receiver actions changes. This is more generally formalized in Proposition 1 (with proof in Appendix A.1).

Proposition 1 [Discontinuous Utility]. *The sender's utility function $\overline{u}_i(\boldsymbol{\pi})$ is discontinuous and piecewise non-linear in $(\pi_1, \ldots, \pi_n)$. Fixing $\boldsymbol{\pi}_{-i}$, $\overline{u}_i(\pi_i, \boldsymbol{\pi}_{-i})$ is discontinuous and piecewise linear in π_i.*

Maximizing $u_i(\pi_i, \boldsymbol{\pi}_{-i})$ by enumerating all linear pieces is infeasible because, by a rough estimate, the number of linear pieces can be as large as $O\big((|\mathcal{S}|^n|\mathcal{A}|^2)^{|\Omega||\mathcal{S}|}\big)$. Instead of enumerating all $O\big((|\mathcal{S}|^n|\mathcal{A}|^2)^{|\Omega||\mathcal{S}|}\big)$ linear pieces, we design a continuous bi-linear program to solve the best response problem (with proof in Appendix A.2).

Proposition 2 [Best Response Program]. *Let $\Delta v(\omega, a, a') \triangleq v(\omega, a) - v(\omega, a')$ for actions a, a'. Then given others' signaling schemes $\boldsymbol{\pi}_{-i}$, sender i's best response can be solved by the following optimization program with $|\Omega||\mathcal{S}| + |\mathcal{S}|^n|\mathcal{A}|$ continous variables and $O(|\mathcal{S}|^n|\mathcal{A}|)$ constraints:*

$$\max_{\pi_i, y} \sum_{\omega \in \Omega} \sum_{\boldsymbol{s} \in \mathcal{S}^n} \mu_0(\omega) \boldsymbol{\pi}_{-i}(\boldsymbol{s}_{-i}|\omega) \pi_i(s_i|\omega) \sum_{a \in \mathcal{A}} u_i(\omega, a) y_{\boldsymbol{s},a}$$

$$\text{s.t.}\ \forall \omega : \sum_{s} \pi_i(s_i|\omega) = 1 \ \textit{and}\ \forall s_i, \omega : \pi_i(s_i|\omega) \geq 0$$

$$\forall \boldsymbol{s} : \sum_{a \in \mathcal{A}} y_{\boldsymbol{s},a} = 1 \ \textit{and}\ \forall \boldsymbol{s}, a \in \mathcal{A} : y_{\boldsymbol{s},a} \in [0, 1]$$

$$\forall \boldsymbol{s}, a' : \sum_{\substack{\omega \in \Omega \\ a \in \mathcal{A}}} \mu_0(\omega) \boldsymbol{\pi}_{-i}(\boldsymbol{s}_{-i}|\omega) \pi_i(s_i|\omega) \Delta v(\omega, a, a') y_{\boldsymbol{s},a} \geq 0.$$

The $y_{\boldsymbol{s},a} \in \{0, 1\}$ in the above program means whether the receiver takes action a given joint signal $\boldsymbol{s}$, which can be relaxed to the continuou range $[0, 1]$. Briefly, the program above takes inspiration from the persuasion setting with (1) a single sender and (2) $|\mathcal{S}| = |\mathcal{A}|$, where signals can be interpreted as an *action recommendation* and optimal signaling expressed as a linear program with an incentive compatibility constraint to ensure that the receiver follows the recommended action. In our setting, even if $|\mathcal{S}| = |\mathcal{A}|$, the receiver observes joint signals of size $|\mathcal{S}|^n$ and a single sender cannot unilaterally specify the joint scheme; only the marginal. Correspondingly, the program needs to resolve the action taken by the receiver and becomes a bi-linear optimization problem.

We next show that the best-response problem is NP-Hard, even with just two senders. This means that the above bi-linear program is not computationally tractable. This is a key result in our work and rules out even additively approximating to the best-response in polynomial time.

Theorem 3 [NP-hardness of Best Response]. *It is NP-hard to solve the best-response problem in multi-sender persuasion, even with additive approximation error $\frac{1}{|\Omega|^6}$ and only $n = 2$ senders (while $|\Omega|$ and $|\mathcal{A}|$ are large).*

The proof (in Appendix A.3) is technical and based on a non-trivial reduction from the NP-hard problem *public persuasion with multiple receivers* (Dughmi & Xu, 2017). Intuitively, each signal $\boldsymbol{s}_{-i}$ from other non-responding senders induces a different belief about the state. From the best-responding sender's perspective, this can be correspondingly interpreted as facing multiple receivers with different prior beliefs and needing to design a single signaling scheme π_i for all of them. With carefully crafted utilities, this problem can encode the public persuasion problem (Dughmi & Xu, 2017), which involves multiple receivers with the same belief but different utility functions. Our proof formally establishes this connection, which leads to the NP-hardness of our problem.

The NP-hardness of computing best response, however, does not imply the hardness of *equilibrium verification*: i.e., determining whether a given strategy profile of the senders constitutes a Nash equilibrium. We conjecture that the equilibrium verification problem is Co-NP hard, whose formal proof would be an intriguing direction for future work.

3.2. Equilibrium Characterization

A simple observation from previous works on multi-sender Bayesian persuasion (Gentzkow & Kamenica, 2017b; Ravindran & Cui, 2022) is, if for every state $\omega \in \Omega$ there is a unique optimal action for the receiver, then a simple equilibrium can be achieved by all senders fully revealing the state - i.e. $\mathcal{S} = \Omega$ and $\pi_i(s_i = \omega|\omega) = 1, \forall i$. Observe that this reveals the exact state realization to the receiver and thus no sender can unilaterally affect the receiver's belief and thus their action. However, for this equilibrium to exist, every sender's signal space must be as large as the state space ($|\mathcal{S}| \geq |\Omega|$), which is impractical if there are many states. Theorem 4 relaxes this assumption, and shows that an equivalent equilibrium exists under a much weaker assumption, $|\mathcal{S}| \geq \min(|\mathcal{A}|^{\frac{1}{n-1}}, |\Omega|^{\frac{1}{n-1}})$, which can be easily satisfied when there are many senders. The proof of the theorem (in App. A.4) is constructive and builds a mapping between signals and actions inspired by grey codes (Wilf, 1989).

Theorem 4 [Full-Revelation Equilibrium]**.** *Suppose $|\mathcal{S}| \geq$ min$(|\Theta|^{1/(n-1)}|, \mathcal{A}|^{1/(n-1)}$, and for every state $\omega \in \Omega$ there is a unique optimal action for the receiver. Then, the multi-sender persuasion game has an NE that fully reveals the optimal action for the realized state to the receiver. This equilibrium, however, is not necessarily unique.*

While the result above generalizes an explicit equilibrium characterization to a much larger setting with limited signals, the corresponding equilibrium is optimal for the receiver and not necessarily the senders. Indeed, if the preferences of the senders do not perfectly align with the receiver (which is common and in-fact the premise behind persuasion), this will not be beneficial for the senders. Further, the construction above is based on the assumption that for every state, the receiver has a unique optimal action. This may not hold in many scenarios, with receivers being indifferent between multiple actions. We show that in such scenarios and thus the general case, finding an equilibrium is PPAD-Hard, even with constant number of senders, states, and actions. The proof (in App. A.5) relies on a reduction from finding equilibrium in two-player games with binary utilities.

Theorem 5 [PPAD-Hardness]**.** *In multi-sender persuasion games that do not satisfy the condition "the receiver has a unique optimal action for every state ω", under some tie-breaking rules, finding NE is PPAD-hard. This holds even if $n = 2$, $|\Omega| = 2$, $|\mathcal{A}| = 4$ (while $|\mathcal{S}|$ is large).*

4. Deep Learning for Local Equilibrium

The strong computational hardness results established in the previous section motivate us to find methods to efficiently calculate *local* Nash equilibrium, a strategy profile wherein any small unilateral deviation cannot improve a player's utility. This has been promoted as an attractive solution concept for a plethora of settings (Fiez et al., 2020; 2021; Jin et al., 2020). In doing so, we also relax the assumption of having access to the exact utility model and take a sample-based approach popularized in the nascent literature on differentiable economics. This is especially prescient as it gracefully generalizes to settings where action space is rich (or even continuous) and it may only be possible to sample utilities for arbitrary policies. Correspondingly, we introduce a computational framework based on deep learning. It consists of a novel discontinuous neural network architecture approximating the senders' utility functions and a local equilibrium solver running extra-gradients on the learned discontinuous networks. We describe the learning framework in detail and compare the found local NE against those obtained by strong baseline network structures as well as the full revelation solution (Theorem 4).

Definition 3 [ϵ-Local Nash Equilibrium]**.** *An ϵ-local Nash equilibrium for a multi-sender persuasion game is a profile of signaling policies $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$ such that for any sender j and deviating policy $\pi'_j \in \{\pi' \mid \|\pi' - \pi_j\| \leq \epsilon\}$, it holds that $\overline{u}_j(\boldsymbol{\pi}) \geq \overline{u}_j(\pi'_j, \boldsymbol{\pi}_{-j})$.*

4.1. Method

We aim to use the extra-gradient (Korpelevich, 1976) method to find an ϵ-local NE. However, the major challenge in applying this, or indeed any other gradient-based learning algorithm, is that the senders' utility function is discontinuous and non-differentiable in their signaling policy, as per Proposition 1. Conventional neural networks well approximate continuous functions but are not expressive enough to express discontinuous functions (Scarselli & Tsoi, 1998). To solve this problem, we extend a fully connected feedforward network with ReLU activation (Agarap, 2018) to learn a differentiable representation of discontinuous functions. To describe our method, we first introduce the activation pattern and the piecewise linearity of ReLU networks.

ReLU networks Suppose there are L hidden layers. Layer l has weights $\boldsymbol{W}^{(l)} \in \mathbb{R}^{n_l \times n_{l-1}}$ and biases $\boldsymbol{b}^{(l)} \in \mathbb{R}^{n_l}$. $n_0 = d$ is the input dimension. The output layer has weights $\boldsymbol{W}^{(L+1)} \in \mathbb{R}^{d' \times n_L}$ and biases $\boldsymbol{b}^{(L+1)} \in \mathbb{R}^{d'}$. With input $\boldsymbol{x} \in \mathbb{R}^d$, we have the pre- and post-activation output of layer l: $\boldsymbol{h}^{(l)}(\boldsymbol{x}) = \boldsymbol{W}^{(l)}\boldsymbol{o}^{(l-1)}(\boldsymbol{x}) + \boldsymbol{b}^{(l)}$ and $\boldsymbol{o}^{(l)}(\boldsymbol{x}) = \sigma\left(\boldsymbol{h}^{(l)}(\boldsymbol{x})\right)$, where $\sigma(x) = \max\{x, 0\}$ is the ReLU activation. For each hidden unit, the ReLU *activation status* has two values, defined as 1 when pre-activation h

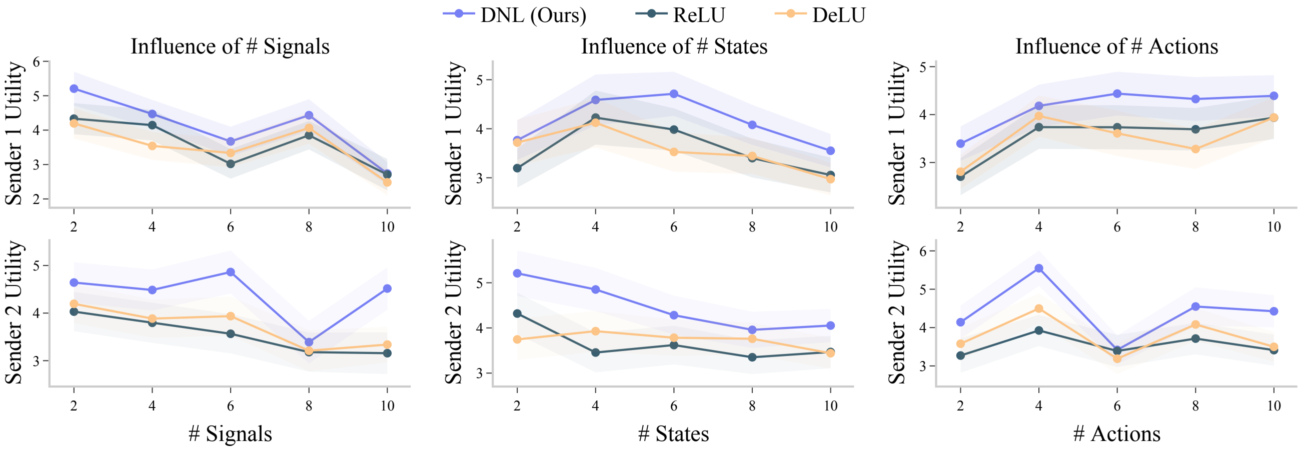

Figure 2. Our method finds better ϵ-local Nash equilibrium than the baseline DeLU (Wang et al., 2023) and ReLU networks.

is positive and 0 when h is strictly negative. The activation pattern of the entire network is defined as follows.

Definition 4 [Activation Pattern]**.** *An* activation pattern *of a ReLU network is a binary vector* $\boldsymbol{r} = [\boldsymbol{r}^{(1)}, \cdots, \boldsymbol{r}^{(L)}] \in \{0,1\}^{\sum_{l=1}^{L} n_l}$*, where* $\mathbf{r}^{(l)}$ *is a* layer activation pattern *including the activation status of each unit in layer* l*.*

The activation pattern depends on the input $\boldsymbol{x}$. Given an activation pattern $\boldsymbol{r}(\boldsymbol{x})$, the ReLU network is a linear function (Croce et al., 2019)

$$\boldsymbol{h}^{(L+1)}(\boldsymbol{x}) = \boldsymbol{M}^{(L+1)}\boldsymbol{x} + \boldsymbol{z}^{(L+1)},$$

where $\boldsymbol{M}^{(L+1)} = \boldsymbol{W}^{(L+1)}(\prod_{k=1}^{L} \boldsymbol{R}^{(L+1-k)}(\boldsymbol{x})\boldsymbol{W}^{(L+1-k)})$, $\boldsymbol{z}^{(L+1)} = \boldsymbol{b}^{(L+1)} + \sum_{k=1}^{L}(\prod_{j=0}^{L-k} \boldsymbol{W}^{(L+1-j)}\boldsymbol{R}^{(L-j)}(\boldsymbol{x}))\boldsymbol{b}^{(k)}$, and $\boldsymbol{R}^{(k)}$ is a diagonal matrix with diagonal elements equal to the layer k's activation pattern $\boldsymbol{r}^{(k)}$.

Previous work To introduce discontinuity, DeLU (Wang et al., 2023) proposes to generate the bias of the last layer $\boldsymbol{b}^{(L+1)}$ by an auxiliary network that is conditioned on the activation pattern $\boldsymbol{r}(\boldsymbol{x})$. The idea is that inputs with the same $\boldsymbol{r}(\boldsymbol{x})$ come from a polytope that is the intersection of half-spaces: $\mathcal{D}(\boldsymbol{x}) = \cap_{l=1,\cdots,L} \cap_{i=1,\cdots,n_l} \Gamma_{l,i}$, where $\Gamma_{l,i}$ corresponding to unit i of layer l defined as:

$$\Gamma_{l,i} = \left\{ \boldsymbol{y} \in \mathbb{R}^d | \Delta_i^{(l)} \left(\boldsymbol{M}_i^{(l)}\boldsymbol{y} + \boldsymbol{z}_i^{(l)} \right) \geq 0 \right\}. \quad (1)$$

Here $\boldsymbol{M}_i^{(l)}\boldsymbol{y} + \boldsymbol{z}_i^{(l)}$ is the output of unit i at layer l, and $\Delta_i^{(l)}$ is 1 if $\boldsymbol{h}_i^{(l)}(\boldsymbol{x})$ is positive, and is -1 otherwise.

In this way, different pieces $\mathcal{D}(\boldsymbol{x})$ has different biases, introducing discontinuity at piece boundaries. However, since inputs in the same piece share the same weights, DeLU is a linear function in a piece and does not have enough expressivity to represent the utility function in the multi-sender persuasion games, which is piecewise non-linear (Proposition 1).

Network architecture We enable a fully-connected network to be piecewise Discontinuous and Non-Linear (DNL) by dividing the network into a lower part and a higher part. The lower part consists of the first $K < L$ linear layers and is a normal network with ReLU activation. During a forward pass, we get the activation pattern

$$\boldsymbol{r}^{(\leq K)} = [\boldsymbol{r}^{(1)}, \cdots, \boldsymbol{r}^{(K)}]$$

of this lower network and generate the weights and biases of the higher part via a hyper-network g whose input is $\boldsymbol{r}^{(\leq K)}$.

Looking at the lower part, inputs with the same $\boldsymbol{r}^{(\leq K)}$ reside in the intersection of half-spaces:

$$\mathcal{D}^{(\leq K)}(\boldsymbol{x}) = \cap_{l=1,\cdots,K} \cap_{i=1,\cdots,n_l} \Gamma_{l,i},$$

with $\Gamma_{l,i}$ defined in Eq. 1. By introducing the hyper-network, inputs in $\mathcal{D}^{(\leq K)}(\boldsymbol{x})$ share a non-linear higher network. Therefore, within this piece, the utility approximation can be non-linear. Furthermore, different pieces have different $\boldsymbol{r}^{(\leq K)}$, so the higher part can be different, introducing discontinuity at boundaries.

Formally, we train a network $f_j(\boldsymbol{\pi}; \theta_j)$, parameterized by θ_j, for each sender j to approximate its ex-ante utility (Definition 1) under the joint signaling policy $\boldsymbol{\pi}$. The input to the lower part of f_j is the joint signaling policy $\boldsymbol{\pi}$. The hyper-network g takes the activation pattern $\boldsymbol{r}^{(\leq K)}(\boldsymbol{x})$ of the lower part as input and outputs $\{(\boldsymbol{W}^k, \boldsymbol{b}^k)\}_{k=K+1}^{L+1}$ as the weights and biases for layer $K+1$ to $L+1$. After obtaining its weights and biases, the higher part then takes the output of the lower part network as input and generates an approximation of the ex-ante utility. It is worth noting that $K < L$, and we have at least two linear layers at the higher part, so that piecewise non-linearity can be ensured. The whole network $f_j(\boldsymbol{\pi}; \theta_j)$ is end-to-end differentiable and updated by the MSE loss function:

$$\mathcal{L}(\theta_j) = \mathbb{E}_{\boldsymbol{\pi}} \left([f_j(\boldsymbol{\pi}; \theta_j) - \overline{u}_j(\boldsymbol{\pi})]^2 \right). \quad (2)$$

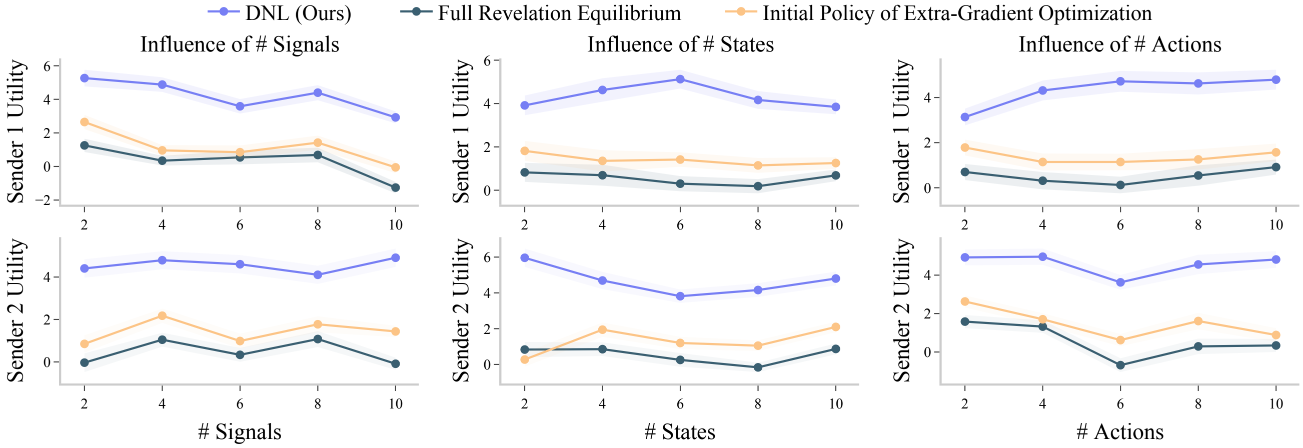

Figure 3. The ϵ-local Nash equilibria found by our method typically Pareto dominate the full revelation equilibria and improve the random initial policies of extra-gradient by a large margin.

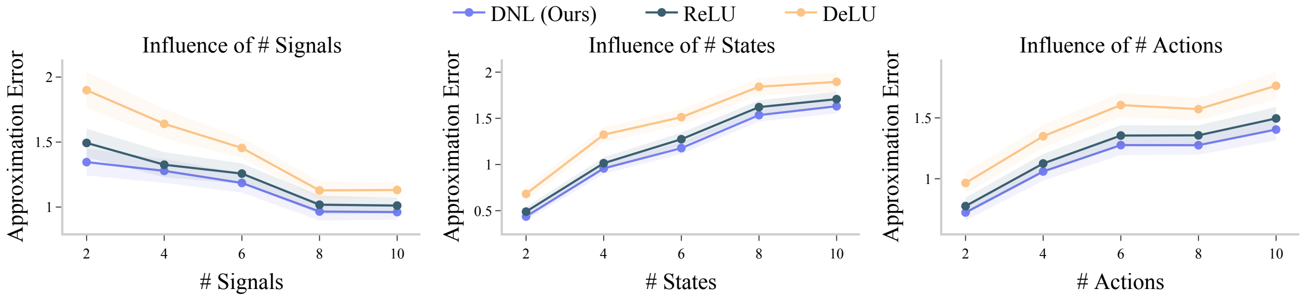

Figure 4. Our network achieves lower approximation errors compared to baseline network structures.

To calculate this loss, we uniformly sample joint policies $\boldsymbol{\pi}$ and obtain the corresponding ex-ante utility $\overline{u}_j(\boldsymbol{\pi})$ by running a game simulator.

Extra-gradient With f_j as a differentiable representation of the senders' ex-ante utility, we can run extra-gradients to find ϵ-local NE. We directly parameterize the signaling policy π_j of sender j by a learnable matrix ϕ_j residing in $\Phi \subset \mathbb{R}^{|\Omega|\times|\mathcal{S}|}$. A matrix in Φ has all of its elements in the range $[0, 1]$, and each row summed to 1.

The extra-gradient update can be written as

$$\text{(extrapolation) } \phi_j^{\tau+1/2} = p_\Phi(\phi_j^\tau - \gamma_\tau \nabla_{\phi_j^\tau} f_j(\boldsymbol{\pi}_{\phi^\tau}; \theta_j)),$$
$$\text{(update) } \phi_j^{\tau+1} = p_\Phi(\phi_j^\tau - \gamma_\tau \nabla_{\phi_j^{\tau+1/2}} f_j(\boldsymbol{\pi}_{\phi^{\tau+1/2}}; \theta_j)).$$

Here, $p_\Phi[\cdot]$ is the projection to the constraint set Φ, and we use a SoftMax projection in practice. The parameters θ_j of f_j is fixed during extra-gradient updates. $\boldsymbol{\pi}_\phi$ is the joint parameterized signaling policy, and γ_τ is the learning rate.

5. Empirical Results

In this section, we evaluate our deep learning method by comparing against continuous neural networks with ReLU activation, discontinuous neural networks DeLU (Wang et al., 2023), and full-revelation strategies on a illustrative example and a synthetic benchmark.

5.1. Didactic Example

We first demonstrate the representational capacity of our method on a simple multi-sender game with 2 senders, 2 signals, 2 actions, and 2 states. The utility matrix of the receiver is $\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$, where each row corresponds to a state and each column corresponds to an action. The utilities for two senders are $\left[\begin{smallmatrix}1&1\\-1&3\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}4&1\\1&1\end{smallmatrix}\right]$, respectively.

In the first row of Fig. 1, we fix both of the two senders' signaling policies at `State 2` to $(0.5, 0.5)$ and vary their signaling policies at `State 1`. The x-axis is the probability of `Sender 1` sending `Signal 1` at `State 1`, the y-axis is the probability of `Sender 2` sending `Signal 1` at `State 1`, and the z-axis is the (possibly approximated) ex-ante utility of `Sender 2`. The second row is similar to the first, but the two senders' signaling policies at `State 2` are $(0.2, 0.8)$ and $(0.8, 0.2)$, respectively.

The first column shows `Sender 2`'s actual ex-ante utility. This utility function displays discontinuities, effectively captured by our method (second column). In contrast, ReLU approximations in the third column are not accurate at piece boundaries, and we can observe that the approximated DeLU function in the fourth column is linear in each piece,

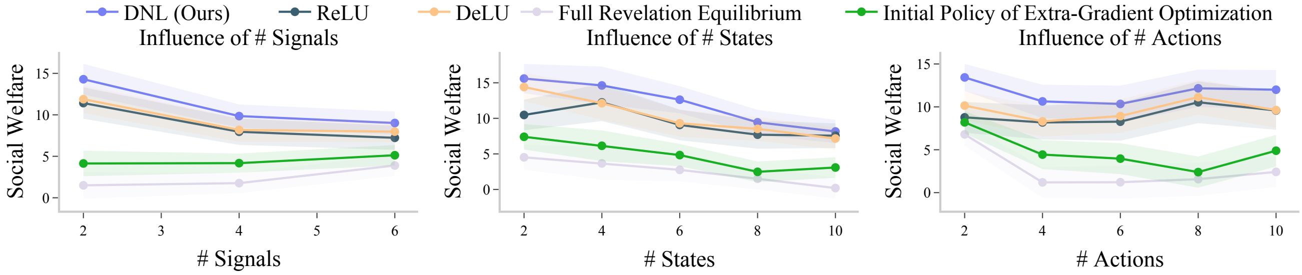

Figure 5. Our method achieves higher social welfare compared against baselines and full-revelation solutions in games with 4 senders.

limiting its representational power for this game.

To ensure a fair comparison, networks used in this study, including ours, ReLU, and DeLU, are standardized in terms of architecture, featuring three hidden layers with 64 units each. The training process involves a dataset of 500,000 randomly selected samples (pairs of signaling policies and corresponding ex-ante utilities), over which the networks are trained for a total of 200 epochs. For our network, the lower part has the first hidden layer. This layer's activation pattern is used to generate the weights and biases of the subsequent two layers by a hyper-network, which is itself composed of two hidden layers, each containing 32 units.

5.2. Synthetic Benchmark

In this section, we generate synthetic problems to test whether our network can find ϵ-local Nash equilibria that are better (in terms of sender utility) than those found by baseline network architectures as well as the full-revelation equilibria mentioned in Theorem 4.

Setup The size of a problem is determined by the tuple $(n, |\Omega|, |\mathcal{S}|, |\mathcal{A}|)$, and our evaluation encompasses a range of problem sizes to thoroughly assess the efficacy of our method. Specifically, we consider 2 and 4 senders, and for each, $(|\Omega|, |\mathcal{S}|, |\mathcal{A}|)$ are drawn from a three-dimensional grid $\{2, 4, 6, 8, 10\}^3$. For each problem size, we randomly generate 5 problem instances. In total, we have 1,250 problem instances to benchmark the proposed learning framework. The utility matrices for the receiver and senders, as well as the prior belief of states, are randomly sampled from a Gaussian distribution with variance at 100 and mean at 0, with a SoftMax applied to generate prior beliefs. We employ random numbers featuring significant variance to enhance the complexity of the benchmark, thereby facilitating a more effective evaluation of different solutions.

We standardize the network architecture of our method and baselines mirroring the configuration delineated in the didactic example to ensure a fair comparative analysis. The training setup is described in detail in Appendix B. To test whether a joint signaling policy profile $\boldsymbol{\pi}$ is an ϵ-local Nash equilibrium, we randomly sample K policies $\boldsymbol{\pi}'_j$ for each sender j in the neighborhood $\{\pi'_j \mid \|\pi'_j - \pi_{\phi_j}\|_\infty \leq \epsilon\}$ and check whether it can gain a higher utility at $\boldsymbol{\pi}'_j$. In our experiments, the number of test samples K grows with the problem size. We set the neighborhood size ϵ to 0.005 and find that our experimental results are robust with the value of ϵ up to 0.01 as evidenced by more results in Appendix B.

Representational capacity In Fig. 4, we fix the number of senders to 2 and compare the approximation errors (Eq. 2) achieved by our method and the two baseline architectures. We show the influence of the numbers of signals, states, and actions in three subplots, respectively, by presenting the average (solid lines) and the 95% confidence interval (shaded areas) of approximation errors. In the first subplot for example, we iterate the number of signals, and present results on all problem instances for each number of signals.

The results suggest that our algorithm provides a more accurate approximation than ReLU and DeLU. The advantage of our method is consistently maintained across all the range evaluated. It is also interesting to observe that the approximation error decreases for all three algorithms as the number of signals increases, but it increases as the numbers of states and actions increase. This observation indicates that the multi-sender persuasion game becomes more challenging with fewer signals, aligning with existing theoretical results on persuasion with limited signals (Dughmi et al., 2016).

Equilibrium In Fig. 2, we conduct a comparison between the equilibrium derived from our method against those produced by baselines. We run the verification process to ascertain whether the extra-gradient outcomes are indeed ϵ-local NE. We present the mean and the 95% confidence interval of the sender utilities at the best solutions that satisfy the criteria. Notably, our findings indicate that for each of the two senders, the ex-ante utility achieved in our model consistently outperforms that of the baselines, exhibiting Pareto dominance. In Fig. 3, we provide additional evidence demonstrating that extra-gradient with our trained networks can significantly enhance the senders' utility from the initial starting points. Furthermore, DNL successfully generates solutions that surpass full-revelation equilibria by a large margin. This improvement underscores the synergisitic benefit of integrating the extra-gradient approach with our networks. Similar results can be observed for games with

4 senders, and in Fig. 5, we show the welfare (the sum of senders' utilities) in these games.

5.3. Real-World Scenarios

Setting In this section, we extend the evaluation of our method to the following real-world scenarios.

Scenario 1: Advertising of Quality Prior economic research on multi-sender persuasion explored an advertising problem (Gentzkow & Kamenica, 2017a). In this problem, a total of n competing firms (senders) market their products to a single consumer. The product of each firm i can be of high quality ($\omega_i = 5$) or low quality ($\omega_i = -5$). The consumer wants to buy at most one product. The quality of the products is the state, known to firms but not the consumer. By sending signals, i.e., verifiable advertisements about their product's quality, a firm tries to persuade the consumer into purchasing from it, which induces utility 1 for the firm. The firm's utility is 0 if the consumer doesn't purchase from it. The consumer is faced with $n + 1$ actions, purchasing from any one of the firms, or none at all. The consumer's utility of purchasing from firm i is $\omega_i + \epsilon_i$, where ϵ_i is a shock (Gaussian-distributed zero-mean noise). If the consumer makes no purchase, their utility is 0. In our experiments, we set n to 7 and generate 20 instances randomly.

Scenario 2: Advertising of Multiple Products We make the previous advertising example more realistic by incorporating multiple products of different quality and prices. Specifically, we consider the following problem.

There are n firms (senders), each of which i sells a product of price p_i and quality ω_i. The true state is the prices and quality of all products. The consumer (receiver) has a partial observation of the state, as it has no access to the quality of products. The receiver has $n+1$ actions, which are buying a product from one of the firms or buying nothing. The utility of firm i is p_i if the receiver buys from it, or -1 otherwise. The senders use signals to strategically reveal the quality information to the receiver, trying to sway their purchase decisions in their favor. The receiver wants to maximize its utility, which is $\omega_i - p_i + \epsilon_i$ if purchasing product i, or 0 if buying nothing. Here ϵ_i is the shock defined in the same way as in Scenario 1. We test the case with $n = 2$ firms. Price p_i and quality ω_i are uniformly random integers in the range [1, 10] and [-8, 12], respectively.

Scenario 3: Uber or Lyft In this last scenario, we move beyond advertising and consider the competition among real-world ride-hailing apps, and a single driver subscribed to both platforms. There are two senders, Uber and Lyft, who receive m and n orders from users, respectively. Each order has four features (1) The price charged to the user; (2) The payment to the driver; (3) The true utility to the app, which is the price minus the payment; and (4) The true cost for the driver, which is known to the app and is influenced by many factors, such as the user rating indicating whether they are friendly, the expected travel time and distance, the expected waiting time, etc.

Table 1. The social welfare (avg±95% confidence interval) at the ϵ-local equilibria found by our method and baseline networks.

Scenario	ReLU	DeLU	Ours
1	0.498±0.004	0.599±0.003	**0.699±0.003**
2	0.407±0.176	0.467±0.179	**0.526±0.004**
3	3.216±0.790	3.783±0.894	**4.344±0.885**

The true state is the joint feature of all orders. Feature (4), the true cost to the driver, is invisible to the driver when they must decide the pickup. Uber and Lyft can send signals to strategically reveal this information in order to persuade the driver into picking up their orders. The driver has $m+n+1$ actions, which are picking up one of the $m + n$ orders or doing nothing. The utility of the driver is the price minus the true cost of the selected order, or -1 if they don't select any order. In our experiments, we set m and n to 4 and the number of signals to $m + 1$.

Results We test the performance of our method and the baseline neural network structures. Table 1 shows the social welfare of the senders at the ϵ-local equilibria found by different methods. Mean and 95% confidence intervals with 20 random instances are presented. We can observe that our method consistently outperforms other methods, indicating that our discontinuous, piecewise nonlinear network structure allows us to effectively tackle these richer settings that prior literature could not.

6. Discussion

We provide a comprehensive computational study of multi-sender Bayesian persuasion, a model for a wide range of real-world phenomena. The complex interplay of simultaneous sender actions and sequential receiver responses makes this game challenging. Our work formalizes this challenge by proving computational hardness results for both best response and equilibrium computation. Relaxing the equilibrium concept, however, offers hope, even without complete information. We propose a novel class of neural networks that can approximate the non-linear, discontinuous utilities in this game; paired with the extra-gradient algorithm, it is highly effective at finding local equilibria. Indeed, our network may be of broader interest to many games with discontinuous utility as it facilitates any downstream optimization algorithm. More broadly, BP is part of the principal-agent model of economics which also includes problems like contract design and Stackelberg games. Insights developed here can be instrumental to multi-principal variants of those problems which, despite their importance, have long eluded robust computational solutions.

Acknowledgements

We thank the reviewers for the helpful comments. This work is supported by NSF awards No. IIS-2147187 and No. CCF-2303372, Army Research Office Award No. W911NF-23-1-0030 and Office of Naval Research Award No. N00014-23-1-2802.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Abbott, T., Kane, D., and Valiant, P. On the Complexity of Two-PlayerWin-Lose Games. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pp. 113–122, Pittsburgh, PA, USA, 2005. IEEE. ISBN 978-0-7695-2468-9. doi: 10.1109/SFCS.2005.59. URL http://ieeexplore.ieee.org/document/1530706/.

Acemoglu, D., Ozdaglar, A., and Siderius, J. A model of online misinformation. Technical report, National Bureau of Economic Research, 2021.

Agarap, A. F. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

Aussel, D., Brotcorne, L., Lepaul, S., and von Niederhäusern, L. A trilevel model for best response in energy demand-side management. *European Journal of Operational Research*, 281(2):299–315, 2020.

Azizian, W., Mitliagkas, I., Lacoste-Julien, S., and Gidel, G. A tight and unified analysis of gradient-based methods for a whole spectrum of differentiable games. In *International conference on artificial intelligence and statistics*, pp. 2863–2873. PMLR, 2020.

Badanidiyuru, A., Bhawalkar, K., and Xu, H. Targeting and signaling in ad auctions. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2545–2563. SIAM, 2018.

Bai, Y., Jin, C., Wang, H., and Xiong, C. Sample-efficient learning of stackelberg equilibria in general-sum games. *Advances in Neural Information Processing Systems*, 34: 25799–25811, 2021.

Balduzzi, D., Racaniere, S., Martens, J., Foerster, J., Tuyls, K., and Graepel, T. The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, pp. 354–363. PMLR, 2018.

Basilico, N., Coniglio, S., and Gatti, N. Methods for finding leader–follower equilibria with multiple followers. *arXiv preprint arXiv:1707.02174*, 2017.

Bergemann, D. and Morris, S. Information Design: A Unified Perspective. *Journal of Economic Literature*, 57 (1):44–95, March 2019. ISSN 0022-0515. doi: 10.1257/jel.20181489. URL https://pubs.aeaweb.org/doi/10.1257/jel.20181489.

Bichler, M., Fichtl, M., Heidekrüger, S., Kohring, N., and Sutterer, P. Learning equilibria in symmetric auction games using artificial neural networks. *Nature machine intelligence*, 3(8):687–695, 2021.

Böhmer, W., Kurin, V., and Whiteson, S. Deep coordination graphs. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.

Bowling, M. Convergence and no-regret in multiagent learning. *Advances in neural information processing systems*, 17, 2004.

Brero, G., Eden, A., Chakrabarti, D., Gerstgrasser, M., Li, V., and Parkes, D. C. Learning stackelberg equilibria and applications to economic design games. *arXiv preprint arXiv:2210.03852*, 2022.

Bro Miltersen, P. and Sheffet, O. Send mixed signals: earn more, work less. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 234–247, 2012.

Calvete, H. I. and Galé, C. Linear bilevel multi-follower programming with independent followers. *Journal of Global Optimization*, 39(3):409–417, 2007.

Candogan, O. and Drakopoulos, K. Optimal signaling of content accuracy: Engagement vs. misinformation. *Operations Research*, 68(2):497–515, 2020.

Castiglioni, M., Celli, A., Marchesi, A., and Gatti, N. Online bayesian persuasion. *Advances in Neural Information Processing Systems*, 33:16188–16198, 2020.

Chen, X., Deng, X., and Teng, S.-H. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 56(3):1–57, May 2009. ISSN 0004-5411, 1557-735X. doi: 10.1145/1516512.1516516. URL https://dl.acm.org/doi/10.1145/1516512.1516516.

Cheng, C., Zhu, Z., Xin, B., and Chen, C. A multi-agent reinforcement learning algorithm based on stackelberg game. In *2017 6th Data Driven Control and Learning Systems (DDCLS)*, pp. 727–732. IEEE, 2017.

Christianos, F., Schäfer, L., and Albrecht, S. Shared experience actor-critic for multi-agent reinforcement learning.

Advances in neural information processing systems, 33: 10707–10717, 2020.

Conitzer, V. and Sandholm, T. Computing the optimal strategy to commit to. In *Proceedings of the 7th ACM conference on Electronic commerce*, pp. 82–90, 2006.

Croce, F., Andriushchenko, M., and Hein, M. Provable robustness of ReLU networks via maximization of linear regions. In *AISTATS 2019*, pp. 2057–2066, 2019.

Ding, B., Feng, Y., Ho, C.-J., Tang, W., and Xu, H. Competitive information design for pandora's box. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 353–381. SIAM, 2023.

Dong, H., Wang, T., Liu, J., and Zhang, C. Low-rank modular reinforcement learning via muscle synergy. *Advances in Neural Information Processing Systems*, 35: 19861–19873, 2022.

Dong, H., Zhang, J., Wang, T., and Zhang, C. Symmetry-aware robot design with structured subgroups. In *International Conference on Machine Learning*, pp. 8334–8355. PMLR, 2023.

Dughmi, S. Algorithmic information structure design: a survey. *ACM SIGecom Exchanges*, 15(2):2–24, February 2017. ISSN 1551-9031. doi: 10.1145/3055589.3055591. URL https://dl.acm.org/doi/10.1145/3055589.3055591.

Dughmi, S. and Xu, H. Algorithmic bayesian persuasion. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pp. 412–425, 2016.

Dughmi, S. and Xu, H. Algorithmic Persuasion with No Externalities. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 351–368, Cambridge Massachusetts USA, June 2017. ACM. ISBN 978-1-4503-4527-9. doi: 10.1145/3033274.3085152. URL https://dl.acm.org/doi/10.1145/3033274.3085152.

Dughmi, S., Kempe, D., and Qiang, R. Persuasion with limited communication. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 663–680, 2016.

Dütting, P., Feng, Z., Narasimhan, H., Parkes, D. C., and Ravindranath, S. S. Optimal auctions through deep learning: Advances in differentiable economics. *Journal of the ACM*, 2023.

Emek, Y., Feldman, M., Gamzu, I., PaesLeme, R., and Tennenholtz, M. Signaling schemes for revenue maximization. *ACM Transactions on Economics and Computation (TEAC)*, 2(2):1–19, 2014.

Fiez, T., Chasnov, B., and Ratliff, L. Implicit learning dynamics in stackelberg games: Equilibria characterization, convergence analysis, and empirical study. In *International Conference on Machine Learning*, pp. 3133–3144. PMLR, 2020.

Fiez, T., Ratliff, L., Mazumdar, E., Faulkner, E., and Narang, A. Global convergence to local minmax equilibrium in classes of nonconvex zero-sum games. *Advances in Neural Information Processing Systems*, 34:29049–29063, 2021.

Gan, J., Elkind, E., Kraus, S., and Wooldridge, M. Mechanism design for defense coordination in security games. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 402–410, 2020.

Gan, J., Han, M., Wu, J., and Xu, H. Generalized principal-agency: Contracts, information, games and beyond, 2024.

Gentzkow, M. and Kamenica, E. Bayesian persuasion with multiple senders and rich signal spaces. *Games and Economic Behavior*, 104:411–429, 2017a.

Gentzkow, M. and Kamenica, E. Bayesian persuasion with multiple senders and rich signal spaces. *Games and Economic Behavior*, 104:411–429, July 2017b. ISSN 08998256. doi: 10.1016/j.geb.2017.05.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0899825617300817.

Gerstgrasser, M. and Parkes, D. C. Oracles & followers: Stackelberg equilibria in deep multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 11213–11236. PMLR, 2023.

Goktas, D., Parkes, D. C., Gemp, I., Marris, L., Piliouras, G., Elie, R., Lever, G., and Tacchetti, A. Generative adversarial equilibrium solvers. *arXiv preprint arXiv:2302.06607*, 2023.

Guan, D.-J. Generalized gray codes with applications. In *PROC NATL SCI COUNC REPUB CHINA PART A PHYS SCI ENG*, volume 22, pp. 841–848. Citeseer, 1998.

Guestrin, C., Koller, D., and Parr, R. Multiagent planning with factored mdps. In *Advances in neural information processing systems*, pp. 1523–1530, 2002a.

Guestrin, C., Lagoudakis, M., and Parr, R. Coordinated reinforcement learning. In *ICML*, volume 2, pp. 227–234. Citeseer, 2002b.

Haghtalab, N., Lykouris, T., Nietert, S., and Wei, A. Learning in stackelberg games with non-myopic agents. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 917–918, 2022.

Hu, K., Ren, Z., and Yang, J. Principal-agent problem with multiple principals. *Stochastics*, 95(5):878–905, 2023.

Jaderberg, M., Czarnecki, W. M., Osindero, S., Vinyals, O., Graves, A., Silver, D., and Kavukcuoglu, K. Decoupled neural interfaces using synthetic gradients. In *International conference on machine learning*, pp. 1627–1635. PMLR, 2017.

Jelassi, S., Domingo-Enrich, C., Scieur, D., Mensch, A., and Bruna, J. Extragradient with player sampling for faster nash equilibrium finding. In *Proceedings of the International Conference on Machine Learning*, 2020.

Jiang, A. X., Procaccia, A. D., Qian, Y., Shah, N., and Tambe, M. Defender (mis) coordination in security games. In *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

Jiang, J., Dun, C., Huang, T., and Lu, Z. Graph convolutional reinforcement learning. In *International Conference on Learning Representations*, 2019.

Jin, C., Netrapalli, P., and Jordan, M. What is local optimality in nonconvex-nonconcave minimax optimization? In *International Conference on Machine Learning*, pp. 4880–4889. PMLR, 2020.

Kamenica, E. Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272, 2019.

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

Kang, Y., Wang, T., and de Melo, G. Incorporating pragmatic reasoning communication into emergent language. *Advances in Neural Information Processing Systems*, 33: 10348–10359, 2020.

Kang, Y., Wang, T., Yang, Q., Wu, X., and Zhang, C. Nonlinear coordination graphs. *Advances in Neural Information Processing Systems*, 35:25655–25666, 2022.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Korpelevich, G. M. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

Kuba, J. G., Chen, R., Wen, M., Wen, Y., Sun, F., Wang, J., and Yang, Y. Trust region policy optimisation in multi-agent reinforcement learning. In *International Conference on Learning Representations*, 2021.

Li, C., Wang, T., Wu, C., Zhao, Q., Yang, J., and Zhang, C. Celebrating diversity in shared multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34:3991–4002, 2021.

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

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

Mansour, Y., Slivkins, A., Syrgkanis, V., and Wu, Z. S. Bayesian exploration: Incentivizing exploration in bayesian games. *arXiv preprint arXiv:1602.07570*, 2016.

Myerson, R. B. Optimal coordination mechanisms in generalized principal–agent problems. *Journal of mathematical economics*, 10(1):67–81, 1982.

Naghizadeh, P. and Liu, M. Voluntary participation in cyber-insurance markets. In *Workshop on the Economics of Information Security (WEIS)*, 2014.

Peng, B., Rashid, T., Schroeder de Witt, C., Kamienny, P.-A., Torr, P., Böhmer, W., and Whiteson, S. Facmac: Factored multi-agent centralised policy gradients. *Advances in Neural Information Processing Systems*, 34: 12208–12221, 2021.

Pfau, D. and Vinyals, O. Connecting generative adversarial networks and actor-critic methods. *arXiv preprint arXiv:1610.01945*, 2016.

Rashid, T., Samvelyan, M., Witt, C. S., Farquhar, G., Foerster, J., and Whiteson, S. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 4292–4301, 2018.

Ravindran, D. and Cui, Z. Competing Persuaders in Zero-Sum Games, June 2022. URL http://arxiv.org/abs/2008.08517. arXiv:2008.08517 [econ].

Scarselli, F. and Tsoi, A. C. Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results. *Neural networks*, 11(1): 15–37, 1998.

Shi, Z., Yu, R., Wang, X., Wang, R., Zhang, Y., Lai, H., and An, B. Learning expensive coordination: An event-based deep rl approach. In *International Conference on Learning Representations*, 2019.

Shu, T. and Tian, Y. M^{3}RL: Mind-aware multi-agent management reinforcement learning. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019.

Tharakunnel, K. and Bhattacharyya, S. Leader-follower semi-markov decision problems: theoretical framework and approximate solution. In *2007 IEEE International*

Symposium on Approximate Dynamic Programming and Reinforcement Learning, pp. 111–118. IEEE, 2007.

Wang, K., Xu, L., Perrault, A., Reiter, M. K., and Tambe, M. Coordinating followers to reach better equilibria: End-to-end gradient descent for stackelberg games. *arXiv preprint arXiv:2106.03278*, 2021a.

Wang, T., Wang, J., Wu, Y., and Zhang, C. Influence-based multi-agent exploration. In *International Conference on Learning Representations*, 2019a.

Wang, T., Wang, J., Zheng, C., and Zhang, C. Learning nearly decomposable value functions via communication minimization. In *International Conference on Learning Representations*, 2019b.

Wang, T., Dong, H., Lesser, V., and Zhang, C. ROMA: Multi-agent reinforcement learning with emergent roles. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.

Wang, T., Gupta, T., Mahajan, A., Peng, B., Whiteson, S., and Zhang, C. RODE: Learning roles to decompose multi-agent tasks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021b.

Wang, T., Zeng, L., Dong, W., Yang, Q., Yu, Y., and Zhang, C. Context-aware sparse deep coordination graphs. In *International Conference on Learning Representations*, 2021c.

Wang, T., Dütting, P., Ivanov, D., Talgam-Cohen, I., and Parkes, D. C. Deep contract design via discontinuous piecewise affine neural networks. *arXiv preprint arXiv:2307.02318*, 2023.

Wang, Y., Han, B., Wang, T., Dong, H., and Zhang, C. Dop: Off-policy multi-agent decomposed policy gradients. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021d.

Waterman, R. W. and Meier, K. J. Principal-agent models: an expansion? *Journal of public administration research and theory*, 8(2):173–202, 1998.

Wen, M., Kuba, J., Lin, R., Zhang, W., Wen, Y., Wang, J., and Yang, Y. Multi-agent reinforcement learning is a sequence modeling problem. *Advances in Neural Information Processing Systems*, 35:16509–16521, 2022.

Wilf, H. S. *Combinatorial algorithms: an update*. SIAM, 1989.

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.

Yang, Q., Dong, W., Ren, Z., Wang, J., Wang, T., and Zhang, C. Self-organized polynomial-time coordination graphs. In *International Conference on Machine Learning*, pp. 24963–24979. PMLR, 2022.

Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., and Wu, Y. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35:24611–24624, 2022.

Zhang, E., Zhao, S., Wang, T., Hossain, S., Gasztowtt, H., Zheng, S., Parkes, D. C., Tambe, M., and Chen, Y. Social environment design. *arXiv preprint arXiv:2402.14090*, 2024.

Zhang, H., Xiao, Y., Cai, L. X., Niyato, D., Song, L., and Han, Z. A multi-leader multi-follower stackelberg game for resource management in lte unlicensed. *IEEE Transactions on Wireless Communications*, 16(1):348–361, 2016.

Zheng, S., Trott, A., Srinivasa, S., Parkes, D. C., and Socher, R. The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning, 2022. URL https://www.science.org/doi/abs/10.1126/sciadv.abk2607.

Zhu, B., Bates, S., Yang, Z., Wang, Y., Jiao, J., and Jordan, M. I. The sample complexity of online contract design. In *EC '23*, pp. 1188. ACM, 2023. ISBN 9798400701047. doi: 10.1145/3580507.3597673. URL https://doi.org/10.1145/3580507.3597673.

A. Proofs

A.1. Proof of Proposition 1

Proof. For a joint scheme $\boldsymbol{\pi}$, each signal realization $\boldsymbol{s}$ induces a posterior belief $\mu_{\boldsymbol{s}}$, wherein receiver take optimal action $a^*(\mu_{\boldsymbol{s}})$. We can equivalently write the function a^* in terms $\boldsymbol{\pi}(s|\cdot)$, ad note that $a^*(\boldsymbol{\pi}(s|\cdot)) \in \mathcal{A}$. When the signaling scheme changes sufficiently such that the new actions are optimal for a given realized posterior $\mu_{\boldsymbol{s}}$, the mapping $a^*(\boldsymbol{\pi}(\boldsymbol{s}|\cdot))$ changes accordingly. Thus, the function $a^*(\boldsymbol{\pi}(s|\cdot))$ is piece-wise constant with the boundary between pieces representing this changed mapping. The utility of sender i is given by:

$$\sum_{\omega\in\Omega}\sum_{\boldsymbol{s}\in\mathcal{S}^n} \mu_0(\omega) u_i(\omega, a^*(\boldsymbol{\pi}(\boldsymbol{s}|\omega)))\boldsymbol{\pi}(\boldsymbol{s}|\omega) \tag{3}$$

$$\sum_{\omega\in\Omega}\sum_{\boldsymbol{s}\in\mathcal{S}^n} \mu_0(\omega) u_i(\omega, a^*(\boldsymbol{\pi}(\boldsymbol{s}|\omega)))\prod_i \pi(s_i|\omega) \tag{4}$$

where we note that since u_i is essentially indexing a matrix, $u_i(\omega, a^*(\boldsymbol{\pi}(\boldsymbol{s}|\omega)))$ is piece-wise constant with the same boundaries as $a^*(\boldsymbol{\pi}(\boldsymbol{s}|\cdot))$. It is evident from the last expression above the utility is piece-wise bi-linear in $(\pi_1, \dots, \pi_n)$ and upon fixing $\boldsymbol{\pi}_{-i}$ is it piecewise linear in π_i. □

A.2. Proof of Proposition 2

Proof. Note that $\boldsymbol{\pi}_{-i}$ refer to the signaling of others and is fixed, with the optimization variables being π_i and $y_{\boldsymbol{s},a}$. Next, observe that if $y_{\boldsymbol{s},a} \in \{0,1\}$ then this optimization program can be interpreted as follows. π_i denotes the signaling scheme of influence i, and $y_{\boldsymbol{s},a}$ denotes whether action a is the optimal action for the user upon receiving the joint signal $\boldsymbol{s}$ and computing the corresponding posterior belief. The sum constraint on $y_{\boldsymbol{s},a}$ ensure $[y_{\boldsymbol{s},1}, \dots, y_{\boldsymbol{s},|\mathcal{A}|}]$ is a one hot vector. To ensure that the choice of $y_{\boldsymbol{s},a}$ are indeed correct, we need to ensure incentive-compatible. That is, we require the following holds for the posterior induced by any joint signal $\boldsymbol{s}$, and any action a': $\sum_\omega P(\omega|\boldsymbol{s}) \sum_{a\in\mathcal{A}} [v(a,\omega) - v(a',\omega)]\, y_{\boldsymbol{s},a}$. By Bayes rule, $P(\omega|\boldsymbol{s}) = \frac{\boldsymbol{\pi}(\boldsymbol{s}|\omega)\mu_0(\omega)}{P(\boldsymbol{s})}$ and since $P(\boldsymbol{s})$ is constant for the whole sum, we can multiply both sides by $P(\boldsymbol{s})$ and arrive at the first constraint in the above LP. Since this constraint enforces the choice of user action at each posterior indeed correct, the objective simply maximizes the sender's ex-ante expected utility.

The only difference between the presented optimization problem and the best-response sketched above is that the variables $y_{\boldsymbol{s},a}$ are now relaxed to within the continuous range $[0,1]$. We now show that this relaxation does not change the optimal solution. That is, an optimal solution to the binary-constrained setting is also an optimal solution to the relaxed continuous setting. Fix any signaling scheme π_i and any joint signal realization $\boldsymbol{s}$. Let $a^*_{\boldsymbol{s}}$ denote a best action for the user at the posterior induced by signal realization $\boldsymbol{s}$ with the schemes $(\pi_i, \boldsymbol{\pi}_{-i})$. Then we can rewrite the incentive compatibility constraint (first constraint) as follows (for brevity, we will write $\boldsymbol{\pi}(\boldsymbol{s}|\omega) = \pi_i(s_i|\omega) * \boldsymbol{\pi}_{-i}\boldsymbol{s}_{-i}|\omega$:

$$\sum_{a\in\mathcal{A}} y_{\boldsymbol{s},a}\left\{\sum_{\omega\in\Omega}\mu_0(\omega)\boldsymbol{\pi}(\boldsymbol{s}|\omega)v(w,a) - \sum_{\omega\in\Omega}\mu_0(\omega)\boldsymbol{\pi}(\boldsymbol{s}|\omega)v(w,a^*_{\boldsymbol{s}})\right\} \geq 0 \tag{5}$$

Note that the first summation term inside the inner bracket is proportional to the expected utility for action a under the posterior induced by $\boldsymbol{s}$, while the second summation term is the expected utility for action a under this same posterior. If $a^*_{\boldsymbol{s}}$ is the unique action that maximizes expected user utility at this posterior, then the only way this can be satisfied is by setting $y_{\boldsymbol{s},a^*_{\boldsymbol{s}}} = 1$ and 0 to all others. If multiple actions may be optimal for the receiver at this belief, then let $a^*_{\boldsymbol{s}}$ be the action among these that is most preferred by sender i (if there is a tie here, pick arbitrarily). Thus, by setting the corresponding $y_{\boldsymbol{s},a^*_{\boldsymbol{s}}} = 1$ and 0 for the rest satisfies the constraint while also maximizing user utility. Thus it follows that relaxing the domain of $y_{\boldsymbol{s},a}$ does not change the optimal solution since these still occur at the endpoints 0 or 1, and it follows that the continuous bi-linear optimization problem above corresponds to sender i's best response. □

A.3. Proof of Theorem 3

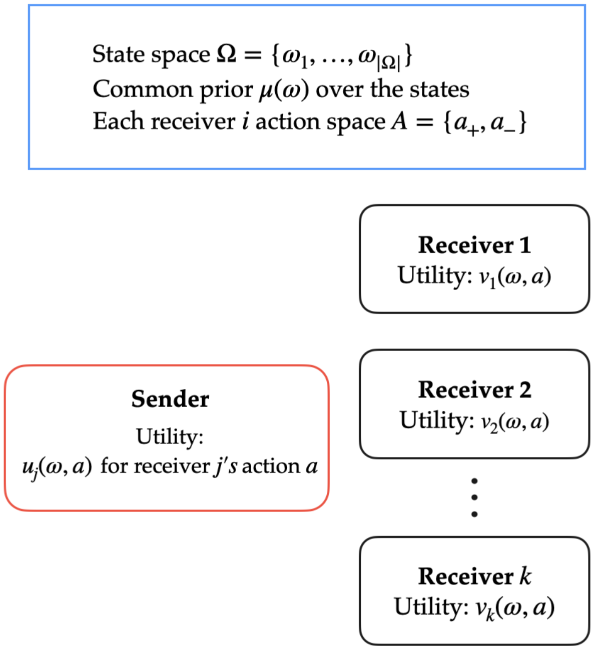
State space $\Omega = \{\omega_1, \ldots, \omega_{|\Omega|}\}$
Common prior $\mu(\omega)$ over the states
Each receiver i action space $A = \{a_+, a_-\}$
Receiver 1
Utility: $v_1(\omega, a)$
Receiver 2
Utility: $v_2(\omega, a)$
Receiver k
Utility: $v_k(\omega, a)$
Sender
Utility:
$u_j(\omega, a)$ for receiver j's action a

Figure 6. Public persuasion with k receivers, each with binary actions

The proof uses a reduction from the following problem called *public persuasion* (Dughmi & Xu, 2017):

Definition 5. *A* public persuasion (**Pub**) *problem (with multiple receivers with binary actions) is described by tuple* $\langle k, \Omega, \mu, \{v_j(\omega), u_j(\omega)\}_{j\in[k],\omega\in\Omega}\rangle$, *where:*

- *There are* k *receivers denoted by* $[k] = \{1, \ldots, k\}$ *each having two actions* $\{+, -\}$*.*
- $\mu \in \Delta(\Omega)$ *is a prior distribution of states* $\omega \in \Omega$*.*
- *Let* $v_j(\omega, +), v_j(\omega, -) \in [0, 1]$ *be the utilities of receiver* $j \in [k]$ *when taking actions* $+, -$ *and the state is* ω*. Let* $v_j(\omega) = v_j(\omega, +) - v_j(\omega, -) \in [-1, 1]$ *be the utility difference.*
- $u_j(\omega, +), u_j(\omega, -) \in [0, 1]$ *are the utilities of the sender when receiver* $j \in [k]$ *takes action* $+, -$*, respectively.*

Let $\pi : \Omega \to \Delta(S)$ be a signaling scheme of the sender. Let $x_s \in \Delta(\Omega)$ denote the posterior distribution over states induced by signal $s \in S$:

$$x_s(\omega) = \frac{\mu(\omega)\pi(s|\omega)}{\pi(s)} \quad \text{where} \quad \pi(s) = \sum_{\omega\in\Omega} \mu(\omega)\pi(s|\omega), \qquad \forall \omega \in \Omega.$$

In the public persuasion problem, given an induced posterior x_s, each receiver $j \in [k]$ is willing to take action $+$ if and only if $\sum_{\omega\in\Omega} x_s(\omega)v_j(\omega) \geq 0$. Let $a_j^*(x_s) \in \{+, -\}$ denote the action taken by receiver $j \in [k]$ given posterior x_s:

$$a_j^*(x_s) = \begin{cases} + & \text{if } \sum_{\omega\in\Omega} x_s(\omega)v_j(\omega) \geq 0 \\ - & \text{if } \sum_{\omega\in\Omega} x_s(\omega)v_j(\omega) < 0. \end{cases} \tag{6}$$

The sender's (expected) utility is the average utility obtained across all k receivers:

$$u^{\mathbf{Pub}}(\pi) = \sum_{s\in S} \pi(s) \frac{1}{k} \sum_{j=1}^{k} \sum_{\omega\in\Omega} x_s(\omega) u_j(\omega, a_j^*(x_s)). \tag{7}$$

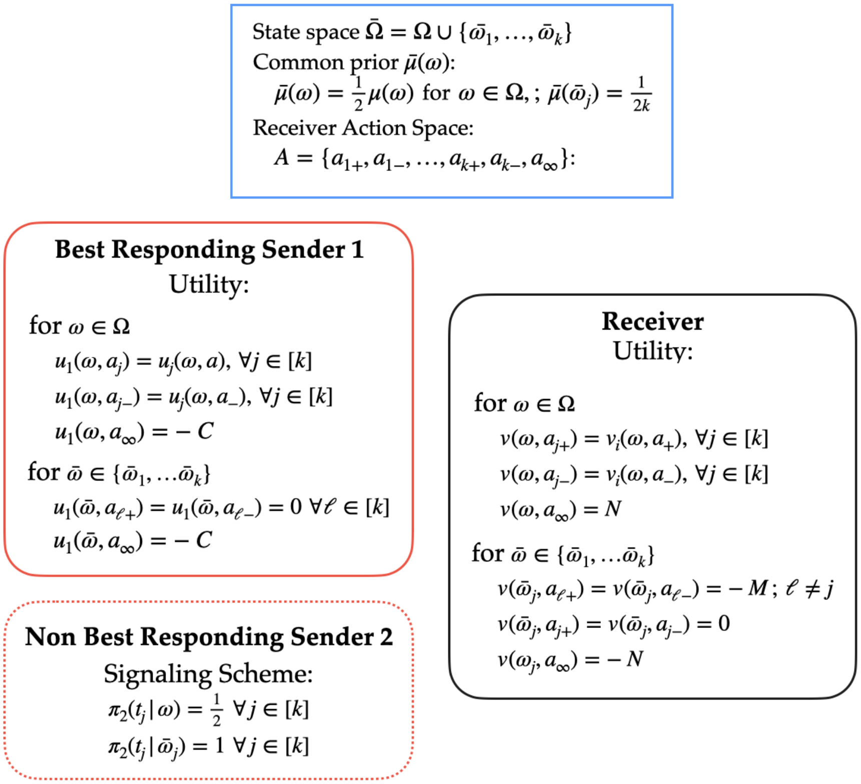
State space $\bar{\Omega} = \Omega \cup \{\bar{\varpi}_1, \ldots, \bar{\varpi}_k\}$
Common prior $\bar{\mu}(\omega)$:
$\bar{\mu}(\omega) = \frac{1}{2}\mu(\omega)$ for $\omega \in \Omega, ;$ $\bar{\mu}(\bar{\varpi}_j) = \frac{1}{2k}$
Receiver Action Space:
$A = \{a_{1+}, a_{1-}, \ldots, a_{k+}, a_{k-}, a_\infty\}$:
Best Responding Sender 1
Utility:
for $\omega \in \Omega$
$u_1(\omega, a_j) = u_j(\omega, a), \forall j \in [k]$
$u_1(\omega, a_{j-}) = u_j(\omega, a_-), \forall j \in [k]$
$u_1(\omega, a_\infty) = -C$
for $\bar{\varpi} \in \{\bar{\varpi}_1, \ldots \bar{\varpi}_k\}$
$u_1(\bar{\varpi}, a_{\ell+}) = u_1(\bar{\varpi}, a_{\ell-}) = 0 \ \forall \ell \in [k]$
$u_1(\bar{\varpi}, a_\infty) = -C$
Non Best Responding Sender 2
Signaling Scheme:
$\pi_2(t_j \mid \omega) = \frac{1}{2} \ \forall j \in [k]$
$\pi_2(t_j \mid \bar{\varpi}_j) = 1 \ \forall j \in [k]$
Receiver
Utility:
for $\omega \in \Omega$
$v(\omega, a_{j+}) = v_i(\omega, a_+), \forall j \in [k]$
$v(\omega, a_{j-}) = v_i(\omega, a_-), \forall j \in [k]$
$v(\omega, a_\infty) = N$
for $\bar{\varpi} \in \{\bar{\varpi}_1, \ldots \bar{\varpi}_k\}$
$v(\bar{\varpi}_j, a_{\ell+}) = v(\bar{\varpi}_j, a_{\ell-}) = -M \,; \ell \neq j$
$v(\bar{\varpi}_j, a_{j+}) = v(\bar{\varpi}_j, a_{j-}) = 0$
$v(\omega_j, a_\infty) = -N$

Figure 7. The multi-sender persuasion problem reduced from public persuasion. k additional states are added with the sole receiver having $2k+1$ actions. The receiver and best-responding sender's utility are chosen such that for all possible k possible signal realization of non-best responding sender, the single receiver's plausible actions mimic that of the k^{th} receiver in public persuasion.

The goal is to find a signaling scheme π to maximize $u^{\mathbf{Pub}}(\pi)$.

Theorem 6 [Dughmi & Xu (2017)]**.** *For any constant $c \in [0, \frac{1}{9}]$, it is NP-hard to solve, within additive approximation error c, public persuasion problems with $|\Omega| = k$ states and uniform prior $\mu(\omega) = \frac{1}{k}$, $\forall \omega \in [k]$.*

We reduce the public persuasion problem to the best-response problem in multi-sender persuasion, which will prove that the latter problem is NP-hard. Given the public persuasion problem $\langle k, \Omega, \mu, \{v_j(\omega), u_j(\omega)\}_{j \in [k], \omega \in \Omega} \rangle$, we construct the following best-response problem with two senders, where we fix sender 2's signaling scheme π_2 and find sender 1's best response: Let $C, N > 0$ and $M \geq \frac{N}{N-1}$ be some large numbers to be chosen later.

- There are $|\Omega| + k$ states, denoted by $\overline{\Omega} = \Omega \cup \{\overline{\varpi}_1, \ldots, \overline{\varpi}_j\}$, with prior $\overline{\mu}(\omega) = \frac{\mu(\omega)}{2}$ for $\omega \in \Omega$ and $\overline{\mu}(\overline{\varpi}_j) = \frac{1}{2k}$ for $j = 1, \ldots, k$.
- The (single) receiver has $2k+1$ actions, denoted by $A = \{a_{1+}, a_{1-}, \ldots, a_{k+}, a_{k-}\} \cup \{a_\infty\}$.
- The receiver's utility is:
  - For any state $\omega \in \Omega$, let

$$v(\omega, a_{j+}) = v_j(\omega)$$
$$v(\omega, a_{j-}) = 0$$
$$v(\omega, a_\infty) = N.$$

  In words, for any state $\omega \in \Omega$, the receiver's two actions a_{j+}, a_{j-} mimics receiver j's actions $+, -$ in the public persuasion problem. And a_∞ is very attractive to the receiver.

- For any state $\overline{\omega}_j$, $j \in [k]$, let

$$v(\overline{\omega}_j, a_{\ell+}) = v(\overline{\omega}_j, a_{\ell-}) = -M \text{ for } \ell \neq j$$
$$v(\overline{\omega}_j, a_{j+}) = v(\overline{\omega}_j, a_{j-}) = 0$$
$$v(\overline{\omega}_j, a_\infty) = -N.$$

In words, under state $\overline{\omega}_j$, the receiver is extremely unwilling to take actions other than $a_{j\pm}$. And a_∞ is very harmful to the receiver.

- Sender 1's utility $u_1(\cdot, \cdot)$ is:
  - For any state $\omega \in \Omega$,

$$u_1(\omega, a_{j+}) = u_j(\omega, +) \quad \forall j \in [k]$$
$$u_1(\omega, a_{j-}) = u_j(\omega, -) \quad \forall j \in [k]$$
$$u_1(\omega, a_\infty) = -C.$$

In words, when the receiver takes actions $a_{j\pm}$, the sender obtains the same utility as if receiver j takes action ± in the public persuasion problem. But the sender suffers a large loss if the receiver takes a_∞.

  - For any state $\overline{\omega}_j \in \Omega$, $j \in [k]$,

$$u_1(\overline{\omega}_j, a_{\ell+}) = u_1(\overline{\omega}_j, a_{\ell-}) = 0 \quad \forall \ell \in [k]$$
$$u_1(\overline{\omega}_j, a_\infty) = -C.$$

- Sender 2's signaling scheme π_2 is the following: it sends k possible signals $\{t_1, \ldots, t_k\}$ with probability:

$$\pi_2(t_j|\omega) = \frac{1}{k}, \ \ \forall j \in [k], \ \forall \omega \in \Omega.$$
$$\pi_2(t_j|\overline{\omega}_j) = 1, \ \ \forall j \in [k].$$

We sketch both the public persuasion framework and the equivalent multi-sender construction outlined above in Fig 6 and 7.

A.3.1. Useful Claims Regarding Receiver's Behavior

Before proving Theorem 3, we present some useful claims regarding the receiver's taking-best-action behavior. First, we characterize the receiver's expected utilities when taking different actions in the multi-sender persuasion problem:

Claim 1. *Let $x \in \Delta(\overline{\Omega})$ be a distribution on the enlarged state space $\overline{\Omega}$. Suppose sender 2 sends signal t_j. Then, the receiver's expected utilities of taking different actions $a \in A$, denoted by $v(x, t_j, a)$, are:*

- $v(x, t_j, a_{j+}) = \frac{1}{k}\sum_{\omega \in \Omega} x(\omega) v_j(\omega)$;
- $v(x, t_j, a_{j-}) = 0$;
- $v(x, t_j, a_{\ell+}) = \frac{1}{k}\sum_{\omega \in \Omega} x(\omega) v_\ell(\omega) - x(\overline{\omega}_j)M$ *for* $\ell \neq j$;
- $v(x, t_j, a_{\ell-}) = 0$ *for* $\ell \neq j$;
- $v(x, t_j, a_\infty) = N\big(\frac{1}{k}\sum_{\omega \in \Omega} x(\omega) - x(\overline{\omega}_j)\big)$.

Proof. For any $a \in A$, by definition,

$$\begin{aligned} v(x, t_j, a) &= \sum_{\omega \in \overline{\Omega}} x(\omega)\pi_2(t_j|\omega)v(\omega, a) \\ &= \sum_{\omega \in \Omega} x(\omega)\frac{1}{k}v(\omega, a) + \sum_{\ell=1}^{k} x(\overline{\omega}_\ell)\pi_2(t_j|\overline{\omega}_\ell)v(\overline{\omega}_\ell, a) \\ &= \frac{1}{k}\sum_{\omega \in \Omega} x(\omega)v(\omega, a) + x(\overline{\omega}_j)v(\overline{\omega}_j, a). \end{aligned}$$

Plugging in the definitions of utilities $v(\omega, a)$ and $v(\overline{\omega}_j, a)$ for different a proves the claim. □

As corollaries of the above claim, we have some guarantees when the receiver takes a best action:

Claim 2. *Given belief $x \in \Delta(\overline{\Omega})$ and sender 2's signal t_j, if the receiver does not take a_∞ as the best action, then it must be $\frac{1}{k}\sum_{\omega\in\Omega} x(\omega) \le \frac{N}{N-1} x(\overline{\omega}_j)$.*

Proof. If $\frac{1}{k}\sum_{\omega\in\Omega} x(\omega) > \frac{N}{N-1} x(\overline{\omega}_j)$, then by Claim 1, the receiver's utility of taking action a_∞ is

$$v(x, t_j, a_\infty) = N\Big(\frac{1}{k}\sum_{\omega\in\Omega} x(\omega) - x(\overline{\omega}_j)\Big) > N\Big(\frac{1}{k}\sum_{\omega\in\Omega} x(\omega) - \frac{N-1}{N}\frac{1}{k}\sum_{\omega\in\Omega} x(\omega)\Big) = \frac{1}{k}\sum_{\omega\in\Omega} x(\omega) \ge v(x, t_j, a)$$

for any other actions $a \neq a_\infty$. So, the receiver should take a_∞, a contradiction. □

Claim 3. *Given belief $x \in \Delta(\overline{\Omega})$ and sender 2's signal t_j, if the receiver is unwilling to take a_∞, then the receiver's utility of taking $a_{\ell\pm}$ for $\ell \neq j$ is $v(x, t_j, a_{\ell\pm}) \le 0$. So, we can assume that the receiver will take a_{j+} or a_{j-}. (Tie-breaking does not affect our conclusion.)*

Proof. According to Claim 2, if the receiver is unwilling to take a_∞, then $\frac{1}{k}\sum_{\omega\in\Omega} x(\omega) \le \frac{N}{N-1} x(\overline{\omega}_j)$. This implies that the receiver's utility of taking $a_{\ell+}$ is, by Claim 1

$$v(x, t_j, a_{\ell+}) = \frac{1}{k}\sum_{\omega\in\Omega} x(\omega) v_\ell(\omega) - x(\overline{\omega}_j) M \le \frac{N}{N-1} x(\overline{\omega}_j) v_\ell(\omega) - x(\overline{\omega}_j) M \le x(\overline{\omega}_j)\Big(\frac{N}{N-1} - M\Big) \le 0,$$

under the assumption of $v_\ell(\omega) \le 1$ and $M \ge \frac{N}{N-1}$. □

Claim 4. *Let $x \in \Delta(\overline{\Omega})$ be a belief on $\overline{\Omega}$. And let $\widetilde{x} \in \Delta(\Omega)$ be the conditional belief on Ω: $\widetilde{x}(\omega) = \frac{x(\omega)}{\sum_{\omega\in\Omega} x(\omega)}$, $\forall \omega \in \Omega$. Fix any $j \in [k]$. Suppose the receiver does not take action a_∞ under signal t_j in the multi-sender persuasion problem. Then, the receiver takes action a_{j+} (and a_{j-}) if and only if the receiver j in the public persuasion problem takes action $+$ (and $-$) under belief $\widetilde{x}$.*

Proof. By Claim 3, the receiver in the multi-sender problem must take action a_{j+} or a_{j-} if not taking a_∞. The receiver is willing to take a_{j+} if and only if, by Claim 1,

$$\frac{1}{k}\sum_{\omega\in\Omega} x(\omega) v_j(\omega) \ge 0 \iff \sum_{\omega\in\Omega} \frac{x(\omega)}{\sum_{\omega\in\Omega} x(\omega)} v_j(\omega) \ge 0 \iff \sum_{\omega\in\Omega} \widetilde{x}(\omega) v_j(\omega) \ge 0,$$

which means that the receiver j in the public persuasion problem is willing to take action $+$ under belief $\widetilde{x}$ (see (6)). □

A.3.2. Proof of Theorem 3

Consider a signaling scheme $\pi_1 : \overline{\Omega} \to \Delta(S)$ of sender 1, where S is the signal space. For a signal $s \in S$, let $x_s \in \Delta(\overline{\Omega})$ be the posterior distribution over $\overline{\Omega}$ given s. And let $\pi_1(s) = \sum_{\omega\in\overline{\Omega}} \overline{\mu}(\omega)\pi_1(s|\omega)$ be the probability of sender 1 sending signal s. A valid signaling scheme π_1 must satisfy the following Bayesian plausibility condition:

$$\sum_{s\in S} \pi_1(s) x_s = \overline{\mu} \iff \begin{cases} \sum_{s\in S} \pi_1(s) x_s(\omega) = \overline{\mu}(\omega) = \frac{\mu(\omega)}{2} & \text{for } \omega \in \Omega \\ \sum_{s\in S} \pi_1(s) x_s(\overline{\omega}_j) = \overline{\mu}(\overline{\omega}_j) = \frac{1}{2k} & \text{for } j \in [k] \end{cases}. \tag{8}$$

Let $a^*(x_s, t_j)$ be the best action that the receiver will take when the posterior induced by sender 1 is x_s (namely, sender 1 sends signal s) and sender 2 sends signal t_j. According to Claim 3, we have

$$a^*(x_s, t_j) \in \{a_\infty, a_{j+}, a_{j-}\}. \tag{9}$$

Let S_∞ be the set of signals of sender 1 for which the receiver will take action a_∞ given some signal t_j from sender 2:

$$S_\infty = \Big\{ s \in S \;\Big|\; a^*(x_s, t_j) = a_\infty \text{ for some } j \in [k] \Big\}.$$

Since a_∞ is very harmful to sender 1 (causing utility $-C$), we show that the total probability of S_∞ cannot be too large.

Lemma 1. *If sender 1's expected utility under signaling scheme π_1 is ≥ 0, then $\pi_1(S_\infty) = \sum_{s\in S_\infty} \pi_1(s) \leq \frac{2k}{C} + \frac{1}{2N}$.*

Proof. Sender 1's expected utility is (fixing sender 2's scheme),

$$\begin{aligned}
u_1(\pi_1) &= \sum_{s\in S} \pi_1(s) \Big[\sum_{\omega\in\overline{\Omega}} x_s(\omega) \sum_{j=1}^{k} \pi_2(t_j|\omega) u_1(\omega, a^*(x_s, t_j)) \Big] \\
&= \sum_{s\in S} \pi_1(s) \Big[\sum_{\omega\in\Omega} x_s(\omega) \sum_{j=1}^{k} \frac{1}{k} u_1(\omega, a^*(x_s, t_j)) + \sum_{j=1}^{k} x_s(\overline{\omega}_j) u_1(\overline{\omega}_j, a^*(x_s, t_j)) \Big] \\
&= \sum_{s\in S} \pi_1(s) \sum_{j=1}^{k} \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) u_1(\omega, a^*(x_s, t_j)) + x_s(\overline{\omega}_j) u_1(\overline{\omega}_j, a^*(x_s, t_j)) \Big] \qquad (10) \\
&= \sum_{s\in S_\infty} \pi_1(s) \sum_{j=1}^{k} \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) u_1(\omega, a^*(x_s, t_j)) + x_s(\overline{\omega}_j) u_1(\overline{\omega}_j, a^*(x_s, t_j)) \Big] \\
&\quad + \sum_{s\in S\setminus S_\infty} \pi_1(s) \sum_{j=1}^{k} \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) u_1(\omega, a^*(x_s, t_j)) + x_s(\overline{\omega}_j) u_1(\overline{\omega}_j, a^*(x_s, t_j)) \Big].
\end{aligned}$$

Since the utility $u_1(\omega, a)$ is always ≤ 1, and when receiver takes action a_∞ sender 1 gets utility $-C$,

$$\begin{aligned}
u_1(\pi_1) \leq & \sum_{s\in S_\infty} \pi_1(s) \sum_{j: a^*(x_s, t_j) = a_\infty} \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega)(-C) + x_s(\overline{\omega}_j)(-C) \Big] \\
& + \sum_{s\in S_\infty} \pi_1(s) \sum_{j=1}^{k} \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) \cdot 1 + x_s(\overline{\omega}_j) \cdot 1 \Big] \\
& + \sum_{s\in S\setminus S_\infty} \pi_1(s) \sum_{j=1}^{k} \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) \cdot 1 + x_s(\overline{\omega}_j) \cdot 1 \Big] \\
\leq & -C \sum_{s\in S_\infty} \pi_1(s) \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) + x_s(\overline{\omega}_j) \Big] \; + \; \underbrace{\sum_{s\in S} \pi_1(s) \sum_{j=1}^{k} \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) + x_s(\overline{\omega}_j) \Big]}_{=1 \text{ by } (12)}.
\end{aligned}$$

Using $u_1(\pi_1) \geq 0$ and rearranging, we get $\sum_{s\in S_\infty} \pi_1(s) \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) + x_s(\overline{\omega}_j) \Big] \leq \frac{1}{C}$, which implies

$$\sum_{s\in S_\infty} \pi_1(s) \Big[\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) \Big] \leq \frac{1}{C} \quad \Longrightarrow \quad \sum_{s\in S_\infty} \pi_1(s) \sum_{\omega\in\Omega} x_s(\omega) \leq \frac{k}{C}.$$

By the Bayesian plausibility condition (8), we have

$$\sum_{s\in S} \pi_1(s) \sum_{\omega\in\Omega} x_s(\omega) = \sum_{\omega\in\Omega} \sum_{s\in S} \pi_1(s) x_s(\omega) = \sum_{\omega\in\Omega} \overline{\mu}(\omega) = \frac{1}{2}.$$

So,

$$\sum_{s\in S\setminus S_\infty} \pi_1(s) \sum_{\omega\in\Omega} x_s(\omega) = \frac{1}{2} - \sum_{s\in S_\infty} \pi_1(s) \sum_{\omega\in\Omega} x_s(\omega) \geq \frac{1}{2} - \frac{k}{C}. \qquad (11)$$

For any signal $s \in S \setminus S_\infty$, the receiver does not take a_∞ under any t_j, which by Claim 2 implies

$$\frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega) \leq \frac{N}{N-1} x_s(\overline{\omega}_j) \quad \Longrightarrow \quad x_s(\overline{\omega}_j) \leq \frac{N-1}{N} \frac{1}{k} \sum_{\omega\in\Omega} x_s(\omega)$$

for all $j \in [k]$. Moreover, because

$$\sum_{j=1}^{k}\Big[\frac{1}{k}\sum_{\omega\in\Omega}x_s(\omega)+x_s(\overline{\omega}_j)\Big]=\sum_{\omega\in\overline{\Omega}}x(\omega)\sum_{j=1}^{k}\pi_2(t_j|\omega)=1, \tag{12}$$

we have for any $s \in S \setminus S_\infty$,

$$1\geq\sum_{j=1}^{k}\Big[\frac{1}{k}\sum_{\omega\in\Omega}x_s(\omega)+\frac{N-1}{N}\frac{1}{k}\sum_{\omega\in\Omega}x_s(\omega)\Big]=\Big(2-\frac{1}{N}\Big)\sum_{\omega\in\Omega}x_s(\omega)\quad\Longrightarrow\quad\sum_{\omega\in\Omega}x_s(\omega)\leq\frac{1}{2-\frac{1}{N}}. \tag{13}$$

From (11) and (13) we get

$$\frac{1}{2}-\frac{k}{C}\leq\sum_{s\in S\setminus S_\infty}\pi_1(s)\frac{1}{2-\frac{1}{N}}\quad\Longrightarrow\quad\sum_{s\in S\setminus S_\infty}\pi_1(s)\geq 1-\frac{2k}{C}-\frac{1}{2N},$$

which proves the lemma since $\sum_{s\in S}\pi_1(s)=1$. □

We give another characterization of π_1: for most of the signals in $S \setminus S_\infty$, the total posterior probability for states in Ω, $x_s(\Omega)=\sum_{\omega\in\Omega}x_s(\omega)$, should be close to $\frac{1}{2}$. Inequality (13) has shown an upper bound $\sum_{\omega\in\Omega}x_s(\omega)\leq\frac{1}{2-\frac{1}{N}}$. The following lemma gives a lower bound:

Lemma 2. *Fix any $\Delta > 0$. Let*

$$S_{\geq}=\Big\{s\in S\setminus S_\infty\;\Big|\;\frac{1}{2-\frac{1}{N}}\geq\sum_{\omega\in\Omega}x_s(\omega)\geq\frac{1}{2}-\Delta\Big\},\qquad S_{<}=\Big\{s\in S\setminus S_\infty\;\Big|\;\sum_{\omega\in\Omega}x_s(\omega)<\frac{1}{2}-\Delta\Big\}.$$

(Note that $S_{\geq}\cup S_{<}=S\setminus S_\infty$ by (13)). We have $\pi_1(S_{\geq})$ is large while $\pi_1(S_{<})$ is small:

$$\pi_1(S_{\geq})=\sum_{s\in S_{\geq}}\pi_1(s)\geq 1-\frac{2k}{C}-\frac{1}{2N}-\frac{1}{\Delta}\Big(\frac{1}{4N-2}+\frac{k}{C}\Big),$$
$$\pi_1(S_{<})=\sum_{s\in S_{<}}\pi_1(s)\leq\frac{1}{\Delta}\Big(\frac{1}{4N-2}+\frac{k}{C}\Big).$$

Proof. By (11),

$$\begin{aligned}\frac{1}{2}-\frac{k}{C}\leq\sum_{s\in S\setminus S_\infty}\pi_1(s)\sum_{\omega\in\Omega}x_s(\omega)\;&=\;\sum_{s\in S_{\geq}}\pi_1(s)\sum_{\omega\in\Omega}x_s(\omega)+\sum_{s\in S_{<}}\pi_1(s)\sum_{\omega\in\Omega}x_s(\omega)\\&\leq\frac{1}{2-\frac{1}{N}}\sum_{s\in S_{\geq}}\pi_1(s)+\big(\frac{1}{2}-\Delta\big)\sum_{s\in S_{<}}\pi_1(s)\\&\leq-\Delta\sum_{s\in S_{<}}\pi_1(s)+\frac{1}{2-\frac{1}{N}}\Big(\sum_{s\in S_{<}}\pi_1(s)+\sum_{s\in S_{\geq}}\pi_1(s)\Big)\\&\leq-\Delta\sum_{s\in S_{<}}\pi_1(s)+\frac{1}{2-\frac{1}{N}}\cdot 1\end{aligned}$$

So,

$$\sum_{s\in S_{<}}\pi_1(s)\leq\frac{1}{\Delta}\big(\frac{1}{2-\frac{1}{N}}-\frac{1}{2}+\frac{k}{C}\big)=\frac{1}{\Delta}\big(\frac{1}{4N-2}+\frac{k}{C}\big).$$

Together with Lemma 1, this implies $\pi_1(S_{\geq})=1-\pi_1(S_\infty)-\pi_1(S_{<})\geq 1-\frac{2k}{C}-\frac{1}{2N}-\frac{1}{\Delta}\big(\frac{1}{4N-2}+\frac{k}{C}\big)$. □

Now, we construct from π_1 a signaling scheme $\widetilde{\pi} : \Omega \to \Delta(\widetilde{S})$ for the public persuasion problem. The signal space of $\widetilde{\pi}$ is $\widetilde{S} = S_{\geq} \cup \{s_0\}$. For any $s \in S_{\geq}$, let the induced posterior $\widetilde{x}_s \in \Delta(\Omega)$ be

$$\widetilde{x}_s(\omega) = \frac{x_s(\omega)}{\sum_{\omega \in \Omega} x_s(\omega)}$$

(where x_s is the posterior induced by s in the signaling scheme π_1), and denote

$$\widetilde{\pi}(s) = \frac{\pi_1(s)}{\sum_{s \in S_{\geq}} \pi_1(s)} \geq \pi_1(s),$$

so $\sum_{s \in S_{\geq}} \widetilde{\pi}(s) = 1$. We will construct the posterior for s_0 later.

Lemma 3. *For any $\omega \in \Omega$,*

$$\Big| \sum_{s \in S_{\geq}} \widetilde{\pi}(s)\widetilde{x}_s(\omega) - \mu(\omega) \Big| \leq 4\Delta + \frac{2}{N} + \frac{4k}{C} + \frac{1}{\Delta}\Big(\frac{1}{2N-1} + \frac{2k}{C}\Big).$$

Proof. On the one hand,

$$\begin{aligned}
\sum_{s \in S_{\geq}} \widetilde{\pi}(s)\widetilde{x}_s(\omega) &\geq \sum_{s \in S_{\geq}} \pi_1(s) \frac{x_s(\omega)}{\sum_{\omega \in \Omega} x_s(\omega)} \geq \sum_{s \in S_{\geq}} \pi_1(s) \frac{x_s(\omega)}{\frac{1}{2 - \frac{1}{N}}} = \Big(2 - \frac{1}{N}\Big) \sum_{s \in S_{\geq}} \pi_1(s) x_s(\omega) \\
\text{by (8)} &= \Big(2 - \frac{1}{N}\Big)\Big(\frac{\mu(\omega)}{2} - \sum_{s \in S_\infty \cup S_<} \pi_1(s) x_s(\omega)\Big) \\
&\geq \Big(2 - \frac{1}{N}\Big)\Big(\frac{\mu(\omega)}{2} - \sum_{s \in S_\infty} \pi_1(s) - \sum_{s \in S_<} \pi_1(s)\Big) \\
\text{by Lemma 1 and 2} &\geq \Big(2 - \frac{1}{N}\Big)\Big(\frac{\mu(\omega)}{2} - \frac{2k}{C} - \frac{1}{2N} - \frac{1}{\Delta}\big(\frac{1}{4N-2} + \frac{k}{C}\big)\Big) \\
&\geq \mu(\omega) - \frac{\mu(\omega)}{2N} - \frac{4k}{C} - \frac{1}{N} - \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big) \\
&\geq \mu(\omega) - \frac{2}{N} - \frac{4k}{C} - \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\sum_{s \in S_{\geq}} \widetilde{\pi}(s)\widetilde{x}_s(\omega) &= \sum_{s \in S_{\geq}} \frac{\pi_1(s)}{\sum_{s \in S_{\geq}} \pi_1(s)} \frac{x_s(\omega)}{\sum_{\omega \in \Omega} x_s(\omega)} \\
\text{(by definition of } S_{\geq}\text{)} \ &\leq \sum_{s \in S_{\geq}} \frac{\pi_1(s)}{\sum_{s \in S_{\geq}} \pi_1(s)} \frac{x_s(\omega)}{\frac{1}{2} - \Delta} \\
&= \frac{2}{1 - 2\Delta} \frac{1}{\sum_{s \in S_{\geq}} \pi_1(s)} \sum_{s \in S_{\geq}} \pi_1(s) x_s(\omega) \\
&\leq \frac{2}{1 - 2\Delta} \frac{1}{\sum_{s \in S_{\geq}} \pi_1(s)} \sum_{s \in S} \pi_1(s) x_s(\omega) \\
\text{by (8)} \ &= \frac{2}{1 - 2\Delta} \frac{1}{\sum_{s \in S_{\geq}} \pi_1(s)} \frac{\mu(\omega)}{2} \\
\text{by Lemma 2} \ &\leq \frac{2}{1 - 2\Delta} \frac{1}{1 - \frac{2k}{C} - \frac{1}{2N} - \frac{1}{\Delta}\big(\frac{1}{4N-2} + \frac{k}{C}\big)} \frac{\mu(\omega)}{2} \\
&\leq \Big(1 + 4\Delta + \frac{4k}{C} + \frac{1}{N} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big)\Big)\mu(\omega) \\
&\leq \mu(\omega) + 4\Delta + \frac{4k}{C} + \frac{1}{N} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big).
\end{aligned}$$

Two above two cases together prove the lemma. □

As shown in Lemma 3, the signaling scheme $\widetilde{\pi}$ with signals in $S_{\geq}$ may not satisfy the Bayesian plausibility condition $\sum_{s\geq S_{\geq}} \widetilde{\pi}(s)\widetilde{x}_s = \mu(\omega)$. That is why we need the additional signal s_0. We want to find a posterior $y \in \Delta(\Omega)$ for signal s_0, and a coefficient $\alpha \in [0, 1]$ such that the following convex combination of $\{\widetilde{x}_s\}_{s\in S_{\geq}}$ and y satisfies Bayesian plausibility:

$$(1-\alpha)\sum_{s\in S_{\geq}} \widetilde{\pi}(s)\widetilde{x}_s + \alpha y = \mu. \tag{14}$$

Lemma 4. *Suppose* $\min_{\omega\in\Omega} \mu(\omega) \geq p_0 \geq 2\big(4\Delta + \frac{2}{N} + \frac{4k}{C} + \frac{1}{\Delta}(\frac{1}{2N-1} + \frac{2k}{C})\big) > 0$. *Then, there exists* $y \in \Delta(\Omega)$ *and* $\alpha \leq \frac{2}{p_0}\big(4\Delta + \frac{2}{N} + \frac{4k}{C} + \frac{1}{\Delta}(\frac{1}{2N-1} + \frac{2k}{C})\big)$ *that satisfy (14).*

Proof. Let $z = \sum_{s\in S_{\geq}} \widetilde{\pi}(s)\widetilde{x}_s$. By Lemma 3, we have

$$\|z-\mu\|_\infty = \max_{\omega\in\Omega} |z(\omega) - \mu(\omega)| \leq 4\Delta + \frac{2}{N} + \frac{4k}{C} + \frac{1}{\Delta}\Big(\frac{1}{2N-1} + \frac{2k}{C}\Big).$$

To satisfy (14), which is equivalent to

$$(1-\alpha)z + \alpha y = \mu \iff \alpha(y-z) = \mu - z,$$

we can let y be the intersection of the ray starting from z pointing towards μ and the boundary of $\Delta(\Omega)$. By doing this, $y - \mu$ and $z - \mu$ are in the same direction and

$$\alpha = \frac{\|\mu - z\|_\infty}{\|y - z\|_\infty}.$$

Since y is on the boundary of $\Delta(\Omega)$, some $y(\omega)$ must be 0. So,

$$\begin{aligned}\|y-z\|_\infty \geq \min_{\omega\in\Omega} z(\omega) &\geq \min_{\omega\in\Omega} \mu(\omega) - \Big(4\Delta + \frac{2}{N} + \frac{4k}{C} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big)\Big)\\ &\geq p_0 - \frac{p_0}{2} = \frac{p_0}{2}.\end{aligned}$$

This implies

$$\alpha \leq \frac{2}{p_0}\Big(4\Delta + \frac{2}{N} + \frac{4k}{C} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big)\Big).$$

□

With (14) satisfied, $\widetilde{\pi}$ now is a valid signaling scheme for the public persuasion problem, which sends signal $s \in S_{\geq}$ with probability $(1-\alpha)\widetilde{\pi}(s)$, inducing posterior $\widetilde{x}_s$, and sends signal s_0 with probability α, inducing posterior y. Let's consider the sender's utility (7) in the public persuasion problem using $\widetilde{\pi}$:

$$\begin{aligned}u^{\mathbf{Pub}}(\widetilde{\pi}) &= (1-\alpha)\sum_{s\in S_{\geq}} \widetilde{\pi}(s)\frac{1}{k}\sum_{j=1}^{k}\sum_{\omega\in\Omega} \widetilde{x}_s(\omega)u_j(\omega, a_j^*(\widetilde{x}_s)) + \alpha\frac{1}{k}\sum_{j=1}^{k}\sum_{\omega\in\Omega} y(\omega)u_j(\omega, a_j^*(y))\\ &\geq \sum_{s\in S_{\geq}} \widetilde{\pi}(s)\frac{1}{k}\sum_{j=1}^{k}\sum_{\omega\in\Omega} \widetilde{x}_s(\omega)u_j(\omega, a_j^*(\widetilde{x}_s)) - \alpha \qquad \text{because } 0 \leq u_j(\omega, a) \leq 1.\end{aligned}$$

By Claim 4, the receiver j's best action $a_j^*(\widetilde{x}_s)$ is $+$ (and $-$) if and only if the receiver in the multi-sender problem takes action $a^*(x_s, t_j) = a_{j+}$ (and a_{j-}) given posterior x_s from sender 1 and signal t_j from sender 2. So, by the definition of sender 1's utility in the multi-sender problem,

$$u_j(\omega, a_j^*(\widetilde{x}_s)) = u_1(\omega, a^*(x_s, t_j)).$$

Then, we have

$$\begin{aligned}
u^{\mathbf{Pub}}(\widetilde{\pi}) &\geq \sum_{s\in S_{\geq}} \widetilde{\pi}(s)\frac{1}{k}\sum_{j=1}^{k}\sum_{\omega\in\Omega}\widetilde{x}_s(\omega)u_1(\omega, a^*(x_s,t_j)) - \alpha \\
&\geq \sum_{s\in S_{\geq}} \pi_1(s)\frac{1}{k}\sum_{j=1}^{k}\sum_{\omega\in\Omega}\frac{x_s(\omega)}{\sum_{\omega\in\Omega}x_s(\omega)}u_1(\omega, a^*(x_s,t_j)) - \alpha \\
&\geq \Big(2-\frac{1}{N}\Big)\sum_{s\in S_{\geq}} \pi_1(s)\frac{1}{k}\sum_{j=1}^{k}\sum_{\omega\in\Omega}x_s(\omega)u_1(\omega, a^*(x_s,t_j)) - \alpha.
\end{aligned}$$

On the other hand, let's consider sender 1's utility in the multi-sender problem with signaling scheme π_1. By Equation (10),

$$\begin{aligned}
u_1(\pi_1) = &\sum_{s\in S}\pi_1(s)\sum_{j=1}^{k}\Big[\frac{1}{k}\sum_{\omega\in\Omega}x_s(\omega)u_1(\omega, a^*(x_s,t_j)) + x_s(\overline{\omega}_j)u_1(\overline{\omega}_j, a^*(x_s,t_j))\Big] \\
\leq &\sum_{s\in S_\infty\cup S_<}\pi_1(s)\cdot 1 \qquad\qquad \text{because } u_1(\cdot,\cdot)\leq 1 \\
&+ \sum_{s\in S_{\geq}}\pi_1(s)\sum_{j=1}^{k}\Big[\frac{1}{k}\sum_{\omega\in\Omega}x_s(\omega)u_1(\omega, a^*(x_s,t_j)) + x_s(\overline{\omega}_j) \underbrace{u_1(\overline{\omega}_j, a^*(x_s,t_j))}_{=0 \text{ because } a^*(x_s,t_j)\in\{a_{j+},a_{j-}\} \text{ from (9)}}\Big] \\
\leq &\frac{2k}{C} + \frac{1}{2N} + \frac{1}{\Delta}\Big(\frac{1}{4N-2}+\frac{k}{C}\Big) \qquad\qquad \text{by Lemma 1 and 2} \\
&+ \sum_{s\in S_{\geq}}\pi_1(s)\sum_{j=1}^{k}\Big[\frac{1}{k}\sum_{\omega\in\Omega}x_s(\omega)u_1(\omega, a^*(x_s,t_j))\Big]
\end{aligned}$$

So, $\sum_{s\in S_{\geq}}\pi_1(s)\sum_{j=1}^{k}\big[\frac{1}{k}\sum_{\omega\in\Omega}x_s(\omega)u_1(\omega,a^*(x_s,t_j))\big] \geq u_1(\pi_1) - \frac{2k}{C} - \frac{1}{2N} - \frac{1}{\Delta}\big(\frac{1}{4N-2}+\frac{k}{C}\big)$. This implies

$$u^{\mathbf{Pub}}(\widetilde{\pi}) \geq \Big(2-\frac{1}{N}\Big)\Big[u_1(\pi_1) - \frac{2k}{C} - \frac{1}{2N} - \frac{1}{\Delta}\big(\frac{1}{4N-2}+\frac{k}{C}\big)\Big] - \alpha. \tag{15}$$

Finally, we prove that if the signaling scheme π_1 is nearly optimal for the multi-sender best-response problem, then the corresponding scheme $\widetilde{\pi}$ for the public persuasion problem must be nearly optimal as well.

Claim 5. *If π_1 is approximately optimal for sender 1's best-response problem up to additive error c, then the $\widetilde{\pi}$ constructed above is approximately optimal for the public persuasion problem with additive error $2c + \frac{4k}{C} + \frac{2}{N} + \frac{1}{\Delta}\big(\frac{1}{2N-1}+\frac{2k}{C}\big) + \alpha$.*

Proof. Let π^* be the optimal signaling scheme for the public persuasion problem, which induces posterior $x^*_s \in \Delta(\Omega)$ at signal $s\in S^*$. Let π'_1 be the following signaling scheme for sender 1 in the multi-sender problem: for any signal $s\in S^*$, the probability of the signal is $\pi'_1(s) = \pi^*(s)$ and the induced posterior $x'_s \in \Delta(\overline{\Omega})$ is

$$x'_s(\omega) = \frac{x^*_s(\omega)}{2}, \quad x'_s(\overline{\omega}_j) = \frac{1}{2k}.$$

It is easy to verify that π'_1 is valid (satisfying Bayesian plausibility (8)). We then note that, at each posterior x'_s, the receiver's utility of taking action a_∞ is always 0 regardless of sender 2's signal t_j:

$$v(x'_s, t_j, a_\infty) = N\Big(\frac{1}{k}\sum_{\omega\in\Omega}x'_s(\omega) - x'_s(\overline{\omega}_j)\Big) = N\Big(\frac{1}{k}\frac{1}{2} - \frac{1}{2k}\Big) = 0.$$

So, we can assume that the receiver will take a_{j+} or a_{j-} by Claim 3. Moreover, by Claim 4, the receiver takes a_{j+} and a_{j-} if and only if the receiver j in the public persuasion problem with belief x^*_s takes action $+$ and $-$. So,

$$u_1(\omega, a^*(x'_s, t_j)) = u_j(\omega, a^*_j(x^*_s)).$$

This means that the utility of sender 1 in the multi-sender problem satisfies:

$$
\begin{aligned}
u_1(\pi'_1) &= \sum_{s^*\in S} \pi'_1(s) \sum_{j=1}^{k} \Big[\frac{1}{k}\sum_{\omega\in\Omega} x'_s(\omega) u_1(\omega, a^*(x_s, t_j)) + x'_s(\overline{\omega}_j) \underbrace{u_1(\overline{\omega}_j, a^*(x'_s, t_j))}_{=0 \text{ because } a^*(x'_s, t_j) \neq a_\infty} \Big] \\
&= \sum_{s^*\in S} \pi^*(s) \sum_{j=1}^{k} \Big[\frac{1}{k}\sum_{\omega\in\Omega} \frac{x^*_s(\omega)}{2} u_j(\omega, a^*_j(x^*_s))\Big] \\
&= \frac{1}{2}\sum_{s^*\in S} \pi^*(s) \frac{1}{k}\sum_{j=1}^{k}\sum_{\omega\in\Omega} x^*_s(\omega) u_j(\omega, a^*_j(x^*_s)) \;=\; \frac{1}{2} u^{\mathbf{Pub}}(\pi^*).
\end{aligned}
$$

If π_1 is approximately optimal up to additive error c in the multi-sender best-response problem, then

$$u_1(\pi_1) \geq u_1(\pi'_1) - c$$

Plugging this into (15),

$$
\begin{aligned}
u^{\mathbf{Pub}}(\widetilde{\pi}) &\geq \Big(2 - \frac{1}{N}\Big)\Big[u_1(\pi'_1) - c - \frac{2k}{C} - \frac{1}{2N} - \frac{1}{\Delta}\big(\frac{1}{4N-2} + \frac{k}{C}\big)\Big] - \alpha \\
&= \Big(2 - \frac{1}{N}\Big)\Big[\frac{1}{2}u^{\mathbf{Pub}}(\pi^*) - c - \frac{2k}{C} - \frac{1}{2N} - \frac{1}{\Delta}\big(\frac{1}{4N-2} + \frac{k}{C}\big)\Big] - \alpha \\
&\geq u^{\mathbf{Pub}}(\pi^*) - \frac{u^{\mathbf{Pub}}(\pi^*)}{2N} - 2c - \frac{4k}{C} - \frac{1}{N} - \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big)\Big] - \alpha \\
&\geq u^{\mathbf{Pub}}(\pi^*) - 2c - \frac{4k}{C} - \frac{2}{N} - \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big) - \alpha.
\end{aligned}
$$

This means that $\widetilde{\pi}$ is approximately optimal for the public persuasion problem up to additive error $2c + \frac{4k}{C} + \frac{2}{N} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big) + \alpha$. □

We now prove Theorem 3. Let $\langle k, \Omega, \mu, \{v_j(\omega), u_j(\omega)\}_{j\in[k],\omega\in\Omega}\rangle$ be any public persuasion problem with $|\Omega| = k$ states and uniform prior $\mu(\omega) = \frac{1}{k} = p_0$. Construct the multi-sender best-response problem as above (where the range of utility of sender 1 is $[-C, 1]$). If we can find an ϵ-approximately optimal signaling scheme π_1 for sender 1's best-response problem with utility range $[-1, 1]$, with

$$\epsilon = \frac{1}{k^6},$$

then π_1 is a $C\epsilon$-approximately optimal signaling scheme with utility range $[-C, 1]$. Then by Claim 5, the scheme $\widetilde{\pi}$ constructed above is approximately optimal for the public persuasion problem with additive error at most

$$
\begin{aligned}
& 2C\epsilon + \frac{4k}{C} + \frac{2}{N} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big) + \alpha \\
\text{by Lemma 4 } & \leq 2C\epsilon + \frac{4k}{C} + \frac{2}{N} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big) + \frac{2}{p_0}\Big(4\Delta + \frac{2}{N} + \frac{4k}{C} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big)\Big) \\
& \leq 2C\epsilon + \big(2k+1\big)\Big(4\Delta + \frac{2}{N} + \frac{4k}{C} + \frac{1}{\Delta}\big(\frac{1}{2N-1} + \frac{2k}{C}\big)\Big).
\end{aligned}
$$

Let $C = k^5, N = k^4, \Delta = \frac{1}{k^2}$.

$$\leq 2k^5\epsilon + \big(2k+1\big)\Big(\frac{4}{k^2} + \frac{2}{k^4} + \frac{4k}{k^5} + k^2\big(\frac{1}{2k^4-1} + \frac{2k}{k^5}\big)\Big) = O\big(\frac{1}{k}\big) \leq \frac{1}{9},$$

for sufficiently large k. Theorem 6 says that finding $\frac{1}{9}$-approximation for the public persuasion is NP-hard. So, finding $\epsilon = \frac{1}{k^6}$-approximation for the multi-sender best-response problem is NP-hard.

A.4. Proof of Theorem 4

Proof. Since at each state ω, there is a unique optimal action a it suffices to consider $|\mathcal{A}| \leq |\Omega|$. Next, let the signal space be $|S| = |\mathcal{A}|^{\frac{1}{n-1}} \triangleq k$; we shall see this is without loss of generality when the signal space is larger. We first give a construction for a mapping α between all k-ary strings of length n (all possible joint signals) to $|\mathcal{A}|$. Let ζ denote a subset of these strings such that for any two strings $\boldsymbol{s}^1 \in \zeta$, $\boldsymbol{s}^2 \in \zeta$ the hamming distance between them is at least two - $d_H(\boldsymbol{s}^1, \boldsymbol{s}^2) \geq 2$. The k-ary Gray code is an ordering of all unique k-ary strings of length n such that any two consecutive strings are exactly 1 apart in hamming distance. Such a construction is always possible (Guan, 1998). Since there are k different values possible at any position, within at least every k strings in the grey code, we should have two strings that are hamming distance 2 apart. Thus $|\zeta| \geq k^{n-1} = |\mathcal{A}|$. This is indeed tight since k^{n-1} is the total number of unique $n-1$ length k-ary strings possible - thus if $|\zeta| > k^{n-1}$, it would mean there are two strings where that match on $n-1$ positions, violating the construction of ζ. We construct α as follows: map each string in ζ to a unique action in $\mathcal{A}$ and assign the remaining joint signal strings arbitrarily to an action.

Under this mapping, we now give a constructive joint signaling scheme that is (1) a Pure Nash Equilibrium and (2) fully reveals the optimal action to the agent. Let $\alpha^{-1}(a)$ map to the joint signal $\boldsymbol{s} \in \zeta$ such that $\alpha(\boldsymbol{s}) = a$. Further, let $f : \Omega \to \mathcal{A}$ denote the unique agent-optimal action under state ω, with its inverse $f^{-1}(a)$ denoting the set of states for which this action is agent-optimal. Next, consider the following joint signaling scheme: for all $\boldsymbol{s} \in \zeta$, $\pi(\boldsymbol{s}|\omega) = 1$ if $\omega \in f^{-1}(\alpha(\boldsymbol{s}))$. That is for any ω, the joint signal $\boldsymbol{s} \in \zeta$ that corresponds to the optimal agent action under ω, i.e. $\alpha(\boldsymbol{s}) = f(\omega)$, is sent with probability 1. The agent can thus uniquely map each joint signal realization to a set of states wherein a fixed action is optimal. In other words, this fully reveals the optimal action for the agent at any state realization ω. To show this is a Nash equilibrium, observe that since all strings in ζ are hamming distance at least 2 apart, there is in fact a bijection between any $n-1$ sub-signal/sub-string within ζ and the action. Thus, each optimal action is fully specified by signals of just $n-1$ agents. So if a sender unilaterally shifts her signaling, the agent can observe that $n-1$ signals still uniquely map to states that share a common optimal action, and essentially ignore the deviating agent's signal. Thus, no change in agent belief or action occurs, leading the deviation to be non-beneficial. Since the choice of deviating agent here is arbitrary, this presented scheme is a pure Nash equilibrium.

However, full revelation equilibrium is not unique, which we show through an example. Consider $n = 2$ senders, with $|\mathcal{A}| = 4$ actions, and $|\Omega| = 4$ states, with the following prior: $[0.15, 0.35, 0.15, 0.35]$. Sender 1 has utility 1 whenever action 1 is taken and 0 otherwise. Similarly, sender 2 has utility 1 whenever action 3 is taken and 0 otherwise. Note both utilities are agnostic to the state ω. The receiver utility is given by the following matrix:

$$V = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \tag{16}$$

Under a full-revelation or optimal action revelation equilibrium, note that each sender would get utility 0.15. Now consider the following signaling scheme using only 3 signals, which we express as a $|\Omega| \times |\mathcal{S}|$ matrix[1].

$$\pi_1 = \begin{bmatrix} 0 & 1 & 0 \\ \frac{4}{7} & \frac{3}{7} & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \pi_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ \frac{4}{7} & \frac{3}{7} & 0 \end{bmatrix} \tag{17}$$

Joint signal realizations 01 and 12 from such a scheme induces the following posterior beliefs with probability 0.3:

$$\mu_{01} = [0, 0, 0.5, 0.5] \quad \mu_{12} = [0.5, 0.5, 0, 0] \tag{18}$$

Note that for any tie-breaking rule that favors senders, the posteriors above give utility 0.3 to both senders. All other posteriors have dominant actions that give 0 utilities to both senders. We can use the optimization program presented in proposition 2 to verify this is an equilibrium. □

[1] a scheme with 3 signals can without loss of generality be extended to a scheme with 4 signals, which is what the optimal receiver action revelation scheme uses.

A.5. Proof of Theorem 5

Proof. It is known that finding Nash equilibria in 2-player games with 0/1 utilities is PPAD-hard (Abbott et al., 2005; Chen et al., 2009). We reduce this PPAD-hard problem to multi-sender persuasion, which proves that the latter problem is also PPAD-hard. Let $\widehat{u}_1, \widehat{u}_2 \in \{0,1\}^{m \times m}$ be the utility matrices of the 2 players, where m is the number of actions of each player. We construct a multi-sender persuasion game as follows:

- There are 2 states $\Omega = \{0,1\}$ with prior $\mu_0(0) = \mu_0(1) = 1/2$, $|\mathcal{A}| = 4$ actions for the receiver labeled as $\mathcal{A} = \{a_{00}, a_{01}, a_{10}, a_{11}\}$, and $n = 2$ senders each having a signal space $\mathcal{S} = \{1, \ldots, m\}$.
- The receiver's utility is 0 regardless of actions and states, so he is indifferent among taking all actions. Suppose the receiver breaks ties in the following way: given joint signal (s_1, s_2) from the 2 senders, take action

$$\alpha(s_1, s_2) = \begin{cases} a_{00} & \text{if } \widehat{u}_1(s_1, s_2) = 0,\ \widehat{u}_2(s_1, s_2) = 0; \\ a_{01} & \text{if } \widehat{u}_1(s_1, s_2) = 0,\ \widehat{u}_2(s_1, s_2) = 1; \\ a_{10} & \text{if } \widehat{u}_1(s_1, s_2) = 1,\ \widehat{u}_2(s_1, s_2) = 0; \\ a_{11} & \text{if } \widehat{u}_1(s_1, s_2) = 1,\ \widehat{u}_2(s_1, s_2) = 1. \end{cases}$$

- The utility of each sender i is:

$$u_i(a, \omega = 1) = \begin{cases} \widehat{u}_i(s_1, s_2) & \text{if there exist } s_1, s_2 \in \mathcal{S} \text{ such that } \alpha(s_1, s_2) = a; \\ 0 & \text{otherwise.} \end{cases}$$
$$u_i(a, \omega = 0) = 0, \quad \forall a \in \mathcal{A}.$$

  We note that the first equation is well-defined, because for any different joint signals (s_1, s_2) and (s_1', s_2'), if they both satisfy $\alpha(s_1, s_2) = \alpha(s_1', s_2') = a$, then they define the same utility $u_i(a, \omega = 1) = \widehat{u}_i(s_1, s_2) = \widehat{u}_i(s_1', s_2')$.

We note that the expected utility of each sender i under signaling schemes $\boldsymbol{\pi} = (\pi_1, \pi_2)$ is equal to

$$\begin{aligned} \overline{u}_i(\boldsymbol{\pi}) &= \sum_{\omega \in \Omega} \sum_{\boldsymbol{s} \in \mathcal{S}^n} \mu_0(\omega) \boldsymbol{\pi}(\boldsymbol{s}|\omega) u_i(\alpha(\boldsymbol{s}), \omega) \\ &= \frac{1}{2} \cdot 0 + \frac{1}{2} \sum_{s_1, s_2} \pi_1(s_1|\omega = 1) \pi_2(s_2|\omega = 1) u_i(\alpha(s_1, s_2), \omega = 1) \\ &= \frac{1}{2} \sum_{s_1, s_2} \pi_1(s_1|\omega = 1) \pi_2(s_2|\omega = 1) \widehat{u}_i(s_1, s_2) \\ &= \frac{1}{2} \widehat{u}_i(x_1, x_2), \end{aligned}$$

where $\widehat{u}_i(x_1, x_2)$ is the expected utility of player i in the 2-player 0/1 utility game when the two players use mixed strategies x_1, x_2 where player i samples action $s_i \in \{1, \ldots, m\}$ with probability $x_i(s_i) = \pi_i(s_i|\omega = 1)$. If we can find an NE (π_1, π_2) for the multi-sender persuasion game, then the corresponding mixed strategy profile (x_1, x_2) where $x_i(s_i) = \pi_i(s_i|\omega = 1)$ is an NE for the 2-player 0/1 utility game, which is PPAD-hard to find. □

B. Find Local NE via Deep Learning

In this section, we describe the settings of our deep learning experiments and show more results.

For each problem instance, we collect a dataset comprising 50,000 randomly selected samples and train the networks for 30 epochs using the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 0.01. For extra-gradient, we initiate the optimization process from a set of 300 random starting points. For each starting point, we run 20 iterations of extra-gradient updates with the Adam optimizer and a learning rate of 0.1. We then use the result with the highest social welfare to compare the performance of different algorithms.

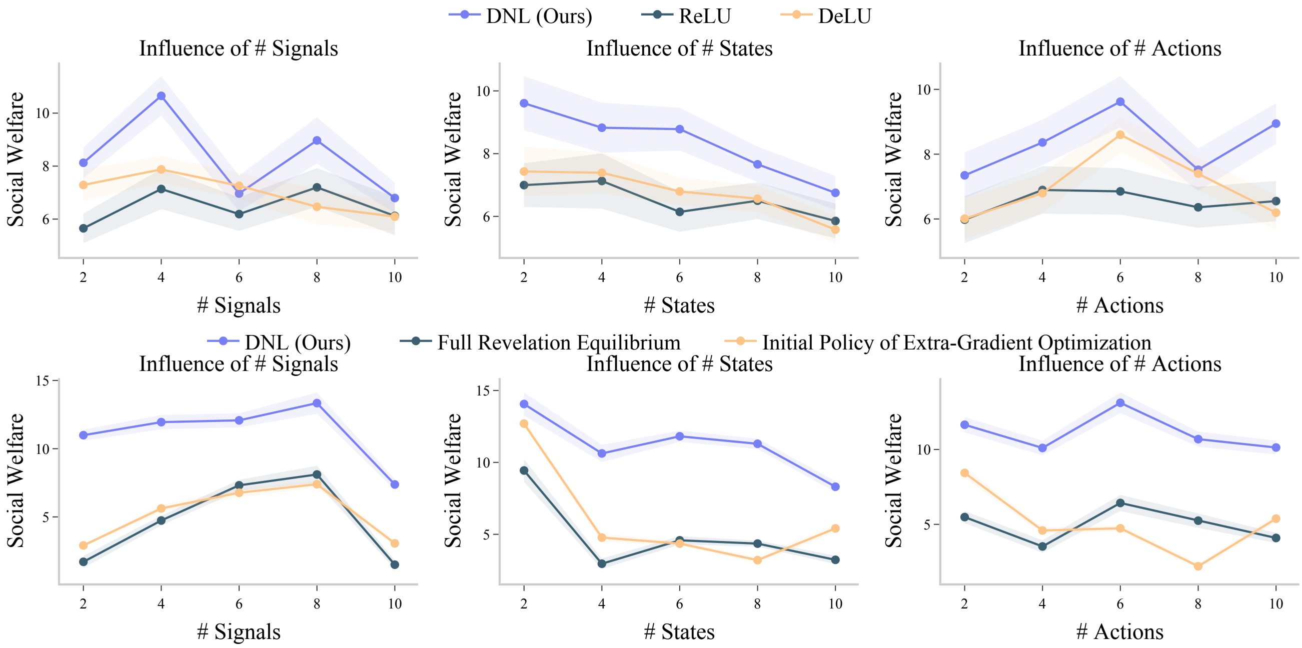

Figure 8. Our method retains its advantage and achieves higher social welfare compared against baselines and full-revelation solutions when we adopt a stricter standard for the local NE check procedure (increase ϵ from 0.005 to 0.01).

To evaluate if a joint signaling policy profile $\boldsymbol{\pi}$ derived from the extra-gradient algorithm constitutes a local NE, we randomly select K policies $\boldsymbol{\pi}'_j$ for each sender j within the vicinity $\{\pi'_j \mid \|\pi'_j - \pi_{\phi_j}\|_\infty \le \epsilon\}$. We then verify if any of these deviations result in increased utility. The number of test samples K grows linearly with the problem size:

$$K = \min\left\{10000, 1000 * (n-1)(|\Omega|-1)(|\mathcal{S}|-1)(|\mathcal{A}|-1)\right\}. \tag{19}$$

In the main text, we set the neighborhood size ϵ to 0.005. Now we apply a more stringent criterion for ϵ-local NE, with ϵ set to 0.01 and the extra-gradient optimization step increased to 30 accordingly. We reassess our method under this setting against baselines in Fig. 8. As we can observe, the performance of our method is still significantly better than other algorithms.

C. More Related Works

Our work focuses on a type of Stackelberg game. Since the signaling strategy of principals could be continuous, this game has a continuous action space. Stackelberg games are employed in various real-world hierarchical scenarios, including taxation (Zheng et al., 2022), security (Jiang et al., 2013; Gan et al., 2020), and business strategies (Naghizadeh & Liu, 2014; Zhang et al., 2016; Aussel et al., 2020). These games typically involve a leader and a follower. In such games with discrete choices, Conitzer & Sandholm (2006) demonstrate that linear programming can efficiently find Stackelberg equilibria using the strategy spaces of both players. For continuous decision spaces, Jin et al. (2020); Fiez et al. (2020) introduce and define local Stackelberg equilibria through first- and second-order conditions, with Jin et al. (2020) also showing that gradient descent-ascent methods can achieve these equilibria under certain conditions, and Fiez et al. (2020) providing specific updating rules that guarantee convergence.

With multiple followers (Zhang et al., 2024), unless they operate independently (Calvete & Galé, 2007), identifying Stackelberg equilibria is significantly harder and becomes NP-hard, even if followers have structured equilibria (Basilico et al., 2017). Wang et al. (2021a) suggest managing an arbitrary equilibrium that the follower may reach through differentiation. Meanwhile, Gerstgrasser & Parkes (2023) develop a meta-learning framework across various follower policies, facilitating quicker adaptations for the principal. This builds on Brero et al. (2022), who pioneered the Stackelberg-POMDP model.

The field of multi-agent reinforcement learning (Yu et al., 2022; Wen et al., 2022; Kuba et al., 2021; Christianos et al., 2020; Peng et al., 2021; Jiang et al., 2019; Wen et al., 2022; Rashid et al., 2018; Wang et al., 2020; 2021b; 2019b; Kang et al., 2020; Li et al., 2021; Wang et al., 2021d; Guestrin et al., 2002b;a; Böhmer et al., 2020; Kang et al., 2022; Wang et al., 2021c;

Yang et al., 2022; Dong et al., 2022; 2023; Wang et al., 2019a) is expanding the application of Stackelberg concepts to more complex, realistic settings. Tharakunnel & Bhattacharyya (2007) introduced a Leader-Follower Semi-Markov Decision Process for sequential learning in Stackelberg settings. Cheng et al. (2017) developed a method known as Stackelberg Q-learning, albeit without proving convergence. Furthermore, Shu & Tian (2019); Shi et al. (2019) have empirically examined these leader-follower dynamics, focusing on the leader's use of deep learning models to predict follower actions.