On the Convergence of First-Order Methods in Signaling Games

Rilind Sahitaj *
Department of Computer Science
Technical University of Munich
rilind.sahitaj@tum.de

Artem Tsikiridis *
Department of Computer Science
Technical University of Munich
artem.tsikiridis@tum.de

Martin Bichler
Department of Computer Science
Technical University of Munich
m.bichler@tum.de

Abstract

We study first-order learning dynamics in Bayesian signaling games, a minimal extensive-form model of strategic interaction under asymmetric information. In these games, a Sender observes a private type and chooses a signal, after which a Receiver forms a posterior belief and selects an action. This Bayesian belief update makes the induced learning field fundamentally different from the affine game-gradient fields of normal-form games. Empirically, we find a sharp dichotomy: standard online learning algorithms reliably converge to strict Perfect Bayesian Equilibria (PE), but not to non-strict ones. We show that this behavior cannot be explained by the global geometric conditions commonly used to prove last-iterate convergence, such as monotonicity or the Minty variational inequality. Indeed, these conditions can fail even in binary signaling games. Instead, we identify the relevant local geometry. Our main result characterizes variational stability in finite signaling games: a PE is variationally stable if and only if it is strict. This yields local last-iterate convergence guarantees for online mirror descent and regularized dual averaging near every strict PE. We then study global average-iterate convergence and identify a class of signaling games in which coarse-trigger regret-minimizing dynamics converge to the unique PE. Finally, we complement the theory with experiments comparing projected gradient ascent, mirror-based methods, optimistic variants, and PPO across canonical signaling-game families. The results position signaling games as a compact testbed for understanding how Bayesian belief formation reshapes the convergence geometry of multi-agent learning.

1 Introduction

A central objective in learning in games is to understand when online learning algorithms converge to equilibrium. In normal-form games, many of the strongest convergence guarantees are obtained by imposing geometric conditions on the game-gradient field, such as monotonicity [10], the Minty variational inequality [27], and variational stability [26]. These conditions connect local payoff improvements to global or local convergence guarantees for first-order learning dynamics, including online mirror descent [43], regularized dual averaging [26], extragradient methods [28], and optimistic variants [9]. A key reason why this theory is tractable in finite normal-form games is that expected

*Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

payoffs are multilinear in mixed strategies, so the induced game-gradient field is affine on the mixed extension of the game. Much less is known about whether this geometric perspective extends to extensive-form games with asymmetric information, where beliefs and strategies interact through Bayesian updating.

We study this question in finite Bayesian signaling games, the canonical extensive-form model of strategic interaction under asymmetric information [8, 34, 36]. A Sender observes a private type and chooses a signal; a Receiver observes the signal, forms a posterior belief over types, and then chooses an action. Signaling games are a minimal setting in which first-order learning must interact with endogenous belief formation. The analogue of the game-gradient field is the *payoff-difference field*, which records the directions of local payoff improvement across signals and actions. Unlike in normal-form games, this field depends on the Receiver's posterior belief, which is determined by Bayes' rule from the Sender's strategy. Consequently, the payoff-difference field is not multilinear, is generally not affine, and is only canonically defined on the region where every signal is reached with positive probability. We call this region the Bayes-defined region.[2]

This change in geometry leads to a convergence puzzle. In our experiments, standard online learning algorithms converge reliably to strict perfect Bayesian equilibria (PE), but fail to converge to non-strict ones (Section 5). Strict PE are pure-strategy equilibria in which each Sender type chooses a unique signal and the Receiver has a unique best response at every signal. They include important separating and pooling equilibria, such as the Riley outcome in Spence's job-market signaling model [32, 36]. However, this observed convergence cannot be explained by the global geometric conditions that are standard tools for analyzing first-order learning in normal-form games, as these do not extend directly to signaling games. Indeed, Theorem 1 shows that even with two types, two signals, and two actions, there are signaling games and strict PE for which neither monotonicity nor the Minty variational inequality holds on the Bayes-defined region. Understanding why learning nevertheless converges near strict PE is the motivating question of this paper.

1.1 Our Results

Our first result shows that the standard global geometry used to prove last-iterate convergence in normal-form games does not extend to signaling games: even in binary signaling games, monotonicity and the Minty variational inequality can fail at a strict PE (Theorem 1). We therefore turn to local geometry. Our main result shows that, in finite signaling games, a separating PE is variationally stable if and only if it is strict (Theorem 2). This parallels the characterization of variationally stable strategy profiles as strict Nash equilibria in finite normal-form games by Mertikopoulos and Zhou [26, Proposition 5.2], but the extension is not formal: in signaling games the payoff-difference field depends on Bayes-updated posterior beliefs, is not multilinear, and is only canonically defined on the Bayes-defined region. Thus, our result shows that strictness remains exactly the right local geometric condition for first-order learning even in this Bayesian extensive-form setting. The proof is quantitative in the strict-PE-to-stability direction: we derive an explicit radius of the stability neighborhood in terms of primitive game parameters and the payoff-difference field (Theorem 3). Combining this characterization with the convergence result of Mertikopoulos and Zhou [26, Theorem 4.10] yields local last-iterate convergence of regularized dual averaging and online mirror descent [43] to every strict separating PE from sufficiently close initializations.

In Section 4, we move beyond local convergence guarantees. We identify a class of signaling games (which includes the example from Theorem 1 that fails the Minty condition) for which, as we show in Theorem 5: the unique PE is also the unique extensive-form coarse correlated equilibrium (EFCCE) [11]. Since the broader EFCCE polytope admits average-iterate convergence guarantees under no-coarse-trigger-regret dynamics [2], and since under our assumptions this polytope collapses to the unique PE, we obtain a *global* average-iterate convergence guarantee to the PE.

Finally, our experiments on canonical binary signaling games compare projected gradient ascent, online mirror ascent, optimistic mirror ascent, and PPO. The results support the theory: strict pure separating PE are reliable attractors of standard first-order learning dynamics, whereas convergence to non-strict equilibria is fragile and algorithm-dependent. Interestingly, convergence is also observed in families with strict pooling or multiple pure PE, both of which lie outside the scope of Theorem 2.

[2] Outside this set, the Receiver's beliefs are not pinned down by Bayes' rule; which beliefs should be considered rational is a central issue in the signaling-games literature [6, 34].

We also observe that optimistic (entropic) mirror descent can converge to a mixed equilibrium. We emphasize that the convergence for projected gradient ascent and PPO is only observed empirically.

Table 3 in Appendix A provides a summary of the results.

1.2 Related Work

Our paper connects three literatures: learning in extensive-form games, convergence in monotone/variational-inequality game settings and signaling games as a minimal model of strategic interaction under asymmetric information. We cover the most relevant work on learning in extensive-form games in this section and the essential references on variational stability in Section 3.1. We discuss further related work in Appendix B.

The dominant algorithmic approach to learning in extensive-form games is based on regret minimization. In two-player zero-sum games, variants of counterfactual regret minimization (CFR) exploit the recursive structure of the game tree and yield average-iterate convergence to Nash equilibrium [44]. In general-sum extensive-form games, however, uncoupled no-regret guarantees typically lead not to Nash equilibrium but to extensive-form correlated solution concepts. In particular, trigger-regret minimization yields convergence of empirical play to extensive-form correlated equilibrium (EFCE), while coarse-trigger-regret minimization over sequence-form strategy spaces yields extensive-form coarse correlated equilibrium (EFCCE) [11–13, 40]. Finally, to our knowledge, last-iterate convergence to equilibrium in extensive-form games has been studied only in the zero-sum case [22, 25, 31].

2 Preliminaries

Finite signaling games A signaling game consists of a Sender and a Receiver. Nature first draws the Sender's type $\theta \in \Theta$ from a commonly known prior $\pi \in \Delta(\Theta)$[3] with $\pi(\theta) > 0$ for all $\theta \in \Theta$. The Sender observes θ and chooses a signal $s \in \Sigma$. The Receiver observes s, and then chooses an action $a \in A$. Utilities are given by functions $u_S, u_R : \Theta \times \Sigma \times A \to \mathbb{R}$. We focus on the finite case, where Θ, Σ, and A are finite sets and we assume $|\Theta| = |\Sigma| = |A| = K$ for a $K \in \mathbb{N}_{\geq 2}$. Explicitly, we denote the sets as $\Theta = \{\theta_1, \dots, \theta_K\}$, $\Sigma = \{s_1, \dots, s_K\}$, $A = \{a_1, \dots a_K\}$. We refer to such a signaling game by the tuple $\mathcal{G} = (\Theta, \Sigma, A, \pi, u_S, u_R)$.

A behavioral strategy of the Sender is a conditional probability distribution $q_S(\cdot \mid \theta) \in \Delta(\Sigma)$ for each $\theta \in \Theta$, and a behavioral strategy of the Receiver is a conditional distribution $q_R(\cdot \mid s) \in \Delta(A)$ for each $s \in \Sigma$. A *strategy profile* is a pair $q = (q_S, q_R)$. Since $\mathcal{G}$ is finite, its strategy space is $X(\mathcal{G}) := \prod_{\theta \in \Theta} \Delta(\Sigma) \times \prod_{s \in \Sigma} \Delta(A)$.

Given a Sender strategy q_S, signal s is said to be *on-path* under q_S if it has positive marginal probability under the prior, $\sum_{\theta \in \Theta} \pi(\theta)\, q_S(s \mid \theta) > 0$. Whenever s is on-path, the Receiver's posterior is determined by Bayes' rule:

$$\mu(\theta \mid s) \;=\; \frac{\pi(\theta)\, q_S(s \mid \theta)}{\sum_{\theta' \in \Theta} \pi(\theta')\, q_S(s \mid \theta')}. \tag{1}$$

Strategy profiles in $X(\mathcal{G})$ for which every signal is on-path under q_S form the *Bayes-defined region*:

$$X^\circ(\mathcal{G}) \;:=\; \Big\{ q \in X(\mathcal{G}) \;:\; \sum_{\theta \in \Theta} \pi(\theta)\, q_S(s \mid \theta) > 0 \text{ for all } s \in \Sigma \Big\}. \tag{2}$$

Lemma 1. *For every signaling game $\mathcal{G}$, $X^\circ(\mathcal{G})$ is convex and open in $X(\mathcal{G})$.*

Lemma 1 will be useful in the analysis that follows; its proof is deferred to Appendix C. Finally, we write $\mathrm{supp}(q_S(\cdot \mid \theta)) := \{s \in \Sigma : q_S(s \mid \theta) > 0\}$ and $\mathrm{supp}(q_R(\cdot \mid s)) := \{a \in A : q_R(a \mid s) > 0\}$.

Solution concepts A *belief system* for $\mathcal{G}$ is a map $\mu : \Sigma \to \Delta(\Theta)$ assigning to every signal $s \in \Sigma$ a posterior $\mu(\cdot \mid s) \in \Delta(\Theta)$ over Sender types. An *assessment* of $\mathcal{G}$ is a pair (q, μ) consisting of a strategy profile $q = (q_S, q_R) \in X(\mathcal{G})$ and a belief system μ for $\mathcal{G}$.

Definition 1. *An assessment (q, μ) of $\mathcal{G}$ is a* Perfect Bayesian Equilibrium *(*PE*) of $\mathcal{G}$ if:*

[3] For a finite set X, $\Delta(X)$ denotes the simplex of probability distributions over X.

(i) For every $\theta \in \Theta$ and every $s \in \Sigma$,

$$\operatorname{supp}\big(q_S(\cdot \mid \theta)\big) \subseteq \arg\max_{s' \in \Sigma} \sum_{a \in A} u_S(\theta, s', a)\, q_R(a \mid s'),$$

$$\operatorname{supp}\big(q_R(\cdot \mid s)\big) \subseteq \arg\max_{a \in A} \sum_{\theta' \in \Theta} u_R(\theta', s, a)\, \mu(\theta' \mid s).$$

(ii) At every signal $s \in \Sigma$ that is on-path under q_S, the belief $\mu(\cdot \mid s)$ is given by Bayes' rule (1)*. For the remaining signals the belief $\mu(\cdot \mid s) \in \Delta(\Theta)$ is unrestricted.*

Condition (i) is known as *sequential rationality* and condition (ii) as *consistency*. We use μ for both the Bayes posterior in (1) and the belief system of an assessment: on $X^\circ(\mathcal{G})$ these coincide, since every signal is on-path, so the belief system μ is uniquely determined by Bayes' rule. Accordingly, we identify a PE assessment (q, μ) on X° with its strategy profile q, and refer to $q^\star \in X^\circ(\mathcal{G})$ simply as a PE. In the signaling game literature, a PE is further classified as separating, pooling, or semi-separating according to whether the Sender's signal reveals her type.

Definition 2. *A* PE (q, μ) *of $\mathcal{G}$ is:*

(i) Separating *if for every $\theta \in \Theta$ there exists a signal $s \in \Sigma$ such that* $\operatorname{supp}\big(q_S(\cdot \mid \theta)\big) = \{s\}$ *and $s \notin \operatorname{supp}\big(q_S(\cdot \mid \theta')\big)$ for every $\theta' \in \Theta \setminus \{\theta\}$.*

(ii) Pooling *if there exists a signal $s \in \Sigma$ such that* $\operatorname{supp}(q_S(\cdot \mid \theta)) = \{s\}$ *for every $\theta \in \Theta$.*

(iii) Semi-separating *if it is neither separating nor pooling.*

In a separating PE, each type plays a pure signal that no other type sends, so observing the signal uniquely identifies the Sender's type. In a pooling PE, all types send the same signal, so observing it conveys no information about the type.

Up to relabeling, the *Riley* PE is the assessment $(q^\star, \mu)$ in which type θ_k plays signal s_k and the Receiver responds with a_k for all $k \in \{1, \dots, K\}$ i.e.,

$$q_S^\star(s_k \mid \theta) = \mathbf{1}\{\theta = \theta_k\}, \qquad q_R^\star(a \mid s_k) = \mathbf{1}\{a = a_k\}, \qquad \mu^\star(\theta \mid s_k) = \mathbf{1}\{\theta = \theta_k\}. \tag{3}$$

We define the set of PE strategy profiles of $\mathcal{G}$ as

$$\mathcal{S}^{\mathrm{PE}} \;=\; \big\{q \in X(\mathcal{G}) \;:\; \exists \mu : \Sigma \to \Delta(\Theta) \text{ such that } (q, \mu) \text{ is a PE of } \mathcal{G}\big\}.$$

We say $\mathcal{G}$ has a *unique* PE if $\mathcal{S}^{\mathrm{PE}}$ is a singleton. We call the strategy profile $q^\star$ of the Riley PE the *Riley outcome*[32]. A PE $q^\star$ is *pure* if $|\operatorname{supp} q_S^\star(\cdot \mid \theta)| = |\operatorname{supp} q_R^\star(\cdot \mid s)| = 1$ for every $\theta \in \Theta$ and $s \in \Sigma$; write $\sigma^\star(\theta)$ and $\alpha^\star(s)$ for the unique elements of these supports. A pure PE is *strict* if $u_S(\theta, \sigma^\star(\theta), \alpha^\star(\sigma^\star(\theta))) > u_S(\theta, s, \alpha^\star(s))$ for every $\theta \in \Theta$ and $s \neq \sigma^\star(\theta)$, and $\sum_\theta \mu^\star(\theta \mid s)\, u_R(\theta, s, \alpha^\star(s)) > \sum_\theta \mu^\star(\theta \mid s)\, u_R(\theta, s, a)$ for every $s \in \Sigma$ and $a \neq \alpha^\star(s)$.

We next define the Bayesian extensive-form coarse correlated equilibrium (EFCCE), introduced by Farina et al. [11]. Given $\mathcal{G}$, a *recommendation device* is a probability distribution $Q \in \Delta(\Sigma^\Theta \times A^\Sigma)$ over pairs of pure strategies (σ, α), with $\sigma : \Theta \to \Sigma$ and $\alpha : \Sigma \to A$. In an EFCCE, a pair $(\sigma, \alpha) \sim Q$ is drawn, and each player commits to follow the recommendation when it is their turn, without observing it.

Definition 3 (11)**.** *A probability distribution $Q \in \Delta(\Sigma^\Theta \times A^\Sigma)$ is a Bayesian extensive-form coarse correlated equilibrium (*EFCCE*) if:*

1. *For every type $\theta \in \Theta$ and every signal $s' \in \Sigma$:*

$$\mathbb{E}_{(\sigma,\alpha)\sim Q}\big[u_S\big(\theta,\, \sigma(\theta),\, \alpha(\sigma(\theta))\big)\big] \;\geq\; \mathbb{E}_{(\sigma,\alpha)\sim Q}\big[u_S\big(\theta,\, s',\, \alpha(s')\big)\big]. \tag{4}$$

2. *For every signal $s \in \Sigma$ with $\sum_{\theta\in\Theta} \pi(\theta) \Pr_{(\sigma,\alpha)\sim Q}[\sigma(\theta) = s] > 0$ and every action $a' \in A$:*

$$\sum_{\theta \in \Theta} \mu(\theta \mid s)\, \mathbb{E}_{(\sigma,\alpha)\sim Q}\big[u_R(\theta,\, s,\, \alpha(s)) \mid \sigma(\theta) = s\big] \;\geq\; \sum_{\theta \in \Theta} \mu(\theta \mid s)\, u_R(\theta, s, a'), \tag{5}$$

with $\mu(\theta|s)$ defined according to (1)[4].

[4]For a recommendation device Q, the marginal $q_S(s \mid \theta)$ in (1) is replaced by the projection $\Pr_{(\sigma,\alpha)\sim Q}[\sigma(\theta) = s]$.

The conditions in Definition 3 require that no Sender profitably deviates to a fixed $s' \in \Sigma$ (4), and that no Receiver, conditional on the recommended signal being $s \in \Sigma$, profitably deviates to a fixed action $a' \in A$ (5).

Every PE strategy $q^\star$ also embeds into the EFCCE polytope as the product recommendation device with $\sigma(\theta) \sim q_S^\star(\cdot \mid \theta)$ and $\alpha(s) \sim q_R^\star(\cdot \mid s)$, with all draws independent. Then, the EFCCE conditions (4)-(5) reduce to the on-path[5] PE sequential rationality conditions in Definition 1. Therefore, $\mathcal{S}^{\mathrm{PE}} \subseteq \mathcal{S}^{\mathrm{EFCCE}}$.

Payoff-difference field To study first-order learning dynamics via geometric conditions, we work with the induced *payoff-difference field*.

Definition 4. *Given a* reference profile $(\bar\sigma, \bar\alpha)$ *with* $\bar\sigma : \Theta \to \Sigma$ *and* $\bar\alpha : \Sigma \to A$*, the* payoff-difference field *of* $\mathcal{G}$ *relative to* $(\bar\sigma, \bar\alpha)$ *is the map* $g : X^\circ(\mathcal{G}) \to \mathbb{R}^{\Theta\times\Sigma} \times \mathbb{R}^{\Sigma\times A}$ *with components, for* $\theta \in \Theta$*,* $s \in \Sigma$*,* $a \in A$*,*

$$g_S(\theta, s;\, q) \;:=\; \sum_{a' \in A} u_S(\theta, s, a')\, q_R(a' \mid s) \;-\; \sum_{a' \in A} u_S(\theta, \bar\sigma(\theta), a')\, q_R\big(a' \mid \bar\sigma(\theta)\big), \tag{6}$$

$$g_R(s, a;\, q) \;:=\; \sum_{\theta \in \Theta} \big[u_R(\theta, s, a) - u_R\big(\theta, s, \bar\alpha(s)\big)\big]\, \mu(\theta \mid s). \tag{7}$$

In words, $g_S(\theta, s; q)$ is the Sender's expected payoff at type θ from signal s minus that from the reference signal $\bar\sigma(\theta)$, and $g_R(s, a; q)$ is the Receiver's posterior expected payoff at signal s from action a minus that from the reference action $\bar\alpha(s)$. Subtracting the reference payoff shifts all coordinates of $g_S(\theta, \cdot\,; q)$ (resp. $g_R(s, \cdot\,; q)$) by the same constant, which changes neither best responses nor any of the variational conditions used in Section 3 (see Proposition 1 in Appendix D.6). Up to this shift and a positive rescaling (by $\pi(\theta)$ for the Sender, and by the probability $\sum_{\theta\in\Theta} \pi(\theta)\, q_S(s \mid \theta)$ that s is reached for the Receiver), g is the gradient of each player's expected utility, i.e., the input to the first-order dynamics we study.

The key difference from the normal-form setting (see e.g., Mertikopoulos and Zhou [26]) is that g depends on posterior beliefs induced by Bayes' rule. As a result, g need not be multilinear, need not be globally defined on all of $X(\mathcal{G})$, and can have geometric properties that differ from those in normal-form games. Nevertheless, as we show in Lemma 2 (the proof is in Appendix C), g retains continuity on $X^\circ(\mathcal{G})$. This is the basic regularity needed for our analysis in Section 3.

Lemma 2. *The payoff-difference field* g *is continuous on* $X^\circ(\mathcal{G})$.

3 Last-iterate Convergence

In this section, we study last-iterate convergence of first-order learning dynamics in signaling games. A classical perspective, which we also adopt, is to establish geometric conditions on g (Definition 4). All missing proofs of this section can be found in Appendix D.

3.1 Limitations to Global Last-iterate Convergence

In normal-form games, the standard route to global last-iterate convergence of first-order learning dynamics is to verify a global geometric condition on the payoff-difference field. In Definition 5, we highlight the two such conditions central to the literature; throughout, $\langle \cdot, \cdot \rangle$ denotes the standard Euclidean inner product on $\mathbb{R}^{\Theta\times\Sigma} \times \mathbb{R}^{\Sigma\times A}$.

Definition 5. *Let* $\mathcal{G}$ *be a signaling game and let* g *be its payoff-difference field.*

(i) g *is* monotone *if* $\langle g(q) - g(q'),\, q - q' \rangle \le 0$ *for all* $q, q' \in X^\circ(\mathcal{G})$.

(ii) $q^\star \in X^\circ(\mathcal{G})$ *satisfies the* Minty condition *if* $\langle g(q),\, q - q^\star \rangle \le 0$ *for every* $q \in X^\circ(\mathcal{G})$.

Standard theory on variational inequalities (see e.g., [10, Chapter 2]) relates the two: if g is monotone and $q^\star$ is a PE of $\mathcal{G}$, then $q^\star$ satisfies the Minty condition; conversely, the Minty condition at $q^\star$ implies that $q^\star$ is a PE of $\mathcal{G}$.

[5] Condition (5) only constrains on-path signals; off-path receiver behavior is unrestricted, in contrast to perfect correlated solution concepts [24].

In normal-form games, monotonicity yields last-iterate convergence of dual averaging and mirror descent [4, 19, 26], while the Minty condition yields last-iterate convergence of extragradient and optimistic methods [9, 28]. Somewhat surprisingly, we show in Theorem 1 that for finite signaling games, even for the simplest case of $|\Theta| = |\Sigma| = |A| = 2$ there are instances where these conditions fail.

Theorem 1. *There exists a signaling game $\mathcal{G}$ with $|\Theta| = |\Sigma| = |A| = 2$ and a strict separating* PE *$q^\star \in X^\circ(\mathcal{G})$ at which the Minty condition fails.*

The instance we use in Theorem 1 is the following: the binary signaling game with prior $\pi(\theta_1) = \pi(\theta_2) = 1/2$ and utilities $u_R(\theta_k, s, a_j) = \mathbf{1}\{j = k\}$ and $u_S(\theta_k, s_i, a_j) = \frac{3}{2}\,\mathbf{1}\{i = k\} + \mathbf{1}\{j = k\}$. It is a remarkably simple instance: the Riley outcome $q^\star$ in (3) is its unique strict PE strategy and the Sender has a dominant strategy to signal her type. Yet, the Minty condition (Definition 5(ii)) fails on $q^\star$. Since monotonicity of g on $X^\circ(\mathcal{G})$ (Definition 5(i)) implies Minty at every PE, Theorem 1 also rules out monotonicity of g on X°.

3.2 Variational Stability

Theorem 1 rules out the global geometric conditions typically used to obtain last-iterate convergence. Therefore, we turn to a *local* condition called *variational stability* [26].

Definition 6 (26)**.** *A strategy profile $q^\star \in X^\circ(\mathcal{G})$ is variationally stable if there exists an open neighborhood $H \subseteq X^\circ(\mathcal{G})$ of $q^\star$ such that $\langle g(q),\, q - q^\star\rangle \le 0$ for every $q \in H$, with strict inequality for $q \in H \setminus \{q^\star\}$.*

For finite signaling games, the notion is well-posed by Lemma 1: openness of $X^\circ(\mathcal{G})$ guarantees that every $q^\star \in X^\circ(\mathcal{G})$ has a non-trivial open neighborhood in $X^\circ(\mathcal{G})$.

We stress that variational stability and the Minty condition of Definition 5(ii) are not directly comparable: the former imposes a *strict* inequality on a neighborhood of $q^\star$, while the latter imposes a non-strict inequality on $X^\circ(\mathcal{G})$. The two conditions also yield different algorithmic guarantees. Under monotonicity or Minty, mirror descent and dual averaging can cycle in last iterate (as in zero-sum bilinear games [9, 27]), while extragradient and optimistic-gradient methods recover last-iterate convergence in these settings [9, 28]. By contrast, under variational stability, FTRL dynamics [26] (in particular, dual averaging with a strongly convex regularizer) and online mirror descent [43] attain last-iterate convergence to $q^\star$ from initializations sufficiently close to $q^\star$; see [26, Section 4] for a detailed discussion.

For finite normal-form games, Mertikopoulos and Zhou [26, Proposition 5.2] show that variationally stable strategy profiles coincide with strict Nash equilibria. In our main result, Theorem 2, we extend this characterization to strict PE for finite signaling games.

Theorem 2. *Let $\mathcal{G}$ be a signaling game and let $q^\star \in X^\circ(\mathcal{G})$. Then, $q^\star$ is variationally stable if and only if it is a strict* PE*.*

We devote the rest of this section to proving Theorem 2, showing first that every strict PE $q^\star \in X^\circ(\mathcal{G})$ is variationally stable.

Fix a strict PE $q^\star \in X^\circ(\mathcal{G})$. Since $|\operatorname{supp} q_S^\star(\cdot \mid \theta)| = |\operatorname{supp} q_R^\star(\cdot \mid s)| = 1$ for every $\theta \in \Theta$ and $s \in \Sigma$ (by definition of strict PE), we write $\sigma^\star(\theta)$ for the unique signal in $\operatorname{supp} q_S^\star(\cdot \mid \theta)$ and $\alpha^\star(s)$ for the unique action in $\operatorname{supp} q_R^\star(\cdot \mid s)$. Equivalently, this defines maps $\sigma^\star\colon \Theta \to \Sigma$ and $\alpha^\star\colon \Sigma \to A$ that capture $q^\star$. Throughout this section, we set $\bar\sigma = \sigma^\star$ and $\bar\alpha = \alpha^\star$.

We now introduce auxiliary notation. Let

$$\delta(\mathcal{G}, q^\star) := \min\Big\{ \min_{\theta\in\Theta,\, s\neq\sigma^\star(\theta)} |g_S(\theta, s; q^\star)|,\ \min_{s\in\Sigma,\, a\neq\alpha^\star(s)} |g_R(s, a; q^\star)| \Big\},$$

$\bar u_S(\mathcal{G}, q^\star) := \max_{\theta,s,a} |u_S(\theta, s, a) - u_S(\theta, s, \alpha^\star(s))|$, and $\bar u_R(\mathcal{G}, q^\star) := \max_{\theta,s,a} |u_R(\theta, s, a) - u_R(\theta, s, \alpha^\star(s))|$. It is not hard to observe that $\delta(\mathcal{G}, q^\star)$ and $\bar u_R(\mathcal{G}, q^\star)$ are strictly positive for a strict PE $q^\star$.

Lemma 3. *For a strict* PE *$q^\star \in X^\circ(\mathcal{G})$, it holds that $\delta(\mathcal{G}, q^\star) > 0$ and $\bar u_R(\mathcal{G}, q^\star) > 0$.*

With Lemma 3 in hand, we obtain a quantitative characterization of the variational stability neighborhood of $q^\star$, with an explicit radius determined by $\delta(\mathcal{G}, q^\star)$, $\bar u_S(\mathcal{G}, q^\star)$, $\bar u_R(\mathcal{G}, q^\star)$, and $\min_{\theta\in\Theta} \pi(\theta)$.

Theorem 3. *Let $q^\star \in X^\circ(\mathcal{G})$ be a strict* PE *of $\mathcal{G}$. Let $L := \max\left\{\bar{u}_S(\mathcal{G}, q^\star),\ \frac{2\bar{u}_R(\mathcal{G},q^\star)}{\min_{\theta\in\Theta}\pi(\theta)}\right\} > 0$. Then, for all $q \in X^\circ(\mathcal{G})$ with $\|q - q^\star\|_1 < \delta(\mathcal{G}, q^\star)/(2L)$, $\langle g(q),\, q - q^\star\rangle \le 0$ holds, with strict inequality whenever $q \neq q^\star$.*

The quantity $\delta(\mathcal{G}, q^\star)/(2L)$ in Theorem 3 has a transparent interpretation: δ measures how strict the PE $q^\star$ is, while L captures how rapidly g varies near $q^\star$: it is their ratio that sets the radius of the variational stability neighborhood. The proof of Theorem 3 relies on two technical lemmas.

Lemma 4. *Let $q^\star \in X^\circ(\mathcal{G})$ be a strict* PE *of $\mathcal{G}$. Then for every $q \in X^\circ(\mathcal{G})$, $\langle g(q^\star),\, q - q^\star\rangle \le -\frac{\delta(\mathcal{G},q^\star)}{2}\,\|q - q^\star\|_1$.*

Lemma 5. *Let $q^\star \in X^\circ(\mathcal{G})$ be a strict* PE *of $\mathcal{G}$. Then, for every $q \in X^\circ(\mathcal{G})$, $\langle g(q) - g(q^\star),\, q - q^\star\rangle \le L \cdot \|q - q^\star\|_1^2$, where $L = \max\left\{\bar{u}_S(\mathcal{G}, q^\star),\ \frac{2\,\bar{u}_R(\mathcal{G},q^\star)}{\min_{\theta\in\Theta}\pi(\theta)}\right\}$.*

We are now ready to prove Theorem 3.

Proof of Theorem 3. By Lemma 3, we have that $\delta(\mathcal{G}, q^\star) > 0$ and $\bar{u}_R(\mathcal{G}, q^\star) > 0$, so $L > 0$ as claimed and the radius $\delta(\mathcal{G}, q^\star)/(2L)$ is well-defined and positive. The case $q = q^\star$ is immediate and we focus on the $q \neq q^\star$ case to complete the proof.

Fix $q \in X^\circ(\mathcal{G})$ with $0 < \|q - q^\star\|_1 < \delta(\mathcal{G}, q^\star)/(2L)$. Then

$$\begin{aligned}
\langle g(q),\, q - q^\star\rangle \;&=\; \langle g(q^\star),\, q - q^\star\rangle \;+\; \langle g(q) - g(q^\star),\, q - q^\star\rangle \\
&\le\; -\frac{\delta(\mathcal{G}, q^\star)}{2}\,\|q - q^\star\|_1 \;+\; L\,\|q - q^\star\|_1^2 \\
&=\; \|q - q^\star\|_1 \left(-\frac{\delta(\mathcal{G}, q^\star)}{2} \;+\; L\,\|q - q^\star\|_1\right) \\
&<\; \|q - q^\star\|_1 \left(-\frac{\delta(\mathcal{G}, q^\star)}{2} \;+\; L \cdot \frac{\delta(\mathcal{G}, q^\star)}{2L}\right) \;=\; 0,
\end{aligned}$$

where the first inequality follows by Lemma 4 and Lemma 5. Finally, the strict inequality holds since, by assumption, $\|q - q^\star\|_1 < \delta(\mathcal{G}, q^\star)/(2L)$. □

We conclude this section by showing that variational stability of a point $q^\star \in X^\circ(\mathcal{G})$ implies that $q^\star$ is a strict PE. This establishes our characterization result and completes the proof of Theorem 2.

Theorem 4. *Let $q^\star \in X^\circ(\mathcal{G})$. If $q^\star$ is variationally stable, then it is a strict* PE*.*

4 Global Average-iterate Convergence

Recall that in Theorem 1 (Section 3.1) we present a simple finite signaling game with a strict PE at which the Minty condition fails (and hence monotonicity also fails; see Definition 5). A feature of this instance is that the sender has a dominant strategy to signal her type, while the receiver has a unique best response upon receiving an "honest" signal. In this section we abstract these features into two assumptions on the signaling game and show that, under them, the unique PE (the Riley assessment) is in fact the unique EFCCE. As we discuss below (Corollary 1), this yields a *global* average-iterate convergence guarantee to the unique PE for this class of signaling games, which is determined by Assumptions 1 and 2 below. We call signaling games satisfying both assumptions *honest-signaling games*.

Assumption 1. *For every $k \in \{1, \dots, K\}$ and every $j \neq k$, $u_R(\theta_k, s_k, a_k) > u_R(\theta_k, s_k, a_j)$.*

Assumption 2. *For every $k \in \{1, \dots, K\}$ and every $j \neq k$, $\min_{a\in A}\, u_S(\theta_k, s_k, a) > \max_{a\in A}\, u_S(\theta_k, s_j, a)$.*

Assumption 1 says that when the receiver believes the sender's type is θ_k, action a_k is the unique best response at the truthful signal s_k. This unique-best-response condition is standard in the signaling-games literature[6] [6, 23, 34]. Assumption 2 says that type θ_k strictly prefers the truthful signal s_k

[6] In continuous-action games, an analogous role is played by the assumption that the receiver's pointwise utility is strictly quasi-concave in the action, see e.g., [34].

over any non-truthful signal s_j, regardless of the receiver's action. The main result of this section is Theorem 5. Its proof can be found in Appendix E.1.

Theorem 5. *For every honest-signaling game, the Riley assessment* $(q^\star, \mu^\star)$ *is the unique* PE. *Moreover,* $q^\star$ *is the unique* EFCCE.

Trigger regret captures the cumulative payoff gain a player could obtain by "triggering" a deviation that may depend on the recommendation; coarse-trigger regret restricts the deviation to be fixed before the recommendation is seen. Algorithms that minimize trigger regret converge in average iterate to the EFCE[7] polytope [13], while algorithms that minimize coarse-trigger regret converge to the (broader) EFCCE polytope [2]. Since Theorem 5 shows that the EFCCE polytope is a singleton for this class of games, we obtain the following corollary.

Corollary 1. *Let* $\mathcal{G}$ *be an honest-signaling game. If both the Sender and the Receiver run algorithms which minimize coarse-trigger regret, the empirical distribution of play converges in average iterate to the Riley assessment* $(q^\star, \mu^\star)$.

Concrete examples include the trigger-regret-minimizing dynamics of Farina et al. [13] and the coarse-trigger-regret-minimizing algorithms of Anagnostides et al. [2], Farina et al. [12].

5 Experiments

Our experiments study whether simple first-order learning dynamics converge to designated equilibrium targets in small signaling games. The experiments are implemented in the behavioral strategy space of binary signaling games and are designed to complement the theoretical analysis. We focus on $K = 2$ and extend two benchmarks to $K = 4$ in Section 5.3. This case is already rich enough to yield multiple equilibrium classes and multiple equilibria. We focus on the equilibrium selection problem, which is predominantly concerned with resolving multiplicity rather than the problem's dimensionality. This setting provides a suitable baseline for the analysis.

5.1 Experimental setup

We study learning dynamics in benchmark $2\times2\times2$ signaling games classified *endogenously* by their equilibrium structure. The seven families span the qualitatively distinct cases that arise in two-type signaling games. Three families admit a unique pure-strategy PE: a *unique separating* family in which each type sends a distinct signal that perfectly reveals private information; a *unique pooling at the low signal* and *unique pooling at the high signal*. Two further families admit multiple pure PE: a Spence-style *pooling-and-separating* family modeled on the job-market signaling problem of Spence [36], in which a pooling equilibrium and a separating equilibrium coexist without belief refinements [e.g. the Intuitive Criterion of 6]; and a *multiple separating*, in which two strict pure separating profiles – the Riley outcome and its permutation coexist as PE. A *multiple-pooling* family admits several pure pooling profiles but no separating equilibrium. Finally, a *unique semi-separating* (see Definition 2) family modeled on the reputation game of Kreps and Wilson [20] has a strictly mixed unique PE in which one sender type plays a pure strategy and the other mixes – the only family in our set whose equilibrium is not pure [8]. For more details on the games, we refer to Appendix F.2. Table 4 lists the hyperparameters per family. Across families we compare four learning algorithms – projected gradient ascent (PGA) [41], entropic online mirror ascent (OMA), optimistic entropic online mirror ascent (Opt. OMA) [38], and proximal policy optimisation (PPO) [33] on a tabular policy/value parameterization – and consider both local and global initializations. For more details on the setting and treatment variables, we refer to Appendix F.1

The simulation design has two stages. First, we generate a candidate signaling game from a parameterized family intended to represent the corresponding equilibrium class. We retain only parameter configurations whose realized equilibrium class matches the intended benchmark family. Second, for each retained game, we simulate learning from randomized initial conditions under the selected algorithm and hyperparameters.

[7] In a (Bayesian) extensive-form correlated equilibrium (EFCE) [40], the player observes the recommendation before deciding whether to follow it; the deviation may then depend on the recommendation. For this reason, it holds that $\mathcal{S}^{\mathrm{EFCE}} \subseteq \mathcal{S}^{\mathrm{EFCCE}}$.

[8] Other families may admit mixed-strategy equilibria as well; what singles out the semi-separating family is the absence of any pure-strategy equilibrium.

Table 1: Aggregate results ($T = 2000$) for one benchmark game per family (Table 6), with preferred constant step sizes.

	Local init.				Global init.			
	PGA	OMA	Opt. OMA	PPO	PGA	OMA	Opt. OMA	PPO
HIT — full convergence classes[†]								
Unique separating	1	1	1	3	4	5	4	11
Unique pooling (m_0)	1	2	1	5	9.5	13	12	20
Unique pooling (m_1)	1	2	1	5[†]	8	13	11	21
Pooling & separating	1	2	1	4	4	6	5	15
Mult. sep.	1	2	1	3	6	7	6	10
Partial convergence classes								
Semi-sep. CONV.	.000	.000	1.000	.000	.000	.000	1.000	.000
Semi-sep. DIST.	.460	.853	.000	.979	.440	.820	.000	.890
Semi-sep. HIT	–	–	40.5	–	–	–	70	–
Mult. pool. CONV.	.560	.520	.480	.880	.820	.820	.840	.980
Mult. pool. DIST.	.019	.020	.020	.060	.082	.055	.053	.001
Mult. pool. HIT	3	11.5	9	29.5	10	16	17.5	16

[†] Exception: PPO, Unique pooling (m_1), local init.: CONV.$= 0.98$.

Convergence is evaluated relative to the equilibrium set associated with the classified game. In unique-equilibrium cases, the target is a singleton; in multiple-pure games, the target is the full set of pure equilibria, so a run counts as successful if it approaches any one of them. For each treatment combination we report convergence frequency (CONV.), average terminal distance to the target set (DIST.), and median hitting time (HIT) over $n = 50$ paired-init seeds. Precisely, DIST. evaluates the Frobenius norm of the difference between the realised agent strategy and the (closest) target equilibrium strategy for the agent. For pooling targets, the receiver's response to the off-path signal is a free parameter of the equilibrium and is excluded. HIT is reported in algorithm-internal rounds, and a PPO round packages a 512-episode rollout plus 16 mini-batch gradient updates per agent, whereas a PGA / OMA / Opt.OMA round is one closed-form expected-payoff gradient step. Convergence is defined as ending within $\delta = 10^{-2}$ of the relevant equilibrium set; distance measures residual terminal error; hitting time records the first iteration at which the trajectory enters the tolerance neighborhood, and is computed only over runs that converge.

5.2 Results

Table 1 shows that all four algorithms perform very well in games with a unique pure equilibrium. For unique separating and the two unique pooling benchmarks (pooling at m_0 and at m_1), convergence is essentially perfect under both local and global initialization, although the mirror-based methods typically require more iterations from global starting points. This indicates that once the equilibrium structure is simple and the target is unique, all four learning rules locate the relevant equilibrium set reliably, with PGA generally reaching it fastest, OMA and Opt. OMA moving more gradually and PPO being the slowest among them. This pattern also holds for the Spence-style signaling game and the multiple separating regime. The algorithms reliably reach an equilibrium, but the equilibrium depends on the basin of the initial condition rather than on the algorithm. Interestingly, our instances in the full-convergence class attain strict equilibria, which aligns with our theoretical results and demonstrates positive outcomes even beyond the formal guarantees.

Performance is weaker in the multiple-pooling and unique semi-separating environments, but for different reasons. In the multiple-pooling case, convergence remains high overall, yet it is somewhat lower under local initialization for the three classical learners than under global initialization (Opt. OMA: 0.48 local vs. 0.84 global; OMA: 0.52 vs. 0.82; PGA: 0.56 vs. 0.82). A natural explanation is that "local" initialization is defined relative to the first profile in the family's pure-pooling target set, so some starting points are placed near boundary regions associated with one equilibrium but not necessarily inside a strong basin of attraction. In these cases, the learner can become trapped near a nearby non-equilibrium boundary configuration or drift toward the wrong corner region before entering the tolerance neighborhood of any pure equilibrium. By contrast, global initialization

Table 2: Aggregate results for the four-type benchmarks of Appendix H.1 ($K = 4$), with the step sizes, initializations and $n = 50$ paired seeds of Table 1. Horizon: $T = 2000$ (240 PPO rounds) for unique separating and $T = 12000$ (1440 PPO rounds) for unique semi-separating.

	Local init.				Global init.			
	PGA	OMA	Opt. OMA	PPO	PGA	OMA	Opt. OMA	PPO
HIT — full convergence class								
Unique separating	1	2	1	9	5	7	5	17
Partial convergence class								
Semi-sep. CONV.	.000	.000	1.000	.000	.000	.000	.820	.000
Semi-sep. DIST.	1.335	1.831	.009	1.904	1.346	1.820	.201	1.896
Semi-sep. HIT	–	–	9936.5	–	–	–	201	–

sometimes places the process in regions from which the overall drift more clearly selects one of the pure equilibria, which can improve eventual convergence even though the hitting times are much longer. PPO is the exception in this regard.
In the unique semi-separating benchmark [20], the learning dynamics rarely settle near the target equilibrium for three of the four algorithms. Instead of converging, most trajectories under PGA, OMA, and PPO continue to move through the state space for the full horizon: the state does not become nearly stationary, the learner keeps updating and exploring different regions, yet it does not enter the tolerance neighborhood of the semi-separating equilibrium. The sole exception is Opt. OMA, which converges for 100% of seeds under both local and global initialization with median hitting times of 40.5 and 70 rounds, respectively.

These results identify the optimistic look-ahead step as a decisive algorithmic ingredient for convergence to the mixed semi-separating equilibrium. Understanding when and why optimistic methods select equilibria in signaling games and whether this extends to broader classes of extensive-form games is a promising direction for future research.

5.3 Scaling to Four Types

We now move from $2\times2\times2$ to $4\times4\times4$ signaling games (four types, signals and actions) to test whether our observations are driven by the equilibrium structure rather than by the dimension of the problem. Classifying games by their equilibrium structure becomes considerably harder in this setting. We therefore focus on two benchmarks with a provably unique PE, one from each convergence class of Table 1. In particular, we choose a four-type extension of the honest-signaling game, whose unique PE is strict and separating, and a multi-threat extension of the reputation game of Kreps and Wilson [20], in which three types each insist on their own signal and the strategic type mixes over all four signals, giving a unique semi-separating PE. More details on these game families are given in Appendix H.1. Algorithms, step sizes, initialization schemes, and the $n = 50$ paired seeds are those of Table 1.

Table 2 shows that the qualitative picture of the binary case carries over. In the unique separating game, all four algorithms converge in every run from both initializations, with hitting times close to those for $K = 2$; only PPO slows down noticeably (9 and 17 rounds vs. 3 and 11). In the semi-separating game, Opt. OMA is again the only algorithm that converges, while PGA, OMA and PPO never enter the tolerance neighborhood and end far from the equilibrium. The main caveat is speed. Under the horizon $T = 2000$ of Table 1, no Opt. OMA run from a local initialization has converged yet, so we extend the horizon to $T = 12000$. All local runs then converge, but they first enter the tolerance neighborhood only after $6{,}033$ to $10{,}630$ iterations, compared with a median of 40.5 for $K = 2$. Under global initialization, 82% of the runs converge. Half of all runs do so within 400 iterations, the remaining converging runs take several thousand, and the other 18% stay far from the equilibrium. Thus, a larger dimension makes convergence to the mixed equilibrium much slower, but it does not change which algorithms succeed.

Acknowledgments and Disclosure of Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program (grant agreement No 101198689). The authors declare no competing interests.

References

[1] Deeksha Adil, Brian Bullins, Arun Jambulapati, and Sushant Sachdeva. Optimal methods for higher-order smooth monotone variational inequalities. *arXiv preprint arXiv:2205.06167*, 2022.

[2] Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Andrea Celli, and Tuomas Sandholm. Faster no-regret learning dynamics for extensive-form correlated and coarse correlated equilibria. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, EC '22, pages 915–916, 2022.

[3] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer International Publishing, 2017.

[4] Sebastian Bervoets, Mario Bravo, and Mathieu Faure. Learning with minimal information in continuous games. *Theoretical Economics*, 15(4):1471–1508, 2020. doi: 10.3982/TE3435.

[5] Benjamin Chasnov, Lillian J Ratliff, Eric Mazumdar, and Samuel A Burden. Convergence analysis of gradient-based learning with non-uniform learning rates in non-cooperative multi-agent settings. *arXiv preprint arXiv:1906.00731*, 2019.

[6] In-Koo Cho and David M Kreps. Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 102(2):179–221, 1987. doi: 10.2307/1885060.

[7] In-Koo Cho and Joel Sobel. Strategic stability and uniqueness in signaling games. *Journal of Economic Theory*, 50(2):381–413, 1990. doi: 10.1016/0022-0531(90)90009-9.

[8] Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6): 1431–1451, 1982.

[9] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with optimism. In *International Conference on Learning Representations*, 2018.

[10] Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*, volume I of *Springer Series in Operations Research*. Springer, 2003.

[11] Gabriele Farina, Tommaso Bianchi, and Tuomas Sandholm. Coarse correlation in extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1934–1941, 2020.

[12] Gabriele Farina, Andrea Celli, Nicola Gatti, and Tuomas Sandholm. Connecting optimal ex-ante collusion in teams to extensive-form correlation: Faster algorithms and positive complexity results. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pages 3164–3173. PMLR, 2021.

[13] Gabriele Farina, Andrea Celli, Alberto Marchesi, and Nicola Gatti. Simple uncoupled no-regret learning dynamics for extensive-form correlated equilibrium. *Journal of the ACM*, 69(6):1–41, 2022. doi: 10.1145/3563772.

[14] Drew Fudenberg and Kevin He. Learning and type compatibility in signaling games. *Econometrica*, 86(4):1215–1255, 2018. doi: 10.3982/ECTA15085.

[15] Drew Fudenberg and Kevin He. Payoff information and learning in signaling games. *Games and Economic Behavior*, 120:96–120, 2020.

[16] Drew Fudenberg and David M. Kreps. Learning in extensive-form games I. Self-Confirming equilibria. *Games and Economic Behavior*, 8(1):20–55, 1995.

[17] Drew Fudenberg and David K. Levine. Self-confirming equilibrium. *Econometrica*, 61(3): 523–545, 1993.

[18] Ebbe Hendon, Hans Jørgen Jacobsen, and Birgitte Sloth. Fictitious play in extensive form games. *Games and Economic Behavior*, 15(2):177–202, 1996. doi: 10.1006/game.1996.0065.

[19] Josef Hofbauer and William H. Sandholm. Stable games and their dynamics. *Journal of Economic Theory*, 144(4):1665–1693, 2009. doi: 10.1016/j.jet.2009.01.007.

[20] David M. Kreps and Robert Wilson. Reputation and imperfect information. *Journal of Economic Theory*, 27(2):253–279, 1982.

[21] Tianyi Lin and Michael I Jordan. Perseus: A simple and optimal high-order method for variational inequalities. *Mathematical Programming*, 209:609–650, 2025.

[22] Mingyang Liu, Asuman E. Ozdaglar, Tiancheng Yu, and Kaiqing Zhang. The power of regularization in solving extensive-form games. In *The Eleventh International Conference on Learning Representations*, 2023.

[23] George J Mailath. Incentive compatibility in signaling games with a continuum of types. *Econometrica: Journal of the Econometric Society*, pages 1349–1365, 1987.

[24] Alberto Marchesi and Nicola Gatti. Trembling-hand perfection and correlation in sequential games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5566–5574, 2021.

[25] Linjian Meng, Tianpei Yang, Youzhi Zhang, Zhenxing Ge, Shangdong Yang, Tianyu Ding, Wenbin Li, Bo An, and Yang Gao. Efficient last-iterate convergence in solving extensive-form games. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[26] Panayotis Mertikopoulos and Zhengyuan Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173:465–507, 2019.

[27] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. Cycles in adversarial regularized learning. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2703–2717. SIAM, 2018.

[28] Aryan Mokhtari, Asuman E. Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108 of *Proceedings of Machine Learning Research*, pages 1497–1507, 2020.

[29] Garey Ramey. D1 signaling equilibria with multiple signals and a continuum of types. *Journal of Economic Theory*, 69(2):508–531, 1996.

[30] Lillian J. Ratliff, Samuel A. Burden, and S. Shankar Sastry. Characterization and computation of local Nash equilibria in continuous games. In *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 917–924, 2013.

[31] Hang Ren, Xiaozhen Sun, Tianzi Ma, Jiajia Zhang, and Xuan Wang. Last-iterate convergence in adaptive regret minimization for approximate extensive-form perfect equilibrium. In *ECAI 2025: 28th European Conference on Artificial Intelligence Proceedings*, pages 4686–4693, 2025.

[32] John G. Riley. Informational equilibrium. *Econometrica*, 47(2):331–359, 1979.

[33] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[34] Joel Sobel. Signaling games. In *Complex social and behavioral systems: Game theory and agent-based models*, pages 251–268. Springer, 2020.

[35] Chaobing Song, Zhengyuan Zhou, Yichao Zhou, Yong Jiang, and Yi Ma. Optimistic dual extrapolation for coherent non-monotone variational inequalities. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 14303–14314. Curran Associates, Inc., 2020.

[36] Michael Spence. Job market signaling. In *Uncertainty in economics*, pages 281–306. Elsevier, 1978.

[37] Jean Jacques Strodiot, Phan Tu Vuong, and Thi Thu Van Nguyen. A class of shrinking projection extragradient methods for solving non-monotone equilibrium problems in Hilbert spaces. *Journal of Global Optimization*, 64(1):159–178, 2016.

[38] Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[39] Paul Tseng. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2):237–252, 1995.

[40] Bernhard von Stengel and Françoise Forges. Extensive-form correlated equilibrium: Definition and computational complexity. *Mathematics of Operations Research*, 33(4):1002–1022, 2008.

[41] Weiran Wang and Miguel A. Carreira-Perpinan. Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. In *arXiv preprint arXiv:1309.1541*, 2013.

[42] Minglu Ye. An infeasible projection type algorithm for nonmonotone variational inequalities. *Numerical Algorithms*, 89(4):1723–1742, 2022.

[43] Zhengyuan Zhou, Panayotis Mertikopoulos, Aris L Moustakas, Nicholas Bambos, and Peter Glynn. Mirror descent learning in continuous games. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 5776–5783. IEEE, 2017.

[44] Martin Zinkevich, Michael Johanson, Michael H. Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. In John C. Platt, Daphne Koller, Yoram Singer, and Sam T. Roweis, editors, *Advances in Neural Information Processing Systems 20 (NIPS 2007)*, pages 1729–1736. MIT Press, 2007.

A Overview of Results

Below we present a tabular overview of our results.

Table 3: Overview of theoretical and empirical convergence results for the signaling-game classes considered in this paper. **Theoretical (Sections 3 and 4)** Here, $X^\circ(\mathcal{G})$ denotes the Bayes-defined region of the game in (2) and $q^\star$ is a PE. **Empirical (Section 5):** ✓ = reliably converges, ○ = initialization-dependent / partial, × = does not converge. We say that the Minty condition (condition (ii) in Definition 5) fails for a class, if there is at least one instance of that class that does not satisfy it.

Class of signaling games	**Structural condition**	**Theoretical**	**Empirical**	**Result**
unique separating	strict PE ($q^\star \in X^\circ(\mathcal{G})$)	Minty fails Variational Stability	PGA ✓ OMA ✓ Opt. OMA ✓ PPO ✓	Thms. 1, 2; experimental
multiple separating; pooling-&-separating	strict PE ($\exists q^\star \in X^\circ(\mathcal{G})$)	Unknown	PGA ✓ OMA ✓ Opt. OMA ✓ PPO ✓	experimental
unique pooling	strict PE ($q^\star \notin X^\circ(\mathcal{G})$)	Unknown	PGA ✓ OMA ✓ Opt. OMA ✓ PPO ✓	experimental
multiple pooling	strict PE ($q^\star \notin X^\circ(\mathcal{G})$)	Unknown	PGA ○ OMA ○ Opt. OMA ○ PPO ✓	experimental
honest-signaling (Assumptions 1 and 2)	unique strict PE ($q^\star \in X^\circ(\mathcal{G})$)	Minty fails Variational Stability Unique EFCCE	PGA ✓ OMA ✓ Opt. OMA ✓ PPO ✓	Thms. 1, 2, 5; experimental
unique semi-separating	mixed PE	Unknown	PGA × OMA × Opt. OMA ○ PPO ×	experimental

B Further Related Work

Signaling games and learning under asymmetric information. Signaling games are a canonical model of strategic interaction under asymmetric information, with applications throughout economics and related fields [6, 23, 34, 36]. They are also a natural minimal testbed for our question: they are simple enough to analyze explicitly, yet already contain the key ingredients absent from normal-form models, namely sequential play, belief updating, and Bayes-consistency constraints. A large economics literature studies equilibrium structure and refinements in signaling games, including perfect Bayesian equilibrium, sequential equilibrium, the Intuitive Criterion, and D1 [6, 7, 29]. Our focus is different. We do not study equilibrium refinement or selection through off-path beliefs per se; instead, we use signaling games as the simplest incomplete-information extensive-form environment in which Bayesian updating changes the learning geometry.

There is also prior work on learning in signaling and sender–receiver games. Some of this literature studies how learning shapes off-path beliefs and equilibrium selection in signaling environments [14, 15], while older work emphasizes that uncoupled learning in extensive-form games may converge only to self-confirming equilibrium rather than Nash equilibrium [16–18].

Variational inequalities and monotone game learning. A large literature explains convergence of first-order methods in normal-form games using structural conditions such as monotonicity and the

Minty variational inequality. These conditions play a role analogous to convexity in optimization and underlie convergence guarantees for projection methods, extragradient methods, mirror descent, and optimistic variants [1, 3, 5, 21, 30, 35, 37, 39, 42]. If a game is strongly monotone, then simple first-order algorithms such as OMA and PGA converge. If a game satisfies the Minty condition globally, extragradient algorithms are known to converge. These are the main known sufficient conditions for obtaining equilibrium convergence. Our paper analyzes monotonicity and the Minty condition in extensive-form signaling games and draws on techniques from dynamical systems to certify last-iterate convergence locally.

C Missing Material of Section 2

C.1 Proof of Lemma 1

Lemma 1. *For every signaling game $\mathcal{G}$, $X^\circ(\mathcal{G})$ is convex and open in $X(\mathcal{G})$.*

Proof. Fix $q', q'' \in X^\circ(\mathcal{G})$ and $t \in [0,1]$, and set $\tilde{q} = (1-t)\,q' + t\,q''$. For every $s \in \Sigma$, we have:

$$\sum_{\theta\in\Theta} \pi(\theta)\,\tilde{q}_S(s \mid \theta) \;=\; (1-t)\sum_{\theta\in\Theta} \pi(\theta)\,q'_S(s \mid \theta) \;+\; t\sum_{\theta\in\Theta} \pi(\theta)\,q''_S(s \mid \theta) \;>\; 0.$$

Here, the strict inequality follows from the fact that both $q', q'' \in X^\circ(\mathcal{G})$. Hence, $\tilde{q} \in X^\circ(\mathcal{G})$ and convexity follows.

To complete the proof, we prove that $X^\circ(\mathcal{G})$ is open in $X(\mathcal{G})$. For each $s \in \Sigma$, $\phi_s(q) := \sum_{\theta\in\Theta} \pi(\theta)\,q_S(s \mid \theta)$ is continuous on $X(\mathcal{G})$. Therefore, the set $\{q \in X(\mathcal{G}) : \phi_s(q) > 0\}$ is open in $X(\mathcal{G})$. As a finite intersection of these sets, $X^\circ(\mathcal{G})$ is open in $X(\mathcal{G})$. □

C.2 Proof of Lemma 2

Lemma 2. *The payoff-difference field g is continuous on $X^\circ(\mathcal{G})$.*

Proof of Lemma 2. For every $\theta \in \Theta$ and every $s \in \Sigma$, $g_S(\theta, s;\, q)$ (defined in (6)) is linear in q_R and does not depend on q_S. Linear functions are continuous, so $g_S(\theta, s;\, \cdot)$ is continuous on $X^\circ(\mathcal{G})$.

For every $s \in \Sigma$ and every $a \in A$, $g_R(s, a;\, q)$ (defined in (7)) depends on q only through the Bayes posterior

$$\mu(\theta \mid s;\, q) \;=\; \frac{\pi(\theta)\,q_S(s \mid \theta)}{\sum_{\theta'\in\Theta} \pi(\theta')\,q_S(s \mid \theta')}.$$

The numerator and denominator are linear in q_S, and the denominator is strictly positive on $X^\circ(\mathcal{G})$ by definition. Therefore, $\mu(\theta \mid s;\, q)$ is continuous in q on $X^\circ(\mathcal{G})$, and so is $g_R(s, a;\, q)$.

Since both components are continuous, g is continuous on $X^\circ(\mathcal{G})$. □

D Missing Material of Section 3

D.1 Proof of Theorem 1

Theorem 1. *There exists a signaling game $\mathcal{G}$ with $|\Theta| = |\Sigma| = |A| = 2$ and a strict separating* PE *$q^\star \in X^\circ(\mathcal{G})$ at which the Minty condition fails.*

Proof. Consider the signaling game $\mathcal{G}$ with $\Theta = \{\theta_1, \theta_2\}$, $\Sigma = \{s_1, s_2\}$, $A = \{a_1, a_2\}$, prior $\pi(\theta_1) = \pi(\theta_2) = 1/2$, and utilities

$$u_R(\theta_k, s, a_j) \;=\; \mathbf{1}\{j = k\}, \qquad u_S(\theta_k, s_i, a_j) \;=\; \tfrac{3}{2}\,\mathbf{1}\{i = k\} + \mathbf{1}\{j = k\}.$$

Let $q^\star$ be the Riley outcome, with $q^\star_S(s_k \mid \theta_k) = 1$ and $q^\star_R(a_k \mid s_k) = 1$, and let $\mu^\star$ be such that $\mu^\star(\theta_k \mid s_k) = 1$. Observe that $q^\star \in X^\circ$ since $\sum_{\theta\in\Theta} \pi(\theta) q^\star_S(s_k \mid \theta) = \pi(\theta_k) = 1/2 > 0$ holds for $k \in \{1, 2\}$.

To verify that $q^\star$ is a PE (Definition 1), we check sequential rationality and consistency. Fix $k \in \{1, 2\}$. For every $i \in \{1, 2\}$, the Sender's expected utility at type θ_k from signal s_i against $q^\star_R$

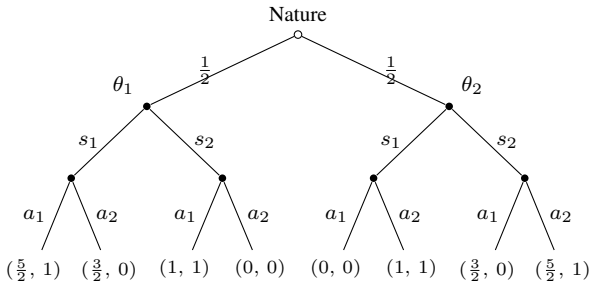


Figure 1: Game tree for the instance used in Theorem 1. Terminal payoffs are (u_S, u_R).

is $\sum_a u_S(\theta_k, s_i, a)\, q_R^\star(a \mid s_i) = u_S(\theta_k, s_i, a_i) = \frac{3}{2}\mathbf{1}\{i = k\} + \mathbf{1}\{i = k\} = \frac{5}{2}\mathbf{1}\{i = k\}$, uniquely maximized at $i = k$. For every $j \in \{1, 2\}$, the Receiver's expected utility at signal s_k from action a_j under belief $\mu^\star(\theta_k \mid s_k) = 1$ is $u_R(\theta_k, s_k, a_j) = \mathbf{1}\{j = k\}$, uniquely maximized at $j = k$. Consistency is immediate, since $\mu^\star$ equals the Bayes posterior at every on-path signal. Hence $q^\star$ is a strict PE.

To show that the Minty condition fails on $X^\circ(\mathcal{G})$, consider the profile $\bar{q}$ defined by $\bar{q}_S(s_j \mid \theta_k) = \mathbf{1}\{j \neq k\}$ and $\bar{q}_R(a_j \mid s_k) = \mathbf{1}\{j \neq k\}$ for every $j, k \in \{1, 2\}$. The signal marginals at $\bar{q}_S$ satisfy $\sum_{\theta \in \Theta} \pi(\theta)\, \bar{q}_S(s_k \mid \theta) = 1/2$ for every $k \in \{1, 2\}$, so $\bar{q} \in X^\circ(\mathcal{G})$, and Bayes' rule gives $\mu(\theta_2 \mid s_1; \bar{q}) = \mu(\theta_1 \mid s_2; \bar{q}) = 1$. With reference profile $(\bar{\sigma}, \bar{\alpha}) = (\sigma^\star, \alpha^\star)$ (observe that this is without loss of generality by Lemma 1), we compute $g_S(\theta_k, s_j; \bar{q})$ and $g_R(s_k, a_j; \bar{q})$ for every $j, k \in \{1, 2\}$ with $j \neq k$. For the Sender,

$$g_S(\theta_1, s_2; \bar{q}) \;=\; u_S(\theta_1, s_2, a_1) - u_S(\theta_1, s_1, a_2) \;=\; 1 - \tfrac{3}{2} \;=\; -\tfrac{1}{2},$$

using $\bar{q}_R(a_1 \mid s_2) = 1$ and $\bar{q}_R(a_2 \mid s_1) = 1$, and by symmetry $g_S(\theta_2, s_1; \bar{q}) = -1/2$. For the Receiver,

$$g_R(s_1, a_2; \bar{q}) \;=\; u_R(\theta_2, s_1, a_2) - u_R(\theta_2, s_1, a_1) \;=\; 1,$$

using $\mu(\theta_2 \mid s_1; \bar{q}) = 1$, and by symmetry $g_R(s_2, a_1; \bar{q}) = 1$. By our choice of reference profile, $g_S(\theta_k, s_k; \bar{q}) = 0$ and $g_R(s_k, a_k; \bar{q}) = 0$ for every $k \in \{1, 2\}$, so these terms vanish in the inner product. The remaining four terms satisfy $\bar{q}_S(s_j \mid \theta_k) - q_S^\star(s_j \mid \theta_k) = 1$ and $\bar{q}_R(a_j \mid s_k) - q_R^\star(a_j \mid s_k) = 1$ for every $j \neq k$. Therefore,

$$\begin{aligned}\langle g(\bar{q}),\, \bar{q} - q^\star \rangle \;&=\; g_S(\theta_1, s_2; \bar{q}) + g_S(\theta_2, s_1; \bar{q}) + g_R(s_1, a_2; \bar{q}) + g_R(s_2, a_1; \bar{q}) \\ &=\; -\tfrac{1}{2} - \tfrac{1}{2} + 1 + 1 \;=\; 1 \;>\; 0,\end{aligned}$$

which violates the Minty condition at $q^\star$ (Definition 5(ii)). □

D.2 Proof of Lemma 3

Lemma 3. *For a strict* PE $q^\star \in X^\circ(\mathcal{G})$*, it holds that* $\delta(\mathcal{G}, q^\star) > 0$ *and* $\bar{u}_R(\mathcal{G}, q^\star) > 0$*.*

Proof. Since $q^\star$ is strict, every type $\theta \in \Theta$ has a unique best response $\sigma^\star(\theta)$ against $q_R^\star$, and therefore $g_S(\theta, s; q^\star) < 0$ holds for every $s \neq \sigma^\star(\theta)$. Similarly, every signal $s \in \Sigma$ has a unique best response $\alpha^\star(s)$ under $\mu^\star(\cdot \mid s)$, and therefore $g_R(s, a; q^\star) < 0$ holds for every $a \neq \alpha^\star(s)$. Taking the minimum of absolute values over the finite index sets gives $\delta(\mathcal{G}, q^\star) > 0$.

To show that $\bar{u}_R(\mathcal{G}, q^\star) > 0$ suppose (toward contradiction) the contrary. If $\bar{u}_R(\mathcal{G}, q^\star) = 0$, then $u_R(\theta, s, a) = u_R(\theta, s, \alpha^\star(s))$ for every θ, s, a, so the receiver is indifferent between every $a \in A \setminus \{\alpha^\star(s)\}$ and $\alpha^\star(s)$ at every signal s. This contradicts that the strict PE $q^\star$ has a unique best response for every $s \in \Sigma$. □

D.3 Proof of Lemma 4

Lemma 4. *Let* $q^\star \in X^\circ(\mathcal{G})$ *be a strict* PE *of* $\mathcal{G}$*. Then for every* $q \in X^\circ(\mathcal{G})$*,* $\langle g(q^\star),\, q - q^\star \rangle \;\leq\; -\frac{\delta(\mathcal{G}, q^\star)}{2}\, \|q - q^\star\|_1$*.*

Proof. Fix $q \in X^\circ(\mathcal{G})$. We have:

$$\begin{aligned}
\langle g(q^\star), q - q^\star \rangle &= \sum_{\theta\in\Theta}\sum_{s\in\Sigma} g_S(\theta, s;\, q^\star)\big[q_S(s \mid \theta) - q_S^\star(s \mid \theta)\big] \\
&\quad + \sum_{s\in\Sigma}\sum_{a\in A} g_R(s, a;\, q^\star)\big[q_R(a \mid s) - q_R^\star(a \mid s)\big] \\
&= \sum_{\theta\in\Theta}\sum_{s\neq\sigma^\star(\theta)} g_S(\theta, s;\, q^\star)\, q_S(s \mid \theta) + \sum_{s\in\Sigma}\sum_{a\neq\alpha^\star(s)} g_R(s, a;\, q^\star)\, q_R(a \mid s) \\
&\leq -\delta(\mathcal{G}, q^\star) \cdot \Big(\sum_{\theta\in\Theta}\sum_{s\neq\sigma^\star(\theta)} q_S(s \mid \theta) + \sum_{s\in\Sigma}\sum_{a\neq\alpha^\star(s)} q_R(a \mid s)\Big) \\
&= -\frac{\delta(\mathcal{G}, q^\star)}{2} \cdot \|q - q^\star\|_1.
\end{aligned}$$

Here, the second equality follows since $q_S^\star(s \mid \theta) = 0$ for $s \neq \sigma^\star(\theta)$ and $q_R^\star(a \mid s) = 0$ for $a \neq \alpha^\star(s)$ (by definition of $\sigma^\star, \alpha^\star$), together with $g_S(\theta, \sigma^\star(\theta); q^\star) = 0$ and $g_R(s, \alpha^\star(s); q^\star) = 0$ (by our choice of reference). The inequality holds since $g_S(\theta, s; q^\star) \leq -\delta(\mathcal{G}, q^\star)$ for $s \neq \sigma^\star(\theta)$ and $g_R(s, a; q^\star) \leq -\delta(\mathcal{G}, q^\star)$ for $a \neq \alpha^\star(s)$. To see this, observe that for every $\theta \in \Theta$ and $s \neq \sigma^\star(\theta)$, and for every $s \in \Sigma$ and $a \neq \alpha^\star(s)$, respectively,

$$\begin{aligned}
g_S(\theta, s; q^\star) &= u_S\big(\theta, s, \alpha^\star(s)\big) - u_S\big(\theta, \sigma^\star(\theta), \alpha^\star(\sigma^\star(\theta))\big) < 0, \\
g_R(s, a; q^\star) &= \sum_{\theta\in\Theta} \mu^\star(\theta \mid s)\, u_R(\theta, s, a) - \sum_{\theta\in\Theta} \mu^\star(\theta \mid s)\, u_R\big(\theta, s, \alpha^\star(s)\big) < 0.
\end{aligned}$$

Here, the equalities follow from (6) and (7), since $q_R^\star$ is pure and $(\bar\sigma, \bar\alpha) = (\sigma^\star, \alpha^\star)$, and the strict inequalities are exactly the Sender's and the Receiver's strictness conditions in the definition of a strict PE (Section 2). As both quantities are strictly negative and $\delta(\mathcal{G}, q^\star)$ is the minimum of their absolute values over all such pairs, the two inequalities follow (this is also the argument in the proof of Lemma 3). The last equality follows from Claim 1, which we state and prove below.

Claim 1. *For every $q \in X^\circ(\mathcal{G})$,*

$$\sum_{\theta\in\Theta}\sum_{s\neq\sigma^\star(\theta)} q_S(s \mid \theta) \;+\; \sum_{s\in\Sigma}\sum_{a\neq\alpha^\star(s)} q_R(a \mid s) \;=\; \tfrac{1}{2}\,\|q - q^\star\|_1.$$

Proof. Fix $\theta \in \Theta$. Since $q^\star$ is a strict PE, $q_S^\star(\cdot \mid \theta)$ is the point mass on $\sigma^\star(\theta)$, i.e., $q_S^\star(\sigma^\star(\theta) \mid \theta) = 1$ and $q_S^\star(s \mid \theta) = 0$ for every $s \neq \sigma^\star(\theta)$. Hence,

$$\begin{aligned}
\|q_S(\cdot \mid \theta) - q_S^\star(\cdot \mid \theta)\|_1 &= |q_S(\sigma^\star(\theta) \mid \theta) - 1| + \sum_{s\neq\sigma^\star(\theta)} |q_S(s \mid \theta) - 0| \\
&= \big(1 - q_S(\sigma^\star(\theta) \mid \theta)\big) + \sum_{s\neq\sigma^\star(\theta)} q_S(s \mid \theta) \\
&= 2 \sum_{s\neq\sigma^\star(\theta)} q_S(s \mid \theta),
\end{aligned}$$

where the first equality expands the ℓ_1 norm componentwise, using the values of $q_S^\star(\cdot \mid \theta)$ above, the second drops the absolute values, since $q_S(\cdot \mid \theta) \in \Delta(\Sigma)$ gives $q_S(\sigma^\star(\theta) \mid \theta) \leq 1$ and $q_S(s \mid \theta) \geq 0$ for every $s \in \Sigma$, and the last uses $1 - q_S(\sigma^\star(\theta) \mid \theta) = \sum_{s\neq\sigma^\star(\theta)} q_S(s \mid \theta)$, again since $q_S(\cdot \mid \theta)$ is a probability distribution. Summing over $\theta \in \Theta$,

$$\sum_{\theta\in\Theta}\sum_{s\neq\sigma^\star(\theta)} q_S(s \mid \theta) \;=\; \tfrac{1}{2}\,\|q_S - q_S^\star\|_1,$$

and an analogous identity for the receiver gives $\sum_{s\in\Sigma}\sum_{a\neq\alpha^\star(s)} q_R(a \mid s) = \frac{1}{2}\,\|q_R - q_R^\star\|_1$. Adding the two and using the decomposition $\|q - q^\star\|_1 = \|q_S - q_S^\star\|_1 + \|q_R - q_R^\star\|_1$ gives the claim. □

The proof of the Lemma follows. □

D.4 Proof of Lemma 5

Lemma 5. *Let $q^\star \in X^\circ(\mathcal{G})$ be a strict* PE *of $\mathcal{G}$. Then, for every $q \in X^\circ(\mathcal{G})$, $\langle g(q)-g(q^\star),\, q-q^\star\rangle \;\le\; L\cdot\|q-q^\star\|_1^2$, where $L=\max\Big\{\bar u_S(\mathcal{G},q^\star),\ \frac{2\,\bar u_R(\mathcal{G},q^\star)}{\min_{\theta\in\Theta}\pi(\theta)}\Big\}$.*

Proof. With reference $(\bar\sigma,\bar\alpha)=(\sigma^\star,\alpha^\star)$, $g_S(\theta,\sigma^\star(\theta);q)=0$ and $g_R(s,\alpha^\star(s);q)=0$ for every $q\in X^\circ(\mathcal{G})$, so $g(q)-g(q^\star)$ vanishes at the equilibrium components and we restrict the analysis to $s\neq\sigma^\star(\theta)$ for the sender and $a\neq\alpha^\star(s)$ for the receiver.

Fix $q\in X^\circ(\mathcal{G})$, $\theta\in\Theta$, and $s\in\Sigma$ with $s\neq\sigma^\star(\theta)$. We have:

$$
\begin{aligned}
&\big|g_S(\theta,s;\,q)-g_S(\theta,s;\,q^\star)\big| \\
&\quad=\Bigg|\sum_{a\in A}u_S(\theta,s,a)\,\big[q_R(a\mid s)-q_R^\star(a\mid s)\big] \\
&\qquad-\sum_{a\in A}u_S\big(\theta,\sigma^\star(\theta),a\big)\,\big[q_R(a\mid\sigma^\star(\theta))-q_R^\star(a\mid\sigma^\star(\theta))\big]\Bigg| \\
&\quad=\Bigg|\sum_{a\in A}\big[u_S(\theta,s,a)-u_S(\theta,s,\alpha^\star(s))\big]\,\big[q_R(a\mid s)-q_R^\star(a\mid s)\big] \\
&\qquad-\sum_{a\in A}\big[u_S\big(\theta,\sigma^\star(\theta),a\big)-u_S\big(\theta,\sigma^\star(\theta),\alpha^\star(\sigma^\star(\theta))\big)\big]\,\big[q_R(a\mid\sigma^\star(\theta))-q_R^\star(a\mid\sigma^\star(\theta))\big]\Bigg| \\
&\quad\le\bar u_S(\mathcal{G},q^\star)\cdot\Big(\|q_R(\cdot\mid s)-q_R^\star(\cdot\mid s)\|_1+\|q_R(\cdot\mid\sigma^\star(\theta))-q_R^\star(\cdot\mid\sigma^\star(\theta))\|_1\Big) \\
&\quad\le\bar u_S(\mathcal{G},q^\star)\cdot\|q_R-q_R^\star\|_1\;\le\;\bar u_S(\mathcal{G},q^\star)\cdot\|q-q^\star\|_1. \qquad (8)
\end{aligned}
$$

The second equality holds since $\sum_{a\in A}[q_R(a\mid s)-q_R^\star(a\mid s)]=0$. Then, the first inequality applies the triangle inequality and the bound $\max_a|u_S(\theta,s',a)-u_S(\theta,s',\alpha^\star(s'))|\le\bar u_S(\mathcal{G},q^\star)$ from the definition of $\bar u_S(\mathcal{G},q^\star)$, valid for every signal s'.

To establish a similar argument for the receiver, we first prove the following Claim.

Claim 2. *For every $q\in X^\circ(\mathcal{G})$ and every $s\in\Sigma$,*

$$\sum_{\theta\in\Theta}\big|\mu(\theta\mid s;\,q)-\mu(\theta\mid s;\,q^\star)\big|\;\le\;\frac{2\,\|q_S-q_S^\star\|_1}{\min_{\theta\in\Theta}\pi(\theta)}.$$

Proof. Fix $q\in X^\circ(\mathcal{G})$ and $s\in\Sigma$. Write $f(\theta):=\pi(\theta)\,q_S(s\mid\theta)$ and $f^\star(\theta):=\pi(\theta)\,q_S^\star(s\mid\theta)$, with $Z:=\sum_\theta f(\theta)>0$ and $Z^\star:=\sum_\theta f^\star(\theta)>0$ (both positive since $q,q^\star\in X^\circ(\mathcal{G})$). Then $\mu(\theta\mid s;\,q)=f(\theta)/Z$ and $\mu(\theta\mid s;\,q^\star)=f^\star(\theta)/Z^\star$.

Since $q^\star$ is a strict PE, $q_S^\star(\cdot\mid\theta)$ has unit support at $\sigma^\star(\theta)$. The signal s being on-path under $q^\star$ means there exists $\theta_s^\star\in\Theta$ with $\sigma^\star(\theta_s^\star)=s$, so that

$$Z^\star\;=\;\sum_{\theta\in\Theta}\pi(\theta)\,q_S^\star(s\mid\theta)\;\ge\;\pi(\theta_s^\star)\;\ge\;\min_{\theta\in\Theta}\pi(\theta). \qquad (9)$$

Adding and subtracting $f(\theta)\,Z$ in the numerator,

$$
\begin{aligned}
\mu(\theta\mid s;\,q)-\mu(\theta\mid s;\,q^\star)\;&=\;\frac{f(\theta)\,Z^\star-f^\star(\theta)\,Z}{Z\,Z^\star}\;=\;\frac{f(\theta)\,(Z^\star-Z)+Z\,(f(\theta)-f^\star(\theta))}{Z\,Z^\star} \\
&=\;\mu(\theta\mid s;\,q)\,\frac{Z^\star-Z}{Z^\star}\;+\;\frac{f(\theta)-f^\star(\theta)}{Z^\star}.
\end{aligned}
$$

Applying the triangle inequality term by term and summing over θ, using $\sum_{\theta\in\Theta}\mu(\theta\mid s;\,q)=1$, and combining with $|Z^\star-Z|\le\|f-f^\star\|_1$, $\|f-f^\star\|_1\le\|q_S-q_S^\star\|_1$, and (9),

$$
\begin{aligned}
\sum_{\theta\in\Theta}\big|\mu(\theta\mid s;\,q)-\mu(\theta\mid s;\,q^\star)\big|\;&\le\;\frac{|Z^\star-Z|}{Z^\star}\;+\;\frac{\|f-f^\star\|_1}{Z^\star}\;\le\;\frac{2\,\|f-f^\star\|_1}{Z^\star} \\
&\le\;\frac{2\,\|q_S-q_S^\star\|_1}{Z^\star}\;\le\;\frac{2\,\|q_S-q_S^\star\|_1}{\min_{\theta\in\Theta}\pi(\theta)},
\end{aligned}
$$

where $|Z^\star - Z| \le \|f - f^\star\|_1$ is the triangle inequality applied to $|\sum_{\theta\in\Theta}(f^\star(\theta) - f(\theta))|$, and $\|f - f^\star\|_1 \le \|q_S - q_S^\star\|_1$ uses $\pi(\theta) \le 1$ together with enlarging the inner sum from the single signal s to all of Σ. □

Fix $q \in X^\circ(\mathcal{G})$, $s \in \Sigma$, and $a \in A$ with $a \neq \alpha^\star(s)$. Then:

$$\begin{aligned}\big|g_R(s,a;\,q) - g_R(s,a;\,q^\star)\big| &= \left|\sum_{\theta\in\Theta}\big[u_R(\theta,s,a) - u_R(\theta,s,\alpha^\star(s))\big]\,\big[\mu(\theta \mid s;\,q) - \mu(\theta \mid s;\,q^\star)\big]\right| \\ &\le \max_{\theta\in\Theta}\big|u_R(\theta,s,a) - u_R(\theta,s,\alpha^\star(s))\big| \cdot \sum_{\theta\in\Theta}\big|\mu(\theta \mid s;\,q) - \mu(\theta \mid s;\,q^\star)\big| \\ &\le \bar{u}_R(\mathcal{G},q^\star)\cdot\frac{2\,\|q_S - q_S^\star\|_1}{\min_{\theta\in\Theta}\pi(\theta)} \;\le\; \frac{2\,\bar{u}_R(\mathcal{G},q^\star)}{\min_{\theta\in\Theta}\pi(\theta)}\cdot\|q - q^\star\|_1. \qquad (10)\end{aligned}$$

The first inequality is the triangle inequality; the second uses the definition of $\bar{u}_R(\mathcal{G},q^\star)$ and Claim 2.

Set $L := \max\big\{\bar{u}_S(\mathcal{G},q^\star),\, 2\bar{u}_R(\mathcal{G},q^\star)/\min_{\theta\in\Theta}\pi(\theta)\big\}$. From (8) and (10),

$$\begin{aligned}\|g(q) - g(q^\star)\|_\infty &= \max_{\theta\in\Theta,\,s\in\Sigma,\,a\in A}\Big\{\big|g_S(\theta,s;\,q) - g_S(\theta,s;\,q^\star)\big|,\;\big|g_R(s,a;\,q) - g_R(s,a;\,q^\star)\big|\Big\} \\ &\le\; L\cdot\|q - q^\star\|_1.\end{aligned}$$

We now use Hölder's inequality to obtain

$$\langle g(q) - g(q^\star),\, q - q^\star\rangle \;\le\; \|g(q) - g(q^\star)\|_\infty\cdot\|q - q^\star\|_1 \;\le\; L\cdot\|q - q^\star\|_1^2$$

and complete the proof. □

D.5 Proof of Theorem 4

Theorem 4. *Let $q^\star \in X^\circ(\mathcal{G})$. If $q^\star$ is variationally stable, then it is a strict* PE.

Proof. Since $q^\star$ is variationally stable, by Definition 6 there exists an open neighborhood $H \subseteq X^\circ(\mathcal{G})$ such that $\langle g(q), q - q^\star\rangle \le 0$ for every $q \in H$, with strict inequality whenever $q \neq q^\star$. Because H is open in $X^\circ(\mathcal{G})$ (which is itself open in $X(\mathcal{G})$ by Lemma 1), there exists an $\varepsilon > 0$ such that $\{q \in X(\mathcal{G}) : \|q - q^\star\|_1 < \varepsilon\} \subseteq H$.

We first consider the Sender. Fix $\theta_k \in \Theta$ and $s \in \operatorname{supp}(q_S^\star(\cdot \mid \theta_k))$. For any $s' \in \Sigma$ with $s' \neq s$, let

$$\lambda_{\max} := \min\left\{q_S^\star(s \mid \theta_k),\, 1 - q_S^\star(s' \mid \theta_k),\, \frac{\sum_{\theta\in\Theta}\pi(\theta)\,q_S^\star(s \mid \theta)}{\pi(\theta_k)},\, \frac{\varepsilon}{2}\right\}.$$

For $\lambda \in (0, \lambda_{\max})$, define q_λ by

$$q_{\lambda,S}(s \mid \theta_k) := q_S^\star(s \mid \theta_k) - \lambda, \qquad q_{\lambda,S}(s' \mid \theta_k) := q_S^\star(s' \mid \theta_k) + \lambda,$$

with all other entries of $q_{\lambda,S}$ and all of $q_{\lambda,R}$ equal to those of $q^\star$. The four terms in $\lambda_{\max}$ ensure, respectively, that $q_{\lambda,S}(s \mid \theta_k) \ge 0$, that $q_{\lambda,S}(s' \mid \theta_k) \le 1$, that $q_\lambda \in X^\circ(\mathcal{G})$ (see (i) below), and that $\|q_\lambda - q^\star\|_1 = 2\lambda < \varepsilon$, so that $q_\lambda \in H$ (see (ii) below). Then:

(i) $q_\lambda \in X^\circ(\mathcal{G})$: by the definition of $\lambda_{\max}$,

$$\sum_{\theta\in\Theta}\pi(\theta)\,q_{\lambda,S}(s \mid \theta) = \sum_{\theta\in\Theta}\pi(\theta)\,q_S^\star(s \mid \theta) - \lambda\pi(\theta_k) > 0,$$

while $\sum_\theta \pi(\theta)\,q_{\lambda,S}(s' \mid \theta) > \sum_\theta \pi(\theta)\,q_S^\star(s' \mid \theta) > 0$, since $q^\star \in X^\circ(\mathcal{G})$.

(ii) $q_\lambda \neq q^\star$ and $\|q_\lambda - q^\star\|_1 = 2\lambda < \varepsilon$.

(iii) By construction of q_λ and the fact that g_S depends on q only through q_R, with $q_{\lambda,R} = q_R^\star$,

$$\langle g(q_\lambda),\, q_\lambda - q^\star\rangle = \lambda\big(g_S(\theta_k,s';\,q^\star) - g_S(\theta_k,s;\,q^\star)\big).$$

By (i)–(ii) and strict variational stability of $q^\star$, $\langle g(q_\lambda),\, q_\lambda - q^\star\rangle < 0$. Combined with (iii), this gives

$$g_S(\theta_k, s';\, q^\star) < g_S(\theta_k, s;\, q^\star) \qquad \text{for every } s' \in \Sigma \setminus \{s\}. \tag{11}$$

Since (11) holds for every $s \in \operatorname{supp}(q_S^\star(\cdot \mid \theta_k))$, the support is a singleton: if it contained two distinct signals s, s', applying (11) to (s, s') and to (s', s) would yield $g_S(\theta_k, s';\, q^\star) < g_S(\theta_k, s;\, q^\star) < g_S(\theta_k, s';\, q^\star)$, a contradiction. Writing $\sigma^\star(\theta_k)$ for the unique element, (11) says $\sigma^\star(\theta_k)$ is the strict best response for type θ_k against $q_R^\star$.

By the analogous argument with $a \in \operatorname{supp}(q_R^\star(\cdot \mid s))$, $a' \in A \setminus \{a\}$, and $\lambda_{\max} := \min\{q_R^\star(a \mid s),\, 1 - q_R^\star(a' \mid s),\, \varepsilon/2\}$ (here membership in $X^\circ(\mathcal{G})$ is automatic since $q_{\lambda,S} = q_S^\star$): for every $s \in \Sigma$, $\operatorname{supp}(q_R^\star(\cdot \mid s))$ is a singleton $\{\alpha^\star(s)\}$, and $\alpha^\star(s)$ is the strict best response under the Bayes posterior $\mu^\star(\cdot \mid s)$.

Bayes consistency holds since $q^\star \in X^\circ(\mathcal{G})$. Therefore $q^\star$ is a strict PE. □

D.6 Independence from the Reference Profile in g

The payoff-difference field of Definition 4 depends on the reference profile $(\bar\sigma, \bar\alpha)$ only conventionally: every quantity used in the variational analysis of Section 3 is invariant under that choice. To this end, we prove the following proposition.

Proposition 1. *Let $(\bar\sigma_1, \bar\alpha_1)$ and $(\bar\sigma_2, \bar\alpha_2)$ be two reference profiles, and let $g^{(1)}, g^{(2)}$ denote the payoff-difference fields induced via Definition 4. Then for every $q, q' \in X^\circ(\mathcal{G})$,*

$$\langle g^{(1)}(q'),\, q - q'\rangle \;=\; \langle g^{(2)}(q'),\, q - q'\rangle.$$

Proof. By Definition 4, for every $\theta \in \Theta$ and $s \in \Sigma$,

$$\begin{aligned} g_S^{(1)}(\theta, s;\, q) - g_S^{(2)}(\theta, s;\, q) &= \sum_{a' \in A} u_S\big(\theta, \bar\sigma_2(\theta), a'\big)\, q_R\big(a' \mid \bar\sigma_2(\theta)\big) \\ &\quad - \sum_{a' \in A} u_S\big(\theta, \bar\sigma_1(\theta), a'\big)\, q_R\big(a' \mid \bar\sigma_1(\theta)\big) \;=:\; c_S(\theta;\, q), \end{aligned}$$

which depends only on θ and q (not on s). Therefore,

$$\begin{aligned} \langle g_S^{(1)}(q'),\, q_S - q'_S\rangle - \langle g_S^{(2)}(q'),\, q_S - q'_S\rangle &= \sum_{\theta \in \Theta} \sum_{s \in \Sigma} c_S(\theta;\, q')\big[q_S(s \mid \theta) - q'_S(s \mid \theta)\big] \\ &= \sum_{\theta \in \Theta} c_S(\theta;\, q') \sum_{s \in \Sigma} \big[q_S(s \mid \theta) - q'_S(s \mid \theta)\big] \\ &= \sum_{\theta \in \Theta} c_S(\theta;\, q') \cdot 0 \;=\; 0, \end{aligned}$$

since $q_S(\cdot \mid \theta), q'_S(\cdot \mid \theta) \in \Delta(\Sigma)$ for every θ.

Analogously for the receiver, for every $s \in \Sigma$ and $a \in A$,

$$g_R^{(1)}(s, a;\, q) - g_R^{(2)}(s, a;\, q) \;=\; \sum_{\theta \in \Theta} \big[u_R\big(\theta, s, \bar\alpha_2(s)\big) - u_R\big(\theta, s, \bar\alpha_1(s)\big)\big]\, \mu(\theta \mid s) \;=:\; c_R(s;\, q),$$

which depends only on s and q (not on a). The same argument as above gives

$$\langle g_R^{(1)}(q'),\, q_R - q'_R\rangle - \langle g_R^{(2)}(q'),\, q_R - q'_R\rangle \;=\; \sum_{s \in \Sigma} c_R(s;\, q') \sum_{a \in A} \big[q_R(a \mid s) - q'_R(a \mid s)\big] \;=\; 0,$$

since $q_R(\cdot \mid s), q'_R(\cdot \mid s) \in \Delta(A)$ for every s.

Summing the sender and receiver identities gives $\langle g^{(1)}(q'),\, q - q'\rangle = \langle g^{(2)}(q'),\, q - q'\rangle$. □

Each of the three conditions in Definition 5 and Definition 6 is stated through inner products of the form $\langle g(q'),\, q - q'\rangle$:

- Monotonicity: $\langle g(q) - g(q'),\, q - q'\rangle \le 0$ for all $q, q' \in X^\circ(\mathcal{G})$.

- Minty at $q^\star$: $\langle g(q),\, q - q^\star\rangle \le 0$ for all $q \in X^\circ(\mathcal{G})$.
- Variational stability at $q^\star$: $\langle g(q),\, q - q^\star\rangle \le 0$ for q in a neighborhood of $q^\star$, with strict inequality whenever $q \neq q^\star$.

By Proposition 1, every inner product of this form is the same for any choice of reference profile. Hence whether $q^\star$ (or g) satisfies any of these three conditions is determined by the game $\mathcal{G}$ and the strategy profile alone, not by the choice of $(\bar\sigma, \bar\alpha)$ in Definition 4.

E Missing Material of Section 4

E.1 Proof of Theorem 5

In this section we prove Theorem 5.

Theorem 5. *For every honest-signaling game, the Riley assessment $(q^\star, \mu^\star)$ is the unique* PE. *Moreover, $q^\star$ is the unique* EFCCE.

In Lemma 6 we show that the Riley outcome is supported on a PE.

Lemma 6. *Under Assumptions 1 and 2, the Riley outcome $q^\star$ is a* PE *strategy.*

Proof. We construct the belief system $\mu^\star$ associated with $q^\star$ via Bayes' rule and show that the assessment $(q^\star, \mu^\star)$ satisfies the conditions of Definition 1 (consistency of beliefs and sequential rationality for both players), so that $q^\star \in \mathcal{S}^{\mathrm{PE}}$.

Every signal $s_k \in \Sigma$ satisfies $\sum_{\theta\in\Theta} \pi(\theta) q_S^\star(s_k \mid \theta) = \pi(\theta_k) > 0$. Hence $q^\star \in X^\circ(\mathcal{G})$ and, by (1), it must be that

$$\mu^\star(\theta \mid s_k) = \frac{\pi(\theta)\, q_S^\star(s_k \mid \theta)}{\sum_{\theta'} \pi(\theta')\, q_S^\star(s_k \mid \theta')} = \frac{\pi(\theta)\, \mathbf{1}\{\theta = \theta_k\}}{\pi(\theta_k)} = \mathbf{1}\{\theta = \theta_k\}, \tag{12}$$

which holds by construction. Hence, consistency of beliefs follows.

We now show sequential rationality per player. For the receiver, we have that $\mathrm{supp}(q_R^\star(\cdot \mid s_k)) = \{a_k\}$ holds for every $k \in \{1, \dots, K\}$. We need to show that

$$a_k \in \arg\max_{a\in A} \sum_{\theta\in\Theta} \mu^\star(\theta \mid s_k) u_R(\theta, s_k, a)\, q_R^\star(a \mid s). \tag{13}$$

Fix $k \in \{1, \dots, K\}$. For every $j \in \{1, \dots, K\} \setminus \{k\}$,

$$\sum_{i=1}^K \mu^\star(\theta_i \mid s_k)\, u_R(\theta_i, s_k, a_k) = \sum_{i=1}^K \mathbf{1}[\theta_i = \theta_k]\, u_R(\theta_i, s_k, a_k) = u_R(\theta_k, s_k, a_k)$$
$$> u_R(\theta_k, s_k, a_j) = \sum_{i=1}^K \mathbf{1}[\theta_i = \theta_k]\, u_R(\theta_i, s_k, a_j) = \sum_{i=1}^K \mu^\star(\theta_i \mid s_k)\, u_R(\theta_i, s_k, a_j).$$

The first and last equalities are by (12) and the strict inequality is Assumption 1. Hence (13) holds.

For the sender, $\mathrm{supp}(q_S^\star(\cdot \mid \theta_k)) = \{s_k\}$ for every $k \in \{1, \dots, K\}$. We need to show that

$$s_k \in \arg\max_{s\in\Sigma} \sum_{a\in A} u_S(\theta_k, s, a)\, q_R^\star(a \mid s). \tag{14}$$

Fix $k \in \{1, \dots, K\}$. For every $j \in \{1, \dots, K\} \setminus \{k\}$, we have that

$$\sum_{a\in A} u_S(\theta_k, s_k, a)\, q_R^\star(a \mid s_k) = \sum_{a\in A} u_S(\theta_k, s_k, a)\, \mathbf{1}[a = a_k] = u_S(\theta_k, s_k, a_k)$$
$$\ge \min_{a\in A} u_S(\theta_k, s_k, a) > \max_{a\in A} u_S(\theta_k, s_j, a) \ge u_S(\theta_k, s_j, a_j)$$
$$= \sum_{a\in A} u_S(\theta_k, s_j, a)\, \mathbf{1}[a = a_j] = \sum_{a\in A} u_S(\theta_k, s_j, a)\, q_R^\star(a \mid s_j).$$

Here, the first and last equalities are by the definition of $q_R^\star$. Finally, the strict inequality is Assumption 2. (14) follows and this concludes the proof. □

We now show that the Riley assessment $(q^\star, \mu^\star)$ is the unique PE. In fact we prove a stronger statement in Lemma 7: $(q^\star, \mu^\star)$ is the unique EFCCE of $\mathcal{G}$.[9]

Lemma 7. *For every signaling game $\mathcal{G}$ satisfying Assumptions 1 and 2, the Riley assessment $(q^\star, \mu^\star)$ is the unique* EFCCE *in $X(\mathcal{G})$.*

Proof. Let $Q \in \Delta(\Sigma^\Theta \times A^\Sigma)$ be any EFCCE. We show that $Q = (q_S^\star, q_R^\star)$ (see (3)). Fix $k \in \{1, \dots, K\}$. By invoking (4) with $\theta = \theta_k$ and $s' = s_k$, we obtain

$$\mathbb{E}_{(\sigma,\alpha)\sim Q}\big[u_S(\theta_k, \sigma(\theta_k), \alpha(\sigma(\theta_k)))\big] \;\geq\; \mathbb{E}_{(\sigma,\alpha)\sim Q}\big[u_S(\theta_k, s_k, \alpha(s_k))\big]. \tag{15}$$

For every realization (σ, α), Assumption 2 implies $u_S(\theta_k, s_k, \alpha(s_k)) \geq u_S(\theta_k, \sigma(\theta_k), \alpha(\sigma(\theta_k)))$, with strict inequality whenever $\sigma(\theta_k) \neq s_k$. Indeed, the two sides coincide if $\sigma(\theta_k) = s_k$, while if $\sigma(\theta_k) = s_j$ for some $j \neq k$, then

$$u_S(\theta_k, s_k, \alpha(s_k)) \;\geq\; \min_{a\in A} u_S(\theta_k, s_k, a) \;>\; \max_{a\in A} u_S(\theta_k, s_j, a) \;\geq\; u_S(\theta_k, s_j, \alpha(s_j)).$$

Hence, if $\Pr_{(\sigma,\alpha)\sim Q}[\sigma(\theta_k) \neq s_k] > 0$, taking expectations gives the strict reverse of (15), a contradiction. We conclude that $\Pr_{(\sigma,\alpha)\sim Q}[\sigma(\theta_k) = s_k] = 1 = q_S^\star(s_k \mid \theta_k)$. Since k was arbitrary, every signal $s \in \Sigma$ is on-path. Therefore, by (12) we also have, for every $\theta \in \Theta$ and every $k \in \{1, \dots, K\}$ that:

$$\mu(\theta \mid s_k) = \frac{\pi(\theta)\,\Pr_{(\sigma,\alpha)\sim Q}[\sigma(\theta) = s_k]}{\sum_{\theta'\in\Theta} \pi(\theta')\,\Pr_{(\sigma,\alpha)\sim Q}[\sigma(\theta') = s_k]} = \frac{\pi(\theta)\,\mathbf{1}\{\theta = \theta_k\}}{\pi(\theta_k)} = \mathbf{1}\{\theta = \theta_k\}.$$

Fix $k \in \{1, \dots, K\}$. Substituting into (5), for every $a' \in A$,

$$\begin{aligned}
\mathbb{E}_{(\sigma,\alpha)\sim Q}\big[u_R(\theta_k, s_k, \alpha(s_k))\big] &= \sum_{\theta\in\Theta} \mu(\theta \mid s_k)\, \mathbb{E}_{(\sigma,\alpha)\sim Q}\big[u_R(\theta, s_k, \alpha(s_k)) \mid \sigma(\theta) = s_k\big] \\
&\geq \sum_{\theta\in\Theta} \mu(\theta \mid s_k)\, u_R(\theta, s_k, a') \\
&= u_R(\theta_k, s_k, a').
\end{aligned}$$

Setting $a' = a_k$ and applying Assumption 1 (which gives $u_R(\theta_k, s_k, a_k) > u_R(\theta_k, s_k, a)$ for every $a \neq a_k$) forces $\alpha(s_k) = a_k$ almost surely. Hence $Q = (q_S^\star, q_R^\star)$ and the proof follows. □

By combining Lemmata 6 and 7 the proof of Theorem 5 follows since every honest-signaling game satisfies Assumptions 1 and 2.

F Missing Material of Section 5

F.1 Treatment Variables and Specification

At $K = 2$ each player's strategy reduces to a 2-dimensional vector, so the joint state is

$$x = (p_L,\, p_H,\, r_L,\, r_H) \in [0, 1]^4,$$

where p_t is the probability that the type-t sender plays the high signal and r_m is the probability that the receiver plays the high action on signal m. *Local* runs start near the family's target equilibrium σ^*, with a per-coordinate Euclidean spread of $\rho \approx 0.07$; *global* runs start uniformly on the interior of the simplex. All four algorithms share the same (σ_0, ρ_0) at each of the $n = 50$ runs within an init. scheme, so cross-algorithm comparisons are paired. In multi-equilibrium families, σ^* refers to the family's fixed reference equilibrium (e.g. Riley outcome for the Spence case, pooling at the low signal for the multiple pooling case and the riley outcome for the multiple separating case).

Across all experiments, each seed independently draws the initial policies (σ_0, ρ_0) per row from $\mathrm{Dirichlet}(50\,\sigma^* + \mathbf{1})$ for LOCAL init (concentrated near the target) or $\mathrm{Dirichlet}(\mathbf{1})$ for GLOBAL init (uniform on the simplex), and seeds any algorithm-internal randomness. The reference profile σ_{ref} is the family's designated centroid — the unique equilibrium when one exists, and the fixed reference equilibrium above in multi-equilibrium families.

[9] Recall that by the hierarchy of solution concepts, $\mathcal{S}^{\mathrm{PE}} \subseteq \mathcal{S}^{\mathrm{EFCCE}}$. Therefore $q^\star \in$ EFCCE follows immediately from Lemma 6.

Table 4: Treatment variables in the preferred specification, using the hyperparameter values that delivered the strongest convergence performance in the robustness analysis

Variable	Value
Families	unique separating; unique pooling (m_0, m_1); pooling and separating; unique semi-separating; multiple pooling; multiple separating (Riley/anti)
Algorithms	PGA, OMA, Opt. OMA, PPO
Step size η	PGA 0.3; OMA / Opt. OMA 1.0; PPO 5×10^{-2}
Step-size Schedule	constant
Initialisation regime	local (near target, $\rho\approx0.07$); global (uniform on simplex)
Horizon T	2000 (classical); 240 rounds (PPO)
Runs per treatment	$n = 50$
Conv. Tolerance δ	10^{-2}

F.2 Game parameterisations

We describe the parameterizations of the different equilibrium classes.

Unique separating (honest-signaling) We define $u_S(\theta, s, a) = \alpha_S \cdot \mathbf{1}[s = \theta] + \varepsilon_S$ and $u_R(\theta, s, a) = \alpha_R \cdot \mathbf{1}[a = \theta] + \varepsilon_R$, where α_S, α_R are sender / receiver reward multiplier for matching behavior. Further, $\varepsilon = (\varepsilon_S, \varepsilon_R)$ are bounded perturbations. Note that for sufficiently high α_S, α_R relative to the perturbation budget, we obtain *sender dominance* (each type strictly prefers its own signal, regardless of the receiver) and a *receiver unique best response* (the matched action strictly beats any alternative on each signal).

Unique pooling Given the pooling signal s_*, we have $u_S(\theta, s_*, a) = \alpha_S$; for all other signals $s \neq s_*$, $u_S(\theta, s, a) = \beta_S < \alpha_S$ (a strictly worse outside option, independent of a). Receiver payoff is the matching premium $u_R(\theta, s, a) = \alpha_R \cdot \mathbf{1}[a = \theta]$ (reward iff the action matches the realised type). Under the on-path posterior on the pooled signal — which equals the prior π — the receiver's expected utility is the probability that the action and the type match. Hence, the best response is unique iff the prior is strictly non-uniform.

Pooling and separating (Spence-style Job-Market Signaling) We consider a discrete job-market signaling model of Spence [36]: the receiver pays a wage w_a to each action, and the sender of type θ pays signaling cost c_θ per unit of high signal. Concretely, for $K = 2$, $u_S(\theta, s, a) = w_a - c_\theta \cdot 1_{\{s=1\}}$, and the receiver earns $+\alpha_R$ when its wage matches the realised type and $-\alpha_R$ otherwise, i.e. $u_R(\theta, s, a) = \alpha_R \cdot (2\,\mathbf{1}[a = \theta] - 1)$. Single-crossing requires the marginal cost to fall in θ: $c_L > c_H$. The single-crossing condition $c_L > c_H$ ensures that a high-signal threshold separates the two types in equilibrium; a non-trivial wage gap $w_H > w_L$ ensures the receiver actually discriminates. With $c_H \leq w_H - w_L \leq c_L$ and $c_H > 0$, both a separating PE and a pooling PE survive.

Unique semi-separating (Kreps–Wilson reputation) We consider a discrete reputation game inspired by Kreps and Wilson [20]. The strategic-type sender chooses between Accommodate ($s = 0$) and Prey ($s = 1$); the commitment-type sender always plays Prey (we use "Prey" throughout, in line with the table below); the receiver chooses to STAY ($a = 0$) or EXIT ($a = 1$). Sender payoffs are parameterised by the deterrence-stay payoff D_1 on Accommodate, the prey-stay payoff P_1, the maximal payoff M_1 from successfully Preying when the receiver Exits, and the accommodate-exit payoff A_1; receiver payoffs by the rate of return $D_2 > 0$ when staying against a Sane sender and $P_2 < 0$ when staying against a Crazy one. The parameter $\delta > 0$ is the Crazy sender's strict-Prey margin (it makes Prey strictly dominant for the Crazy type; it is *not* a discount factor and does not enter receiver payoffs), see also the following table.

In the unique semi-separating PE each player's mixing probability is fixed by the *other* player's indifference condition. The Sane sender's mix probability q^* on Prey is pinned by the receiver's indifference between Stay and Exit on the on-path Prey signal: the posterior on Sane must equal $1 - \tau$, with $\tau \equiv D_2/(D_2 - P_2)$, which yields $q^* = (1 - p)(1 - \tau)/(\tau p)$; this lies in $(0, 1)$ iff $p > 1 - \tau$. The receiver's mix probability r^* on Stay (conditional on the Prey signal) is pinned by the

Table 5: Stage payoffs of the reputation signalling game per sender type; each cell reports $(u_S,\ u_R)$. The *Sane* sender (prior p) chooses freely between Accommodate and Prey; the *Crazy* sender (prior $1-p$) plays Prey by construction (it is strictly dominant by $\delta > 0$, with X the Crazy baseline payoff). The conditions $M_1 > D_1 > P_1$, $A_1 > P_1$, $D_2 > 0 > P_2$ and $pD_2 + (1-p)P_2 > 0$ with $p \in (0,1)$ jointly pin down the unique semi-separating PE.

	Sane sender (prob. p)		**Crazy sender (prob. $1-p$)**	
	STAY (a=0)	EXIT (a=1)	STAY (a=0)	EXIT (a=1)
Accommodate (s=0)	$(D_1,\ D_2)$	$(A_1,\ 0)$	$(X-\delta,\ P_2)$	$(X-\delta,\ 0)$
Prey (s=1)	$(P_1,\ D_2)$	$(M_1,\ 0)$	$(X,\ P_2)$	$(X,\ 0)$

Sane sender's indifference between Accommodate, after which the receiver Stays, and Prey, yielding $r^* = (M_1 - D_1)/(M_1 - P_1)$. Neither expression involves A_1; the condition $A_1 > P_1$ only rules out the pure pooling PE.

Multiple pooling Sender utility depends only on whether the receiver plays the prior-best-response action: $u_S(\theta, s, a) = \beta \cdot \mathbf{1}[a = a^*]$ with $a^* = \arg\max_a \sum_\theta \pi(\theta) u_R(\theta, s, a)$. Receiver matches signal-conditional type as in the pooling family. A sufficiently high β ensures the sender strictly prefers the receiver to play a^*, which makes both pure pooling profiles (pool at s_0, pool at s_1) potential candidates for equilibria. A sufficiently high value for α_R ensures the receiver's prior-BR is unique on every signal.

Multiple separating (Riley + anti-Riley) This class admits two pure-strategy separating PE at $K = 2$ — Riley ($\sigma^* = I$) and anti-Riley ($\sigma^* =$ swap).

K=2 instance (used in table 1). For each type, sender payoffs depend on whether the realised (s, a) pair lies on the Riley or anti-Riley path:

$$u_S(\theta_0, s_0, a_0) = R \qquad u_S(\theta_0, s_1, a_0) = A$$
$$u_S(\theta_0, s_1, a_1) = D_R \qquad u_S(\theta_0, s_0, a_1) = D_A,$$

with type θ_1 symmetric under the type/signal/action swap. The receiver matches: $u_R(\theta, s, a) = \alpha_R \cdot \mathbf{1}[a = \theta]$. The conditions $R > A > D_A > D_R \geq 0$ together with a strictly non-uniform prior imply that both pure separating profiles are *strict* PE: the Riley IC slack is $R - D_R$ and the anti-Riley IC slack is $A - D_A$. (Note that mixed-semi-separating equilibria may still survive in cases where the riley and anti-riley separating equilibrium exist.) To rule out pooling, we additionally require the condition:

$$R + D_R > A + D_A.$$

Finally, for completeness, we provide the used game parameterizations for all equilibrium classes, which we used in the simulations to generate the results that are illustrated in table 1.

Table 6: Exact numeric values used to instantiate every benchmark game in table (table 1).

Family	**Parameter**	**Value**
`unique_separating_honest_signaling`	prior (μ_0, μ_1)	$(0.5,\ 0.5)$
	$\alpha_S,\ \alpha_R$	$2.0,\ 2.0$
`unique_pooling`	pool signal s_*	0 (LOW) or 1 (HIGH)
	prior (μ_0, μ_1)	$(0.6,\ 0.4)$
	$\alpha_S,\ \beta_S$	$2.0,\ 0.5$
	α_R	2.0
`spence_signaling`	prior (μ_L, μ_H)	$(0.4,\ 0.6)$
	wages (w_L, w_H)	$(0.0,\ 2.0)$
	costs (c_L, c_H)	$(3.0,\ 1.0)$
	α_R	2.0

(continued on next page)

(continued from previous page)

Family	Parameter	Value
reputation_signaling	sane prior p	0.6
	D_1, P_1, M_1, A_1	2.0, 1.0, 4.0, 2.0
	D_2, P_2	1.0, -1.0
	δ (Crazy dominance margin)	1.0
multiple_pooling	prior (μ_0, μ_1)	(0.6, 0.4)
	$\beta,\ \alpha_R$	1.0, 1.0
multiple_separating	prior (μ_0, μ_1)	(0.6, 0.4)
	$R,\ A$ (Riley, anti payoffs)	4.0, 2.0
	$D_R,\ D_A$ (deviation payoffs)	0.0, 1.0
	α_R	1.0

For completeness, we provide the game parameterizations for all games in the 7 different classes. The game parameters used in Table 1 are stated in Table 6.

F.3 Algorithm details and hyperparameters

We specify the four algorithms and the hyperparameters used in every cell of table 1. All step sizes are constant ($\eta_S = \eta_R = \eta_0$, no decay schedule).

All four algorithms operate on the same simplex-valued strategies. The three classical learners (PGA, OMA, Opt. OMA) consume full-information feedback from a shared oracle. PPO is the only algorithm that does *not* use this oracle: it samples episodes from the joint distribution induced by $(\sigma, \rho, \text{prior})$, where σ, ρ describe the sender and the receiver (mixed) strategy respectively.

Feedback oracle. With $\sigma[\tau, m] = q_S(m \mid \tau)$ and $\rho[m, a] = q_R(a \mid m)$, the oracle returns

$$Q_S[\tau, m] = \sum_{a \in A} \rho[m, a]\, u_S(\tau, m, a), \qquad Q_R[m, a] = \sum_{\tau \in \Theta} \mu(\tau \mid m)\, u_R(\tau, m, a), \tag{16}$$

where μ is the posterior (1). On $X^\circ(\mathcal{G})$, Q_S and Q_R coincide with g_S and g_R in (6)–(7) up to an additive constant in each row, to which all three updates below are invariant. If a signal m has zero probability under (π, σ), the oracle sets $\mu(\cdot \mid m) := \pi$.

F.3.1 Projected Gradient Ascent (PGA)

Per-row Euclidean ascent followed by exact simplex projection [41]:

$$\sigma_{t+1}[\tau, :] = \Pi_{\Delta_K}\big(\sigma_t[\tau, :] + \eta_0\, Q_S[\tau, :]\big), \qquad \rho_{t+1}[m, :] = \Pi_{\Delta_K}\big(\rho_t[m, :] + \eta_0\, Q_R[m, :]\big).$$

Note that Q_S, Q_R are the full-information feedback gradients from 16. We use the step size: $\eta_0 = 0.3$.

F.3.2 Entropic Online Mirror Ascent (OMA)

Next, we have

$$\sigma_{t+1}[\tau, :] \propto \sigma_t[\tau, :]\, \exp\big(\eta_0\, Q_S[\tau, :]\big), \qquad \rho_{t+1}[m, :] \propto \rho_t[m, :]\, \exp\big(\eta_0\, Q_R[m, :]\big),$$

with row-wise normalisation. We subtract the row-wise maximum of $\eta_0\, Q$ before exponentiation for numerical stability. Iterates remain strictly in the relative interior of the simplex. Step size: $\eta_0 = 1.0$.

F.3.3 Optimistic Entropic OMA (Opt. OMA)

Two iterates per player as in Syrgkanis et al. [38]: a pivot $\bar{\sigma}_t$ and a played strategy σ_t. Per round, observe gradients at the played strategies, take an entropic step on the pivot, and an extra prediction step on top of the new pivot:

$$\bar{\sigma}_{t+1}[\tau, :] \propto \bar{\sigma}_t[\tau, :]\, \exp(\eta_0\, Q_S[\tau, :]), \qquad \sigma_{t+1}[\tau, :] \propto \bar{\sigma}_{t+1}[\tau, :]\, \exp(\eta_0\, Q_S[\tau, :]),$$

and analogously for the receiver. Step size: $\eta_0 = 1.0$.

F.3.4 Proximal Policy Optimisation (PPO)

Episode structure. PPO sees a *single-step episodic environment*: each episode is one play of the signaling game, $(t, m, a) \sim \text{prior} \cdot \sigma \cdot \rho$, and the per-agent reward is the realised one-shot payoff. Because each episode is exactly one transition, neither a discount factor γ nor a GAE bootstrap parameter λ_{GAE} applies: the per-episode return is the realised reward and the advantage is the one-step residual, for more details, see [33].

Tabular policies and values. Each agent has tabular logits $\theta \in \mathbb{R}^{K \times K}$ with $\pi(\cdot \mid s) = \text{softmax}(\theta[s, :])$ and a tabular value head $V \in \mathbb{R}^K$ over the same state set, which matches the function-approximation class to OMA's policy class.

Round. Each PPO round (i) collects one batch of $N_{\text{rollout}} = 512$ joint episodes per agent under the current policies, shared by both agents (ii) computes one-step advantages $A = r - V(s)$, normalised to zero mean and unit variance over the batch; (iii) runs $E = 4$ epochs of mini-batch ($M_{\text{batch}} = 128$) clipped-surrogate updates per agent with

$$\mathcal{L} = -\,\mathbb{E}\Big[\min\big(r_\theta\,\widehat{A},\ \text{clip}(r_\theta, 1{-}\epsilon, 1{+}\epsilon)\,\widehat{A}\big)\Big] + c_v\,\text{MSE}\big(V(s), r\big) - c_e\,H(\pi_\theta),$$

where $r_\theta = \pi_\theta(a \mid s)/\pi_{\theta_{\text{old}}}(a \mid s)$ is the importance-sampling ratio and $H(\pi_\theta)$ the per-state entropy. The two agents are updated independently on this shared batch.

Hyperparameters. Table 7 lists every PPO hyperparameter used in the main cross-tab.

Table 7: PPO hyperparameters used in every PPO cell of table 1. Discount factor γ and GAE parameter λ_{GAE} are not applicable because each episode is a single transition; the advantage is the one-step residual $r - V(s)$.

Symbol	Value	Description
N_{rollout}	512	Episodes per round (per agent)
E	4	Optimisation epochs per round
M_{batch}	128	Mini-batch size
ϵ	0.2	Clip ratio (clipped surrogate)
c_v	0.5	Value-loss coefficient
c_e	0.01	Entropy-bonus coefficient
η_0	5×10^{-2}	Adam learning rate
$g_{\max}$	0.5	Global gradient-norm clip
Optimiser	Adam	PyTorch defaults: $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon_{\text{Adam}} = 10^{-8}$
Advantage normalisation	on	per-batch zero-mean, unit-variance
T_{PPO}	240 rounds	main cross-tab horizon

G Standard deviations

The table table 1 reports point estimates over $n = 50$ paired-init seeds per cell. Table 8 reports the matching across-seed sample standard deviations for each metric. $\text{sd}(\text{CONV.})$ is the sample standard deviation of the binary convergence indicator (so $\text{sd}(\text{CONV.}) \approx \sqrt{p(1-p)}$ for $\text{CONV.} = p$); $\text{sd}(\text{DIST.})$ is the sample standard deviation of the terminal distance over all 50 runs; and $\text{sd}(\text{HIT})$ is the sample standard deviation of the hitting time over the converged subset ("$-$" when no run converged or only one did).

The table 8 makes the qualitative interpretation precise. Cells with $\text{CONV.} \in \{0.000, 1.000\}$ in the main table 1 also have $\text{sd}(\text{CONV.}) = 0$ (every seed had the same outcome). Across-seed variability in CONV. is non-trivial in two locations: (i) the entire multi-pooling family ($\text{sd}(\text{CONV.}) \in [0.141, 0.505]$ across the eight algorithm$\times$init cells), which is the only family where convergence is genuinely seed-dependent for every algorithm; and ii) PPO under local initialisation on unique pooling at m_1 ($\text{sd}(\text{CONV.}) = 0.141$, consistent with 1 out of 50 seeds failing). The remaining 47 of the 56 algorithm$\times$init$\times$family cells in table 8 have $\text{sd}(\text{CONV.}) = 0$, so the qualitative reading "either every seed converges or every seed fails" holds outside those two locations. The $\text{sd}(\text{HIT})$ entries are heavy-tailed in multi-pooling (e.g. OMA local at 34.5 rounds and PPO local at 48.2, against medians of 11.5 and 29.5 respectively).

Table 8: Across-seed sample standard deviations companion to the main results table. Each cell aggregates $n = 50$ paired-init seeds. HIT standard deviations are computed over the converged subset only.

		Local initialization			Global initialization		
Eq. class	**Alg.**	**sd(CONV.)**	**sd(DIST.)**	**sd(HIT)**	**sd(CONV.)**	**sd(DIST.)**	**sd(HIT)**
Unique separating	PGA	0.000	0.000	0.0	0.000	0.000	1.1
	OMA	0.000	0.000	0.4	0.000	0.000	1.1
	Opt. OMA	0.000	0.000	0.0	0.000	0.000	1.0
	PPO	0.000	0.000	0.8	0.000	0.000	1.7
Unique pooling (m_0)	PGA	0.000	0.000	0.2	0.000	0.000	5.5
	OMA	0.000	0.000	1.9	0.000	0.000	6.0
	Opt. OMA	0.000	0.000	1.6	0.000	0.000	4.8
	PPO	0.000	0.000	3.4	0.000	0.000	8.0
Unique pooling (m_1)	PGA	0.000	0.000	0.2	0.000	0.000	5.4
	OMA	0.000	0.000	1.7	0.000	0.000	6.3
	Opt. OMA	0.000	0.000	1.4	0.000	0.000	5.5
	PPO	0.141	0.002	3.1	0.000	0.000	8.3
Pooling and separating	PGA	0.000	0.000	0.0	0.000	0.000	4.4
	OMA	0.000	0.000	0.6	0.000	0.000	3.3
	Opt. OMA	0.000	0.000	0.3	0.000	0.000	2.9
	PPO	0.000	0.000	1.0	0.000	0.000	7.7
Unique semi-separating	PGA	0.000	0.079	–	0.000	0.094	–
	OMA	0.000	0.198	–	0.000	0.234	–
	Opt. OMA	0.000	0.000	10.0	0.000	0.000	41.0
	PPO	0.000	0.203	–	0.000	0.220	–
Multiple pooling	PGA	0.501	0.026	1.5	0.388	0.205	7.8
	OMA	0.505	0.027	34.5	0.388	0.147	10.3
	Opt. OMA	0.505	0.027	8.3	0.370	0.152	22.1
	PPO	0.328	0.222	48.2	0.141	0.005	27.3
Multi-separating (Riley/anti)	PGA	0.000	0.000	0.0	0.000	0.000	3.4
	OMA	0.000	0.000	0.7	0.000	0.000	5.6
	Opt. OMA	0.000	0.000	0.4	0.000	0.000	4.8
	PPO	0.000	0.001	0.9	0.000	0.000	2.1

G.1 Experiments compute resources

The experiments/simulations are small-scale. In other words, no additional computational resources are required and the experiments are feasible on a local machine in a couple of hours (e.g. in our case, all experiments were run on a 14" Apple MacBook Pro (Nov. 2024) with an Apple M4 Pro processor and 48 GB of unified memory, running macOS Tahoe 26.3.1 (a).)

H Missing Material of Section 5.3

H.1 Four-type benchmark games

We extend the unique separating and unique semi-separating benchmarks to $K = 4$. Throughout this section, $\Theta = \{\theta_1, \dots, \theta_4\}$, $\Sigma = \{s_1, \dots, s_4\}$, and $A = \{a_1, \dots, a_4\}$. Here, we write $Q_S = (q_S(s_j \mid \theta_i))_{i,j=1}^4$ and $Q_R = (q_R(a_j \mid s_i))_{i,j=1}^4$ for the Sender and Receiver strategy matrices, respectively. The constructions below admit a unique PE among all behavioral strategy profiles. In both cases every signal is on-path at the equilibrium, so the equilibrium belief system is uniquely determined by Bayes' rule.

H.1.1 Unique separating (honest-signaling game)

We retain the matching rewards of the binary honest-signaling benchmark. For $i, j, \ell \in \{1, \dots, 4\}$, define

$$u_S(\theta_i, s_j, a_\ell) = \alpha_S \mathbf{1}\{j = i\} + \varepsilon_S(\theta_i, s_j, a_\ell), \tag{17}$$
$$u_R(\theta_i, s_j, a_\ell) = \alpha_R \mathbf{1}\{\ell = i\} + \varepsilon_R(\theta_i, s_j, a_\ell). \tag{18}$$

The perturbations are fixed payoff functions satisfying $|\varepsilon_S(\theta_i, s_j, a_\ell)| \le b_S$ and $|\varepsilon_R(\theta_i, s_j, a_\ell)| \le b_R$, where $b_S, b_R \ge 0$. We assume a full-support prior and

$$\alpha_S > 2b_S, \qquad \alpha_R > 2b_R. \tag{19}$$

Under these bounds, truthful signaling is strictly dominant. The Receiver uniquely matches the revealed type.

Proposition 2 (Unique separating PE)**.** *Under* (19)*, the honest-signaling game has a unique PE. It is the strict Riley PE, given by*

$$q_S^\star(s_j \mid \theta_i) = \mathbf{1}\{j = i\}, \qquad q_R^\star(a_\ell \mid s_j) = \mathbf{1}\{\ell = j\}, \qquad \mu^\star(\theta_i \mid s_j) = \mathbf{1}\{i = j\}. \tag{20}$$

Proof. Fix a type θ_i and a signal $s_j \neq s_i$. By (19),

$$\min_{a \in A} u_S(\theta_i, s_i, a) \ge \alpha_S - b_S > b_S \ge \max_{a \in A} u_S(\theta_i, s_j, a).$$

Hence, s_i is strictly dominant for type θ_i, and every PE must have $q_S(s_i \mid \theta_i) = 1$. The prior has full support, so every signal is on-path and Bayes' rule gives $\mu(\theta_i \mid s_j) = \mathbf{1}\{i = j\}$. At signal s_i, for every $a_\ell \neq a_i$,

$$u_R(\theta_i, s_i, a_i) - u_R(\theta_i, s_i, a_\ell) \ge \alpha_R - 2b_R > 0.$$

Therefore the Receiver must choose a_i with probability one. These strict inequalities also establish that the displayed assessment is a strict PE, completing both existence and uniqueness. □

Concrete parameterization. For the experiments, we take the uniform prior $\pi(\theta_i) = 1/4$ and the fixed utilities

$$u_S(\theta_i, s_j, a_\ell) = 2\,\mathbf{1}\{j = i\} u_R(\theta_i, s_j, a_\ell) = 2\,\mathbf{1}\{\ell = i\}. \tag{21}$$

This sets $\alpha_S = \alpha_R = 2$, and $b_S = b_R = 0$, so (19) holds.

H.1.2 Unique semi-separating (multi-threat reputation game)

We extend the binary reputation game of Kreps and Wilson [20] by introducing three distinct types. Type θ_1 is Sane and strategic, whereas θ_j, $j \in \{2, 3, 4\}$, is committed to the corresponding Prey signal s_j. Signal s_1 is Accommodate. The Receiver chooses STAY (labeled a_1) or one of three threat-specific EXIT actions (a_2, a_3, a_4).

Write $p = \pi(\theta_1)$ (assuming $p \in (0, 1)$) and $\pi_j = \pi(\theta_j) > 0$ for $j = 2, 3, 4$, with $p + \pi_2 + \pi_3 + \pi_4 = 1$. The Sane Sender's payoffs are

$$u_S(\theta_1, s, a) = \begin{cases} D_1, & s = s_1,\ a = a_1, \\ A_1, & s = s_1,\ a \neq a_1, \\ P_1, & s \neq s_1,\ a = a_1, \\ M_1, & s \neq s_1,\ a \neq a_1. \end{cases} \tag{22}$$

For each type $j \in \{2, 3, 4\}$, set

$$u_S(\theta_j, s, a) = X - \delta \mathbf{1}\{s \neq s_j\}, \qquad \delta > 0, \tag{23}$$

where $X \in \mathbb{R}$ is a baseline payoff. Thus, each type strictly prefers its designated Prey signal regardless of the Receiver's action. Receiver payoffs are independent of the signal and are specified in Table 9. The appropriate Exit action gives payoff zero, while an Exit action tailored to a different type incurs a penalty $\zeta > 0$.

We impose the same payoff ordering and prior condition as in the binary reputation benchmark, together with a strictly positive penalty for an inappropriate Exit response:

$$\begin{gathered} M_1 > D_1 > P_1, \qquad A_1 > P_1, \qquad D_2 > 0 > P_2, \\ \delta > 0, \qquad \zeta > 0, \qquad pD_2 + (1 - p)P_2 > 0. \end{gathered} \tag{24}$$

Table 9: Receiver payoffs in the four-type reputation game, for every $s \in \Sigma$. Action a_j is the Exit response tailored to type θ_j, $j \in \{2,3,4\}$.

Type	STAY (a_1)	EXIT (a_2)	EXIT (a_3)	EXIT (a_4)
θ_1 (Sane)	D_2	0	0	0
θ_2	P_2	0	$-\zeta$	$-\zeta$
θ_3	P_2	$-\zeta$	0	$-\zeta$
θ_4	P_2	$-\zeta$	$-\zeta$	0

Define

$$b_j := \frac{-\pi_j P_2}{pD_2} \quad (j = 2,3,4), \qquad b := \sum_{j=2}^{4} b_j, \qquad r^\star := \frac{M_1 - D_1}{M_1 - P_1}. \tag{25}$$

The conditions in (24) imply $b_j > 0$, $b < 1$, and $r^\star \in (0,1)$.

Proposition 3 (Unique semi-separating PE)**.** *Under* (24)*, the four-type reputation game has a unique PE. Its strategy matrices are*

$$Q_S^\star = \begin{pmatrix} 1-b & b_2 & b_3 & b_4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad Q_R^\star = \begin{pmatrix} 1 & 0 & 0 & 0 \\ r^\star & 1-r^\star & 0 & 0 \\ r^\star & 0 & 1-r^\star & 0 \\ r^\star & 0 & 0 & 1-r^\star \end{pmatrix}. \tag{26}$$

The equilibrium is semi-separating: the Sane type mixes over all four signals, and each Prey signal is shared with its associated type.

Proof. We first derive the strategy profile of any PE. By (23), each type θ_j, $j \geq 2$, must play s_j with probability one. Hence every Prey signal is on-path. Write $x_j = q_S(s_j \mid \theta_1)$. Bayes' rule gives

$$\mu(\theta_1 \mid s_j) = \frac{px_j}{px_j + \pi_j}, \qquad \mu(\theta_j \mid s_j) = \frac{\pi_j}{px_j + \pi_j}.$$

All other types have posterior probability zero. At s_j, the appropriate Exit action a_j gives expected payoff zero. Each other Exit action has strictly negative expected payoff $-\zeta\pi_j/(px_j + \pi_j)$, while the expected payoff from Stay is

$$\frac{px_j D_2 + \pi_j P_2}{px_j + \pi_j}. \tag{27}$$

Consequently, Stay is uniquely optimal if $x_j > b_j$, the appropriate Exit is uniquely optimal if $x_j < b_j$, and precisely these two actions are best responses if $x_j = b_j$.

Suppose $x_j > b_j$ for some j. Then $x_j > 0$ and the Receiver must Stay at s_j, giving the Sane Sender payoff P_1. However, deviating to Accommodate guarantees at least $\min\{D_1, A_1\} > P_1$, independently of the Receiver's response there. This contradicts Sender optimality. Therefore $x_j \leq b_j$ for every j. Since $b < 1$, it follows that

$$q_S(s_1 \mid \theta_1) = 1 - \sum_{j=2}^{4} x_j \geq 1 - b > 0.$$

Only the Sane type sends s_1, so its posterior probability there is one and the Receiver uniquely chooses Stay. Hence the Sane Sender's equilibrium payoff is D_1.

If $x_j < b_j$ for any j, the appropriate Exit action is uniquely optimal at s_j, and the Sane Sender can obtain $M_1 > D_1$ by choosing that signal. Thus $x_j = b_j$ for every j. Each Prey signal is therefore used with positive probability by the Sane type. Its indifference between Accommodate and s_j requires

$$r_j P_1 + (1 - r_j) M_1 = D_1, \qquad r_j := q_R(a_1 \mid s_j),$$

which uniquely gives $r_j = r^\star$. No other Exit action can receive positive probability, so every PE must have the strategy matrices in (26).

To verify existence, equip this profile with its Bayes beliefs. At s_1 the Receiver uniquely prefers Stay. At each s_j, $j \geq 2$,

$$\mu^\star(\theta_1 \mid s_j) = \frac{-P_2}{D_2 - P_2}, \qquad \mu^\star(\theta_j \mid s_j) = \frac{D_2}{D_2 - P_2}, \tag{28}$$

and every other type has posterior probability zero. Stay and a_j both give expected payoff zero, while the other actions give strictly negative payoffs. The Sane Sender obtains D_1 from all four signals, and each commitment type strictly prefers its own signal. Sequential rationality therefore holds. Finally, all signals are on-path, so the belief system is unique as well. □

Concrete parameterization. We retain the binary reputation benchmark's numerical payoffs and split its commitment-type prior equally among the three commitment types:

$$\begin{gathered}(\pi(\theta_1), \pi(\theta_2), \pi(\theta_3), \pi(\theta_4)) = \left(\frac{3}{5}, \frac{2}{15}, \frac{2}{15}, \frac{2}{15}\right), \\ (D_1, P_1, M_1, A_1) = (2, 1, 4, 2), \qquad (D_2, P_2) = (1, -1), \qquad X = \delta = \zeta = 1.\end{gathered} \tag{29}$$

The prior condition holds since $pD_2 + (1-p)P_2 = 1/5 > 0$. Here $b_j = 2/9$, $b = 2/3$, and $r^\star = 2/3$, so the exact target matrices are

$$Q_S^\star = \begin{pmatrix} \frac{1}{3} & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad Q_R^\star = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \end{pmatrix}. \tag{30}$$

After Accommodate, the Sane posterior is one; after each Prey signal, the posterior assigns probability $1/2$ to the Sane type and $1/2$ to the associated commitment type. The signal probabilities are $(1/5, 4/15, 4/15, 4/15)$, and every Receiver action occurs with positive probability. The displayed target profile is used for local initialization and distance-to-equilibrium evaluation.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We justify the theoretical and empirical results in the main sections. Missing proofs or experimental details are included in the appendix.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: This is included in section 1.1.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation, while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Yes, for every theory section, we have an additional section in the appendix, where we include missing (full) proofs of all statements. In every section, we reference the corresponding section in the appendix that includes the missing proofs. We state all needed assumptions/definitions.

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We state the results in 5 and missing details in F. We also include the source code.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Yes, we provide code to replicate the results. Our simulations do not depend on any data source. Therefore, the latter does not apply to our paper.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide details on the hyperparameters of the algorithms and the studied signaling games parameterizations in the appendix F.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We discuss that in section G.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification:We discuss this in G.1.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper focuses on theoretical results and does not rely on any data sources or human subjects. We are not aware of any foreseeable harmful consequences, and the work conforms to the NeurIPS Code of Ethics in all respects.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: We do not identify direct societal impacts to discuss.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Not applicable

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: Not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: Not applicable

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Not applicable

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: Not applicable

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: Our use of LLMs was limited to standard code assistance and polishing and does not require disclosure under the NeurIPS policy.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.