Beyond Bayesian Nash: Learning Minimax-Regret Equilibria for Adversarial Team Games under Asymmetric Information

Anonymous authors
Paper under double-blind review

Abstract

Adversarial team games (ATGs) with asymmetric information, such as adversarial pathfinding, goal search, and reachability games on graphs, require strategies that are robust to hidden opponent types, such as a hidden goal flag, and to deception. Under asymmetric information, deception is seen as strategic shifts in the type distribution such that the omniscient opponent can collude with Nature and condition its play on the observed type. Existing risk-neutral solution concepts, such as Bayesian Nash equilibrium (BNE), are sensitive to distribution shifts, while distributionally robust approaches provide guarantees only within a prescribed ambiguity set. To address these limitations, we introduce Probabilistically Robust Minimax-Regret Equilibrium (PR-MRE), a novel equilibrium concept that combines the distribution-free robustness of minimax-regret reasoning with probabilistic information from a nominal type distribution. PR-MRE minimizes worst-case regret over a high-confidence subset of the type space, providing protection against strategic redistribution of probability mass while avoiding the conservatism of fully distribution-free approaches. We show that, for normal-form Bayesian games, PR-MRE can be formulated as a robust bilinear program and derive a tractable semidefinite relaxation. We then adapt this relaxation into a novel meta-solver within a robust double-oracle framework, PRMRE-PSRO, enabling population-based learning of approximate PR-MRE strategies via deep reinforcement learning best responses. Experiments on graph-structured adversarial team games demonstrate that PR-MRE discovers strategies with substantially improved worst-case performance across hidden types compared to risk-neutral equilibrium solutions, resulting in more robust behavior under strategic distribution shifts.

1 Introduction

Computation of equilibria in large imperfect-information games has been a central driver of recent advances in artificial intelligence. A key challenge in such settings is determining how strategic agents should reason about uncertainty. Existing equilibrium concepts differ fundamentally in their treatment of risk and robustness, ranging from risk-neutral formulations that optimize expected performance to risk-averse approaches that guard against uncertainty through worst-case reasoning. These differing notions of rationality lead to markedly different behaviors and robustness guarantees. Consequently, the literature on equilibria and equilibrium refinements for imperfect-information games is extensive (see Table 1 for an overview). In Bayesian games, a Nature player moves before play begins and assigns each player a type drawn from a joint type distribution. Harsanyi's seminal framework Harsanyi (1967; 1968) models strategic interactions with incomplete information by introducing the Bayesian Nash Equilibrium (BNE), in which each player maximizes expected utility under the common prior distribution over types. While BNE provides a principled solution concept for games with hidden information, it relies critically on the common prior assumption: although players do not observe each other's private types, the underlying type distribution is assumed to be common knowledge. This assumption can be restrictive in real-world strategic interactions involving asymmetric information and deception. In many settings, the type distribution itself may be uncertain, evolve over time, or even be influenced by an adversary with an information advantage. We refer to such changes as *strategic distribution shifts*. Unlike classical distributionally robust optimization (DRO), which

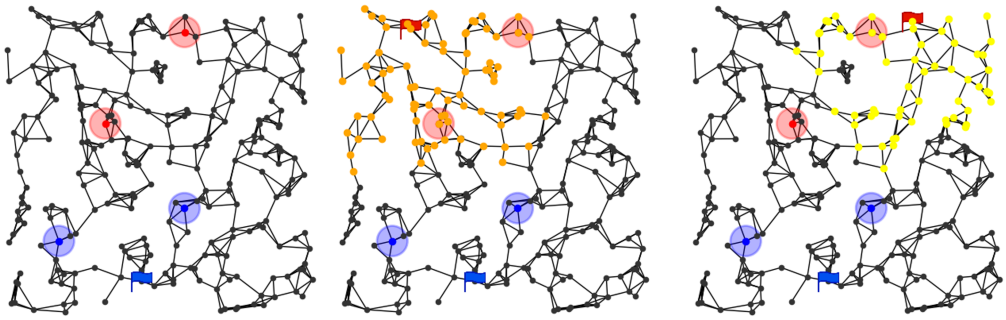

Figure 1: Graph-Structured Capture-the-Flag (Graph CtF) with Blue Team's Imperfect Information about the Red Flag Location: 2v2 CtF on a graph topology with Flag Uncertainty for Blue Team seen as two potential Red Flag locations. Limited Knowledge: Graph assumed known, but limited flag visibility (orange (yellow) indicates nodes on Left (Right) Flag are visible). Asymmetric: Blue agents cannot distinguish between Red flag locations at other nodes, whereas Red agents have perfect information about Blue flag.

models uncertainty through structured ambiguity sets such as KL-divergence, Wasserstein, or f-divergence balls around a nominal distribution, strategic distribution shifts need not admit a local divergence-based characterization. They may instead arise from regime changes, task-level shifts, or adaptive opponents that fundamentally alter the distribution of types.

Figure 1 provides a concrete example of such strategic distribution shifts in an adversarial search game. The Blue team must locate and capture a hidden Red flag before a timeout, while the Red team knows the true flag location and can defend accordingly. Blue only has access to a prior belief over possible flag locations (e.g., 60% right and 40% left), making the flag location a hidden type known only to Red. Consequently, a strategy that is optimized for a nominal prior may become vulnerable when the underlying type distribution changes or when Red exploits Blue's prior assumptions.

We consider imperfect-information games with deception, specifically asymmetric-information games in which the type distribution may undergo strategic distribution shifts and the informed opponent can condition its play on the realized type. The Capture-the-Flag example illustrates a broader challenge: behaviors derived from risk-neutral solution concepts such as Bayesian Nash Equilibrium (BNE) can be highly exploitable because they rely on a nominal type distribution. This motivates the central question addressed in this work: *What is an appropriate notion of robust best response for the information-disadvantaged team, and the resulting equilibrium concept, that provides favorable worst-case performance guarantees under strategic distribution shifts?*

Distributionally robust optimization (DRO) protects against uncertainty through ambiguity sets around a nominal distribution. While effective for modeling local perturbations, DRO-based equilibrium concepts provide guarantees only within the specified ambiguity set and may fail under larger regime changes. For example, in a three-flag version of the search-and-capture game, a nominal type distribution of $(0.6, 0.1, 0.3)$ may shift to $(0.0, 0.05, 0.95)$, fundamentally altering the strategic landscape. Such strategic distribution shifts are not naturally captured by local divergence-based uncertainty models. At the other extreme, a fully worst-case robust approach reasons against an adversarial type distribution over the entire type simplex, but this can be overly conservative because it effectively concentrates on the most difficult type regardless of its likelihood.

To bridge this gap, we consider strategic distribution shifts that preserve *type typicality*: types that are rare under the nominal distribution remain rare under admissible shifts, while probability mass may be redistributed arbitrarily among typical types. We then adopt a regret-based notion of robustness. Regret, or exploitability, measures the performance loss incurred by acting under imperfect information relative to a type-conditioned optimum. A Minimax-Regret Equilibrium (MRE) minimizes worst-case regret across all types and yields a distribution-free certificate in the form of a uniform lower bound on performance under

Table 1: Taxonomy of robust Blue best-responses (BRs) according to the optimization objective and the amount of information trusted about Nature's nominal. Each robust Blue BR coupled with the informed Red BR leads to a unique equilibrium concept (see Section 2.1 for details) and an associated robustness property (see Table 2) for the adversarial game with asymmetric information.

	Trust in Nature's Nominal Distribution: High → Low			
Objective \ **Trust**	**Full**	**Local**	**Coarse**	**None**
Payoff	Bayesian BR (BNE) Nominal Prior	DRO-Ball Local Perturbations	DRO-Structured Typicality-preserving shifts (Structured type ambiguity)	Worst-Case BR Entire type-simplex (Adversarial Nature)
Regret	–	–	**PR-MRE** Typicality-preserving shifts (Structured type ambiguity)	**MRE** Entire type-space

arbitrary distribution shifts. While robust, MRE can be conservative because it ignores probabilistic information about type occurrence. We therefore introduce the Probabilistically Robust Minimax-Regret Equilibrium (PR-MRE), which incorporates coarse probabilistic information through the typicality-preserving threat model. PR-MRE interpolates between risk-neutral and worst-case approaches, yielding tighter performance guarantees while retaining robustness to strategic distribution shifts.

Computing MRE and PR-MRE in general adversarial team games remains an open challenge. Existing solution methods for ATGs Anagnostides et al. (2026); Kalogiannis et al. (2024); Xiao et al. (2026); Zhang et al. (2024); Celli & Gatti (2018); Carminati et al. (2022b); Anagnostides et al. (2023) and extensive-form games (EFGs), such as Counterfactual Regret Minimization (CFR) Zinkevich et al. (2007), are designed to compute Nash equilibria and do not naturally extend to regret-based robustness objectives under structured type uncertainty. Moreover, MRE was originally introduced in the context of finite mechanism-design settings Hyafil & Boutilier (2004); Renou & Schlag (2010; 2011), and existing formulations do not readily generalize to the large strategy spaces such as team treeplexes encountered in adversarial team games. To address this challenge, we propose PR-MRE PSRO, a robust double-oracle framework for computing approximate MRE and PR-MRE strategies via metagame approximation. We show that the resulting metagame problems can be formulated as bilinear and robust bilinear programs, respectively, and derive tractable semidefinite relaxations that serve as novel meta-solvers within a population-based learning framework. Combined with deep reinforcement learning best-response oracles, the proposed approach enables scalable computation of robust strategies in graph-structured adversarial team games.

From Table 1, the optimization objective (payoff vs. regret) and the degree of trust placed in Nature's nominal type distribution define a two-dimensional design space of robust equilibrium concepts, with PR-MRE occupying the previously unexplored regime of regret-based best-response under coarse probabilistic information from the nominal. We formally state our contributions as follows,

1. **New equilibrium concept, PR-MRE**: We introduce a novel robust equilibrium concept for asymmetric information games, Probabilistically-Robust Minimax-Regret equilibrium (PR-MRE) that interpolates between risk-neutral, distributionally-robust equilibria and the distribution-free Minimax-Regret equilibrium (MRE). In contrast to MRE that treats all types equally likely, PR-MRE utilizes coarse probabilistic information to minimize worst-case exploitability across subsets of the type-space weighted by known information about their probability mass under admissible distribution shifts.

2. **Typicality-preserving threat model and robustness guarantees**: We show that the proposed PR-MRE best-response for the information-disadvantaged team provides a robustness certificate in the form of the tightest performance lower bound under strategic distribution shifts that redistribute probability mass arbitrarily among high-confidence subsets of the type-space.

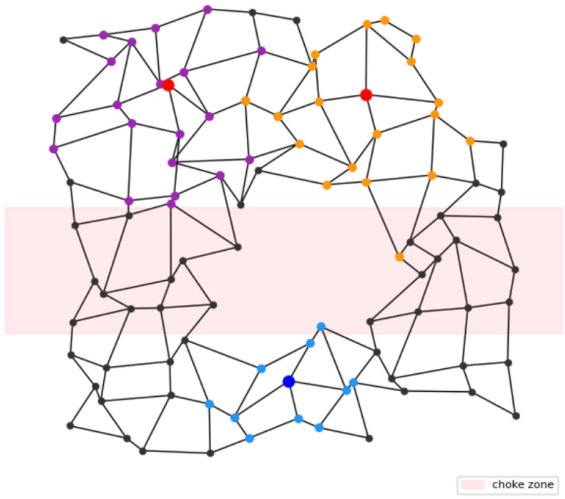

Figure 2: Graph Capture-the-Flag with two Red flag hypotheses – θ_0 (left) and θ_1 (right) (marked as the two red nodes in the top basin), the shaded pink zone represents the choke zone corresponding to the two corridors, and the purple (orange) nodes represent the information frontier or the nodes from which the left (right) flag hypothesis is visible. Red team agents initialize in the top-basin whereas the Blue team agents initialize in the bottom-basin below the choke zone. The Blue team agents cannot distinguish between the two flag hypotheses before they reach the information frontier while passing through the choke zone. See Figure 9 for a behavioral analysis of the learned Blue team policies via the proposed Probabilistically-Robust Minimax-Regret, PR-MRE meta-equilibria versus the Bayesian Nash equilibrium, BNE for a nominal type distribution $(\mu(\theta_0), \mu(\theta_1)) = (0.8, 0.2)$.

3. **Scalable approximation algorithm, PRMRE-PSRO**: We propose a metagame approximation to compute MRE and PR-MRE in adversarial team games under asymmetric information. We show that the metagame problems corresponding to MRE and PR-MRE can be formulated as bilinear and robust bilinear programs respectively. We provide a tractable SDP relaxation that is incorporated as a robust meta-equilibria within a double-oracle framework, PRMRE PSRO to compute robust team policies for graph-structured ATGs under asymmetric information.

4. **Empirical validation in graph-structured adversarial team games**: We demonstrate empirically that PRMRE PSRO discovers behaviors that are substantially more robust to distribution shifts than risk-neutral equilibrium solutions. In particular, the learned policies exhibit scouting behavior prior to commitment, reducing exploitability across competing flag hypotheses rather than overfitting to the dominant hypothesis under the nominal distribution.

2 Graph-Structured Adversarial Team Games under Asymmetric Information

We study graph-structured adversarial team games under asymmetric information, instantiated as Graph Capture-the-Flag (Graph CtF). Two teams of agents move on an undirected graph over a finite horizon and optimize opposing objectives – Blue team wins if the Blue agents capture the Red flag, and Red team wins if they successfully defend their flag for the time-horizon. The key asymmetry is that the Red team observes the true flag location, while the Blue team acts under uncertainty over a finite set of flag hypotheses.

Formally, the game is defined by the tuple $\mathscr{G} = \langle \mathcal{G}, n_r, n_b, \Theta, T \rangle$, where $\mathcal{G} = \langle \nu, \mathcal{E} \rangle$ is an undirected graph with node set ν, n_r and n_b are the number of Red and Blue agents, $\Theta \subseteq \nu$ is the set of possible flag locations, and T is the time horizon. The hidden type is sampled as $\theta \sim \mu \in \Delta(\Theta)$. The game proceeds sequentially over T steps. At each time step, agents select actions based on their available information and move on the graph according to the transition dynamics. The Red team observes the realized type θ, whereas the Blue team does not, and must act based only on its observation history. Accordingly, we consider policy

Table 2: Robustness properties associated with the various Blue best-responses (BRs) corresponding to different optimization objectives and trust levels on Nature (see Table 1 for taxonomy on robust Blue BRs).

	High ⟶ Trust in Nature's Nominal Distribution ⟶ Low			
Objective \ **Trust**	**Full**	**Local**	**Coarse**	**None**
Payoff	No Robustness	Worst-case payoff certificate Local perturbations	Worst-case payoff certificate Typicality-preserving shifts (Structured type ambiguity)	Worst-case payoff certificate Entire type-simplex
Regret	–	–	**Worst-case regret certificate** ⇔ **ULB** (Lemma 4) Typicality-preserving shifts (Structured type ambiguity)	**Worst-case regret certificate** ⇔ **ULB** (Lemma 1) Entire type-space

classes Π_r and Π_b, where Red policies may condition on θ and history, while Blue policies depend only on the Blue team's observation history. Let $V(\pi_r, \pi_b; \theta)$ denote the expected return of the Blue team under policies $\pi_r \in \Pi_r$ and $\pi_b \in \Pi_b$ when the underlying type is θ.

The nominal type distribution $\mu(\cdot)$ induces a standard Bayesian objective. In deceptive settings, however, the true distribution over types may differ substantially from $\mu(\cdot)$, rendering policies optimized for the nominal distribution highly exploitable. This motivates the search for equilibrium concepts and associated best-response notions that remain robust under strategic distribution shifts.

2.1 Equilibrium Concepts and Robustness Properties

In this section, we present a unified treatment of various equilibrium concepts in the specific context of adversarial team games and analyse their robustness properties with respect to shifts in the type distribution. For the case of asymmetric information, the Red team has perfect knowledge of Nature's hidden type (flag location) and can condition its play on the observed type.

Definition 2.1. *The Red team best-response operator* $\mathrm{BR}^{(\mathrm{r})}(\pi_b; \theta) : \Pi_\mathrm{b} \times \Theta \to \Pi_\mathrm{r}$ *to the Blue team strategy* $\pi_b \in \Pi_\mathrm{b}$ *and observed type* θ *is given as* $\mathrm{BR}^{(\mathrm{r})}(\pi_b; \theta) = \operatorname*{arginf}_{\pi \in \Pi_\mathrm{r}} V(\pi(\cdot), \pi_b; \theta)$.

Thus, for a given Blue team strategy, the Red team computes a best-response that minimizes the expected Blue team return conditioned on the observed type. How the Blue team defines as 'rational' play under imperfect information leads to a unique equilibrium concept and an associated robustness property. We now define various best-response operators for the Blue team, each of which, when coupled with the Red team best-response operator stated in Definition 2.1 defines a unique equilibrium.

2.1.1 Risk-Neutral Methods

Definition 2.2. *(Bayesian Best-Response Operator under Asymmetric Information) The Blue team Bayesian best-response operator* $\mathrm{BR}^{(\mathrm{b})}(\pi_r(\cdot); p, \Theta) : \Pi_\mathrm{r} \times \Delta(|\Theta|) \to \Pi_\mathrm{b}$ *given Red team strategy* $\pi_r(\cdot)$ *and the nominal type distribution* $p \in \Delta(|\Theta|)$ *is given as follows,*

$$\mathrm{BR}^{(\mathrm{b})}(\pi_r(\cdot); p, \Theta) = \operatorname*{argsup}_{\pi \in \Pi_\mathrm{b}} \mathbb{E}_{\theta \sim p}[V(\pi_r(\cdot), \pi; \theta)]. \tag{1}$$

The Bayesian best-response operator maximizes expected team performance under a nominal type distribution leading to the risk-neutral equilibrium concept of a Bayesian Nash equilibrium.

Definition 2.3. *(Bayesian Nash Equilibrium, BNE) The ex-ante Bayesian Nash equilibrium* $(\pi^*_{b,\text{BNE}}(p), \pi^*_r)$ *for a nominal type distribution* p *is defined as the fixed point of the best-response operators defined in Definition 2.1 and Definition 2.2 as the following,*

$$\pi^*_{b,\text{BNE}}(p) \in \text{BR}^{(\text{b})}(\pi^*_r(\cdot); p, \Theta) \text{ and } \pi^*_r(\cdot) \in \text{BR}^{(\text{r})}(\pi^*_{b,\text{BNE}}(p); \cdot).$$

The BNE is an ex-ante equilibrium concept in which players commit to strategies before play commences and optimize expected utility under the nominal type distribution. We focus on this setting because it aligns naturally with population-based training methods in which meta-equilibria are computed over sets of competing policies. For ex-post notions that incorporate sequential rationality and belief consistency, we refer the reader to refinements such as the perfect Bayesian Nash equilibrium (PBNE) Ouyang et al. (2016); Konicki et al. (2026).

From Definition 2.3, the BNE can equivalently be interpreted as a solution to the following saddle-point equation,

$$\mathbb{E}_{\theta\sim p}\left[V(\pi^*_r(\cdot), \pi_b; \theta)\right] \leq \mathbb{E}_{\theta\sim p}\left[V(\pi^*_r(\cdot), \pi^*_{b,\text{BNE}}(p); \theta)\right] \leq \mathbb{E}_{\theta\sim p}\left[V(\pi_r(\cdot), \pi^*_{b,\text{BNE}}(p); \theta)\right]. \tag{2}$$

At the saddle point given by the strategy-tuple $(\pi^*_b, \pi^*_r(\cdot))$ (the BNE), no team has any incentive to deviate given the opponent team strategy. An exact computation of the saddle point eq. (2) for the general formulation where strategy spaces Π_r and Π_b are sequence-form polytopes (or team treeplexes) is intractable. Approaches in the MARL community to find tractable approximations to the equilibria of high-dimensional games such as the graph-structured adversarial team game introduced in Section 2 broadly consists of DO-based Carminati et al. (2022a); McAleer et al. (2023a), CFR-based Chakrabarti et al. (2024) and PPO-based Rudolph et al. (2026). Risk-neutral equilibria, such as BNE (2), do not take into account potential strategic shifts in the type distribution, which motivates the following class of distributionally-robust and risk-averse methods to manage risk posed by ambiguity due to deception.

2.1.2 Distributionally-Robust Methods

A distributionally-robust Blue team response given opponent team strategy $\pi_r(\cdot) \in \Pi_\text{r}$ optimizes for the best worst-case performance for type distributions belonging to an ambiguity set. The distributionally-robust best-response operator corresponding to a distributional ambiguity set $\mathscr{P}$ is defined as $\text{BR}^{(\text{b})}_{\text{DRO}}(\pi_r(\cdot); \mathscr{P}, \Theta) = \underset{\pi\in\Pi_b}{\text{argsup}} \inf_{q\in\mathscr{P}} \mathbb{E}_{\theta\sim q}[V(\pi_r(\cdot), \pi; \theta)]$.

$$\mathbb{E}_{\theta\sim f}[V(\pi_r(\cdot), \pi_{b,\text{DRO}}(\pi_r(\cdot), \mathscr{P}); \theta)] \leq \inf_{q\in\mathscr{P}} \mathbb{E}_{\theta\sim q}[V(\pi_r(\cdot), \pi_{b,\text{DRO}}(\pi_r(\cdot), \mathscr{P}); \theta)]. \tag{3}$$

The robustness guarantee of the DRO best-response is restricted to distributions contained within the ambiguity set $\mathcal{P}$. Consequently, DRO does not provide performance guarantees under strategic distribution shifts that fall outside the prescribed uncertainty model.

2.1.3 Adversarial Nature and Worst-Case Robustness

A rational, albeit pessimistic Blue team best-response under adversarial type selection is given by $\text{BR}^{(\text{b})}_{\text{wc}}(\pi_r(\cdot)) = \underset{\pi\in\Pi_\text{b}}{\text{argsup}} \left(\inf_{q\in\Delta(|\Theta|)} \mathbb{E}_{\theta\sim q}[V(\pi_r(\cdot), \pi; \theta)]\right)$. Let q_wc be the adversarial type distribution corresponding to the above worst-case best-response operator on given $\pi_r(\cdot)$ such that for $\pi_{b,\text{wc}}(\pi_r(\cdot)) \in \text{BR}^{(\text{b})}_{\text{wc}}(\pi_r(\cdot))$,

$$\sup_{\pi\in\Pi_\text{b}} \inf_{q\in\Delta(|\Theta|)} \mathbb{E}_{\theta\sim q}[V(\pi_r(\cdot), \pi; \theta)] = \mathbb{E}_{\theta\sim q_\text{wc}}[V(\pi_r(\cdot), \pi_{b,\text{wc}}(\pi_r(\cdot)); \theta)].$$

For a given Red team strategy, the above worst-case best-response operator solves a zero-sum game between the Blue team and Nature to compute a robust Blue team best-response. This leads to the following three-player equilibrium formulation where Nature acts as an adversary and plays an adversarial type distribution such that the Red team and Nature collude (since Red can correlate play with Nature) against the Blue team, $\pi^*_{b,\text{wc}} \in \text{BR}^{(\text{b})}_{\text{wc}}(\pi^*_r(\cdot))$ and $\pi^*_r(\cdot) \in \text{BR}^{(\text{r})}(\pi^*_{b,\text{wc}}; \cdot)$. Worst-case robust equilibrium provides a global performance

guarantee when the Nature is adversarial such that $\mathbb{E}_{\theta\sim q_{\text{wc}}}[V(\pi_r(\cdot),\pi^{'};\theta)] \leq \mathbb{E}_{\theta\sim q_{\text{wc}}}[V(\pi_r(\cdot),\pi_{b,\text{wc}};\theta)]$. This however can lead to overly pessimistic strategies such that,

$$\mathbb{E}_{\theta\sim q_{\text{wc}}}[V(\pi_r(\cdot),\pi_{b,\text{wc}};\theta)] \leq \mathbb{E}_{\theta\sim q\in\mathcal{Q}}[V(\pi_r(\cdot),\pi_{b,\text{wc}};\theta)] \leq \mathbb{E}_{\theta\sim q\in\mathcal{Q}}[V(\pi_r(\cdot),\pi_{b,\text{DRO}};\theta)]. \tag{4}$$

The three-player game formulation treats all types with equal importance independent of likelihood under nominal distribution and computes a robust Blue response to an adversarial type distribution that concentrates mass potentially on the hardest type. Such an approach discards available probabilistic information about the occurrence of types and could lead to behavior with poor typical performance under non-adversarial distribution shifts. In the pursuit of global performance guarantees, we discuss a tangential approach in the following subsection that reasons about regret with respect to type-optimal best-responses.

2.1.4 Minimax-Regret Robust Best-Response and Minimax-Regret Equilibrium (MRE)

Distributionally robust methods offer robustness only up to an ambiguity set centered at the nominal distribution and are suitable for settings when protection is sought against mis-specification of the nominal distribution (prior) and not strategic distribution shifts. *Strategic distribution shifts* refer to a *smart* Nature running an oracle procedure to compute a type distribution that potentially places probability mass adversarially on high-regret types. We characterize regret (or exploitability) with respect to a type as follows,

Definition 2.4. *(Exploitability) We define the exploitability (or regret) of Blue team policy π on type θ given Red team policy $\pi_r(\cdot)$ as,*

$$\mathcal{E}\left(\pi_r(\cdot),\pi;\theta\right) = \sup_{\pi^*\in\Pi_b} V(\pi_r(\cdot),\pi^*;\theta) - V(\pi_r(\cdot),\pi;\theta), \tag{5}$$

and the worst-case exploitability of π given $\pi_r(\cdot)$ over the type-space as $\mathcal{E}\left(\pi_r(\cdot),\pi;\Theta\right) = \sup_{\theta\in\Theta}\mathcal{E}\left(\pi_r(\cdot),\pi;\theta\right)$.

We define the following Blue team robust best-response given opponent team strategy $\pi_r(\cdot)$ that minimizes worst-case regret with respect to type-optimal best-responses across the type-space.

Definition 2.5. *(Minimax-Regret Best-Response Operator under Asymmetric Information)*

$$\text{BR}^{(\text{b})}_{\text{MMR}}(\pi_r;\Theta) = \operatorname*{arginf}_{\pi\in\Pi_\text{b}} \sup_{\theta\in\Theta}\mathcal{E}\left(\pi_r(\cdot),\pi;\theta\right) \tag{6}$$

In other words, the minimax-regret robust best-response penalizes deviation from the performance obtained under perfect information about the type. Such a rational attitude hedges across all types under imperfect information and prevents being overly sub-optimal on any single type. The ex-ante Minimax-Regret (MRE) equilibrium (π_b^*,π_r^*) is defined as the fixed point of the best-response operator defined in Def. 2.5, $\pi^*_{b,\text{MRE}} \in \text{BR}^{(\text{b})}_{\text{MMR}}(\pi_r^*;\Theta)$ and $\pi_r^*(\cdot) \in \text{BR}^{(\text{r})}(\pi^*_{b,\text{MRE}};\cdot)$. The MRE best-response is distribution-free and provides global performance guarantees independent of any restrictive assumptions about the type to lie within an ambiguity set around the nominal distribution. The key consequence of minimax-regret reasoning is that it induces a performance certificate that is valid for every type distribution in the simplex. This certificate is formalized below.

Lemma 1 (*Uniform Performance Lower Bound*)**.** *For a given Red team strategy $\pi_r(\cdot)$, the expected payoff under Blue team strategy π_b and a type distribution $\mu(\cdot)$ can be bounded from below as follows such that the lower bound is tight, i.e., $\exists\ q \in \Delta(|\Theta|)$ for which equality holds,*

$$\mathbb{E}_{\theta\sim\mu}[V(\pi_r(\cdot),\pi_b;\theta] \geq \mathcal{V}^{(-)}(\pi_b,\mu) = \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi\in\Pi_\text{b}} V(\pi_r,\pi;\theta) - \mathcal{E}(\pi_r(\cdot),\pi_b,\Theta)\ \forall\ \mu\in\Delta(|\Theta|). \tag{7}$$

Proof. Proof in Appendix A.1. □

From (7), the lower-bound on performance can be decoupled into a per-type component, free of the decision variable π_b, representing optimal decision under perfect information and a second term denoting the worst-case exploitability of the decision π_b across the type-space. Thus, the MRE best-response $\pi_{b,\text{MRE}} \in \text{BR}^{(\text{b})}_{\text{MMR}}(\pi_r;\Theta)$ from (6) can equivalently be interpreted as the decision $\pi_{b,\text{MRE}}$ that maximizes the performance lower-bound $\mathcal{V}^{(-)}(\cdot,\mu)$ for all type distributions $\mu(\cdot)$ in the type-simplex i.e.,

$$\text{BR}^{(\text{b})}_{\text{MMR}}(\pi_r;\Theta) = \underset{\pi\in\Pi_\text{b}}{\text{arginf}} \sup_{\theta\in\Theta} \mathcal{E}\left(\pi_r(\cdot),\pi;\theta\right) = \underset{\pi\in\Pi_\text{b}}{\text{argsup}}\, \mathcal{V}^{(-)}(\pi,\mu)\ \forall\ \mu(\cdot)\in\Delta(|\Theta|).$$

In contrast to DRO methods that do not provide guarantees for perturbations outside the ambiguity set and worst-case robust equilibria that concentrate mass on hard types agnostic to the likelihood information contained in the nominal type distribution leading to poor typical performance under non-adversarial distribution shifts, MRE provides a uniform lower bound on performance across the type-simplex by hedging regret across all types.

Limitation: MRE minimizes worst-case exploitability across the entire type-space and does not utilize useful probabilistic information from the nominal type distribution. For instances where a type with high exploitability is rare under the nominal type-distribution and admissible perturbations thereof, minimizing worst-case exploitability across the entire type-space might lead to conservative performance under typical distributions. This necessitates the need for a robust best-response and an associated equilibrium concept bridging distributionally robust and risk-averse methods and minimax-regret equilibria that utilizes probabilistic information about type occurrence whilst also providing robustness guarantees to non-adversarial distribution shifts in terms of a performance lower bound for typical distributions.

3 Probabilistically-Robust Minimax-Regret Equilibria

Motivated by the limitations of existing equilibrium concepts as discussed in Section 2.1, we introduce a novel robust best-response operator and the associated equilibrium concept – Probabilistically-Robust Minimax-Regret equilibria for ATGs under asymmetric information that minimizes exploitability across subsets of the type-space weighted by known information about their probability mass under admissible distribution shifts. We first discuss the threat model. We consider distribution shifts that preserve type typicality, meaning that types that are rare under the nominal distribution remain rare under admissible shifts. Admissible distributions therefore preserve coarse probabilistic information encoded in the nominal type distribution while allowing substantial redistribution of probability mass within high-probability regions of the type space. This is formalized via covers of the type-space as follows,

Definition 3.1. *(Probabilistic Covers of Type-Space) For a nominal type distribution $\bar{\mu}(\cdot)$ and confidence parameter δ, a valid high-probability cover of the type-space $\mathscr{C}\subseteq\Theta$ is such that $\sum_{\theta\in\mathscr{C}}\bar{\mu}(\theta)\geq 1-\delta$. The associated low-probability cover $\mathscr{L}$ to the high-probability cover $\mathscr{C}$ is defined as $\mathscr{L}=\mathscr{C}^c$ such that $\sum_{\theta\in\mathscr{L}}\bar{\mu}(\theta)\leq\delta$.*

We characterize permissible perturbations to the nominal type distribution as given by the set of distributions that preserve all high-probability covers (or equivalently, all low-probability covers) of the type-space. For a nominal type distribution $\bar{\mu}(\cdot)$ and confidence parameter δ, the set of all $(1-\delta)$- high-probability covers Θ is defined as $\Theta(\bar{\mu},\delta) \coloneqq \{\mathscr{C}\subseteq\Theta \text{ such that } \sum_{\theta\in\mathscr{C}}\bar{\mu}(\theta)\geq 1-\delta\}$ and the set of all δ- low-probability covers is defined as $\Theta_\ell(\bar{\mu},\delta)\coloneqq\{\mathscr{C}^c \text{ for all } \mathscr{C}\in\Theta(\bar{\mu},\delta)\}$. We also denote the set of all rare types under the nominal distribution and confidence level δ as $\Theta_r(\bar{\mu},\delta)=\{\theta \text{ such that } \bar{\mu}(\theta)\leq\delta\}$. We now formally characterize our threat model as follows,

Definition 3.2. *(Threat Model) The permissible set of perturbed distributions $\mathcal{S}(\bar{\mu},\delta)$ is given as $\mathcal{S}(\bar{\mu},\delta)=\{\mu(\cdot)\in\Delta(|\Theta|) \text{ such that } \sum_{\theta\in\mathscr{L}}\mu(\theta)\leq\delta \text{ for all } \mathscr{L}\in\Theta_\ell(\bar{\mu},\delta)\}$.*

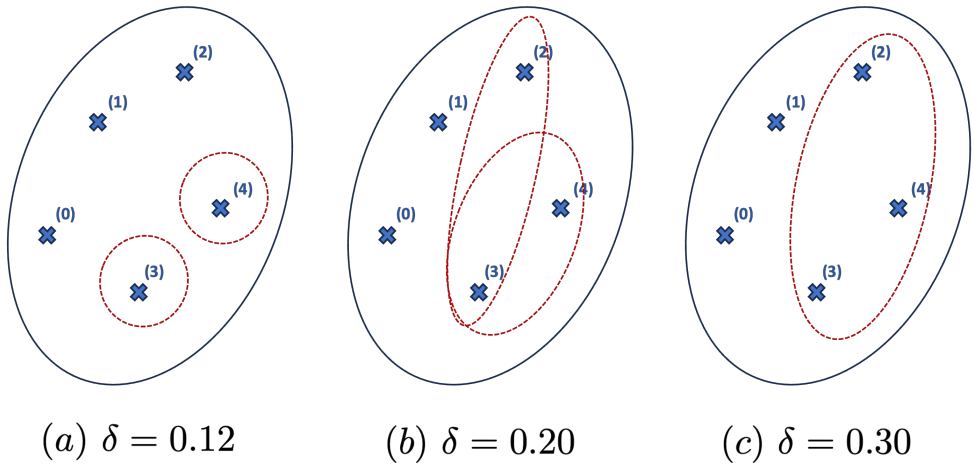

Figure 3: Illustration of PR-MRE, nominal distribution $\bar{\mu}(\cdot) = (0.35, 0.35, 0.15, 0.03, 0.12)$. The blue outer ellipse denotes the type space $\Theta = \{\theta_0, \cdots, \theta_4\}$, the crosses denote specific types and the red dotted ellipses denote the set of low-probability covers $\Theta_l(\bar{\mu}, \delta)$ for various confidence levels δ. For (a), $\Theta_l(\bar{\mu}, 0.12) = \{\{\theta_3\}, \{\theta_4\}\}$ such that under admissible shifts $\mu \in \Delta(|\Theta|)$, $\mu(\delta_3) \leq \delta$ and $\mu(\delta_4) \leq \delta$. Similarly for (b), $\Theta_l(\bar{\mu}, 0.20) = \{\{\theta_2, \theta_3\}, \{\theta_3, \theta_4\}\}$ such that $\mu(\theta_2) + \mu(\theta_3) \leq \delta$ and $\mu(\delta_3) + \mu(\delta_4) \leq \delta$, and for (c) $\Theta_l(\bar{\mu}, 0.30) = \{\theta_2, \theta_3, \theta_4\}$ such that $\mu(\theta_2) + \mu(\theta_3) + \mu(\theta_4) \leq \delta$. Rather than minimizing worst-case exploitability uniformly across the entire type space as in MRE, PR-MRE focuses on high-confidence subsets while retaining robustness to low-probability covers defined through the confidence level δ.

In other words, the permissible perturbed distributions under our threat model allocate probability mass arbitrarily up to respecting typicality of types such that low-probability covers remain low probability.

For instance, for a nominal distribution $\bar{\mu}(\cdot) = (0.35, 0.35, 0.15, 0.03, 0.12)$ and $\delta = 0.12$, it is easy to verify that the low-probability cover is given by $\Theta_\ell(\bar{\mu}, 0.12) = \{(\theta_3), (\theta_4)\}$ (as depicted in Figure 3a) such that for any permissible perturbed distribution $\mu(\cdot) \in \mathcal{S}(\bar{\mu}, \delta)$, $\mu(\theta_3) \leq \delta$ and $\mu(\theta_4) \leq \delta$. Similarly for $\delta = 0.2$, the low probability cover is given by $\Theta_\ell(\bar{\mu}, 0.20) = \{(\theta_2, \theta_3), (\theta_3, \theta_4)\}$ and permissible distributions $\mu(\cdot)$ are such that $\mu(\theta_2) + \mu(\theta_3) \leq \delta$ and $\mu(\theta_3) + \mu(\theta_4) \leq \delta$ (Figure 3b), and for $\delta = 0.3$ (Figure 3c), $\mu(\theta_2) + \mu(\theta_3) + \mu(\theta_4) \leq \delta$. The inequality constraints arise as a direct consequence of the threat model that preserves coarse probabilistic information about type occurrence (all high-probability covers of the type-space) under distribution shift. Such a threat model makes sense for scenarios when the task hypotheses map to actual real-world configurations. For instance, for capture-the-flag, certain flag hypotheses might actually be realized with a bounded probability because of terrain traversability, other physical constraints etc.

Geometry of $\mathcal{S}(\bar{\mu}, \delta)$: The threat model from Definition 3.2 admits a meaningful class of strategic perturbations not captured by ball-shaped ambiguity sets centered at a nominal type distribution (based on for example KL- or generalized f-divergence). Unlike ambiguity-set methods that constrain perturbations to remain close to the nominal distribution, our threat model permits arbitrary redistribution of probability mass within a high-confidence subset of the type space while preserving its aggregate probability mass. For the high-confidence subset $\Theta/\Theta_r \subseteq \Theta$, arbitrary allocations of probability mass across types within Θ/Θ_r is allowed as long as the net mass over rare types Θ_r follows constraints posed by Definition 3.2. Consequently, the model admits highly concentrated ("peaky") distributions that place most of their mass on a single type, while still preserving coarse probabilistic information about which regions of the type space are typical. Figure 3 illustrates the key distinction such that for $\bar{\mu} = (0.35, 0.35, 0.15, 0.03, 0.12)$ and $\delta = 0.12$, $(0, 0, 1, 0, 0)$ is a permissible type distribution under the typicality-preserving threat model where $\mu(\theta_1), \mu(\theta_2)$ shift from $0.35 \to 0$ and $\mu(\theta_3)$ shifts from $0.15 \to 1$. Other example distributions contained in the permissible set $\mathcal{S}(\bar{\mu}, 0.12)$ include $(0.12, 0, 0.76, 0.06, 0.06), (0.90, 0, 0, 0.10, 0)$ and $(0, 1, 0, 0, 0)$.

This motivates a robust best-response operator that balances exploitability on high-probability regions of the type space against exploitability on atypical types. The resulting operator is defined as follows.

Probabilistically-Robust Minimax-Regret (PR-MRE) Best-Response Operator

Definition 3.3. *(PR-MRE Best-Response Operator under Asymmetric Information) For a given Red team strategy $\pi_r(\cdot)$, the probabilistically-robust MRE best-response operator for the Blue team is defined as follows,*

$$\mathrm{BR}^{(\mathrm{b})}_{\mathrm{PRMRE}}(\pi_r(\cdot);\bar{\mu},\delta) = \operatorname*{arginf}_{\pi\in\Pi_{\mathrm{b}}} \sup_{\mu\in\mathcal{S}(\bar{\mu},\delta)} \sum_{\theta\in\Theta} \mu(\theta)\mathcal{E}\left(\pi_r(\cdot),\pi;\theta\right). \tag{8}$$

Lemma 2. *For a given Red team strategy $\pi_r(\cdot)$, the expected payoff for Blue team strategy π_b and any type distribution $\mu(\cdot)\in\mathcal{S}(\bar{\mu},\delta)$ permissible under the threat model can be bounded from below as follows,*

$$\mathbb{E}_{\theta\sim\mu}[V(\pi_r(\cdot),\pi_b;\theta] \geq \mathcal{V}^{(-)}_{\mathcal{S}}(\pi_b,\mu) = \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_{\mathrm{b}}} V(\pi_r,\pi^*;\theta) - \sup_{q\in\mathcal{S}(\bar{\mu},\delta)}\left(\sum_{\theta\in\Theta} q(\theta)\mathcal{E}(\pi_r(\cdot),\pi_b;\theta)\right). \tag{9}$$

It is straightforward to prove the following Lemma that the PR-MRE robust best-response from (8) is equivalent to (10) through a water-filling argument by arguing that the inner supremum for any outer $\pi\in\Pi_b$ corresponds to a type distribution that places all permissible mass on the highest regret type within the high-confidence Θ/Θ_r subset while respecting the remaining constraints posed by the low-probability covers. Lemma 3 shows that the PR-MRE robust best-response weighs regret over rare types by a maximum of δ in contrast to the distribution-free MRE (6) that treats all types as equally likely and minimizes worst-case regret across the type-space.

Lemma 3. *For a given Red team strategy $\pi_r(\cdot)$, the probabilistically-robust MRE best-response operator for the Blue team is defined as follows,*

$$\mathrm{BR}^{(\mathrm{b})}_{\mathrm{PRMRE}}(\pi_r(\cdot);\bar{\mu},\delta) = \operatorname*{arginf}_{\pi\in\Pi_{\mathrm{b}}} \sup_{\substack{\mu(\Theta/\Theta_r)+\sum_{\theta\in\Theta_r}\mu(\theta)=1,\\ \sum_{\theta\in\mathscr{L}}\mu(\theta)\leq\delta\ \forall\ \mathscr{L}\in\Theta_\ell(\bar{\mu},\delta)}} \mu(\Theta/\Theta_r)\cdot\mathcal{E}\left(\pi_r(\cdot),\pi;\Theta/\Theta_r\right) + \sum_{\theta\in\Theta_r}\mu(\theta)\cdot\mathcal{E}\left(\pi_r(\cdot),\pi;\theta\right), \tag{10}$$

where $\mathcal{E}(\pi_r(\cdot),\pi;\Theta/\Theta_r)$ refers to the worst-case exploitability of the Blue team strategy π over the type-space subset Θ/Θ_r where Θ_r is the set of all rare types such that under the nominal distribution $\bar{\mu}(\cdot)$, $\bar{\mu}(\theta)\leq\delta,\ \forall\ \theta\in\Theta_r$.

Probabilistically-Robust Minimax-Regret Equilibrium (PR-MRE)

The objective in (8) minimizes worst-case exploitability on a high-confidence subset of the type space while discounting exploitability on atypical types by the confidence parameter δ. From the novel Blue team best-response operator (8) as proposed above, we obtain the equilibrium concept – Probabilistically-Robust Minimax-Regret equilibria (PR-MRE) characterized by the following fixed-point system,

$$\pi^*_{b,\mathrm{PRMRE}}(\bar{\mu},\delta)\in\mathrm{BR}^{(\mathrm{b})}_{\mathrm{PRMRE}}(\pi^*_r(\cdot);\bar{\mu},\delta) \text{ and } \pi^*_r(\cdot)\in\mathrm{BR}^{(\mathrm{r})}(\pi^*_{b,\mathrm{PRMRE}}(\bar{\mu},\delta);\cdot). \tag{11}$$

The key robustness guarantee of PR-MRE is given below.

Lemma 4. *(Probabilistically-Robust Minimax-Regret Equilibria Robustness Property) For all distributions permissible under the threat model $\mathcal{S}(\bar{\mu}, \delta)$,* PR-MRE *provides the tightest lower-bound on expected performance such that $\mathcal{V}_{\mathcal{S}}^{(-)}(\pi_{b,\text{PRMRE}}, \mu) \geq \mathcal{V}_{\mathcal{S}}^{(-)}(\pi, \mu)$ for all $\mu \in \mathcal{S}(\bar{\mu}, \delta)$, $\pi \in \Pi_b$.*

Proof. Proof in Appendix A.1. □

A corollary of Lemma 4 is that PR-MRE, which minimizes a weighted sum of exploitabilities over high-confidence subsets of the type-space, obtains a tighter performance lower bound on permissible distributions under the threat model as compared to MRE, which minimizes worst-case exploitability over the entire type-space, $\mathcal{V}_{\mathcal{S}}^{(-)}(\pi_{b,\text{PRMRE}}, \mu) \geq \mathcal{V}_{\mathcal{S}}^{(-)}(\pi_{b,\text{MRE}}, \mu) \ \forall \ \mu \in \mathcal{S}(\bar{\mu}, \delta)$.

This motivates PR-MRE as an appropriate robust equilibrium concept under asymmetric information, providing a tighter lower bound on expected performance under admissible distribution shifts than distributionally robust, worst-case, and minimax-regret equilibrium concepts. In the Graph CtF setting, this corresponds to favoring strategies that remain robust across plausible flag hypotheses while avoiding excessive conservatism induced by highly atypical flag locations.

The tighter lower bound does not come for free, as we establish in Section 3.1 that the computation of PR-MRE for a metagame approximation of the ATG can be formulated as a robust bilinear program that couples regrets across types as compared to a simpler bilinear program for MRE and the well-known linear programming formulation for BNE. This is due to the fact that PR-MRE reasons about which regions of the type-space to minimize regret over as compared to MRE that minimizes worst-case regret over the entire type-space.

Programming Formulation

Unlike BNE, whose best response is a single optimization over expected utility, regret based reasoning introduces a nested optimization over type specific optima. Consequently, the optimization couples regrets across types, leading to a robust bilinear program rather than the linear programs familiar from Bayesian equilibria. We next derive this formulation and subsequently develop a tractable semidefinite relaxation for use within PSRO.

From (8) and Definition 3.2, the PR-MRE best-response operator can be reformulated as follows,

$$\inf_{\pi \in \Pi_b} \sup_{\substack{\mu \in \Delta(|\Theta|) \\ \sum_{\theta \in \mathscr{L}} \mu(\theta) \leq \delta \ \forall \ \mathscr{L} \in \Theta_l(\bar{\mu}, \delta)}} \mu^\top \varepsilon(\pi) \iff \inf_{\tilde{\varepsilon}, \pi} \tilde{\varepsilon} \quad \sup_{\substack{\mu \in \Delta(|\Theta|) \\ \sum_{\theta \in \mathscr{L}} \mu(\theta) \leq \delta \ \forall \ \mathscr{L} \in \Theta_l(\bar{\mu}, \delta)}} \mu^\top \varepsilon(\pi) = \tilde{\varepsilon}(\pi) \leq \tilde{\varepsilon} \quad (12)$$

where $\mu^\top \varepsilon(\pi) = \sum_{\theta \in \Theta} \mu(\theta) \mathcal{E}\left(\pi_r(\cdot), \pi; \theta\right)$ is the expected regret under $\mu(\cdot)$ and Blue team policy π.

The robust average regret condition couples regrets across types. Moreover, optimizing over regrets is a nested optimization problem as opposed to BNE, distributionally-robust and worst-case notions of best-response since to optimize the regret, deviation from type-wise optimal solutions is minimized. For $\delta = 0$, $\Theta_l(\bar{\mu}, 0) = \Phi$ such that robust average regret under the threat model reduces to worst-case regret corresponding to MRE best-response. The nested infimum and supremum can be reformulated as shown above. Eq. (12) is a robust formulation for expected regret under type distributions permissible under the threat model.

PR-MRE Best-Response Programming Formulation

For Red team strategy $\pi_r(\cdot)$,

$$\inf_{\tilde{\varepsilon},\pi} \quad \tilde{\varepsilon} \tag{PRMRE-BR}$$

$$\delta\lambda^\top \mathbf{1}_{|\mathscr{L}|\times 1} + \lambda_{\mathrm{u}} \leq \tilde{\varepsilon}, \tag{13}$$

$$-\varepsilon^\top(\pi) + \lambda^\top \mathcal{A}_{\mathscr{L}} - \lambda_{\mathrm{p}}^\top + \lambda_{\mathrm{u}}\mathbf{1} = 0, \tag{14}$$

$$\lambda \in \mathbb{R}_+^{|\mathscr{L}|}, \lambda_{\mathrm{p}} \in \mathbb{R}_+^{|\Theta|};\ \lambda_{\mathrm{u}} \in \mathbb{R}, \tag{15}$$

$$\text{where } \varepsilon(\pi) = \left[\sup_{\pi^* \in \Pi_b} V(\pi_r(\cdot), \pi^*; \theta) - V(\pi_r(\cdot), \pi; \theta)\right]_{\theta\in\Theta}, \tag{16}$$

is the regret vector corresponding to Blue team policy π.

From the Red team best-response operator defined in Definition 2.1 and the PR-MRE equilibrium (11) defined as the fixed-point of the Blue team and Red team best-response operators under asymmetric information, we obtain the following programming formulation for the PR-MRE equilibrium.

PR-MRE Equilibrium Programming Formulation

$$\inf_{\tilde{\varepsilon},\pi_{\mathrm{PRMRE}},\pi_r^*(\cdot)} \quad \tilde{\varepsilon} \tag{PRMRE-EQM}$$

$$\delta\lambda^\top \mathbf{1}_{|\mathscr{L}|\times 1} + \lambda_{\mathrm{u}} \leq \tilde{\varepsilon}, \tag{17}$$

$$\varepsilon(\pi_{\mathrm{PRMRE}}) = \left[\sup_{\pi^* \in \Pi_b} V(\pi_r(\cdot), \pi^*; \theta) - V(\pi_r(\cdot), \pi_{\mathrm{PRMRE}}; \theta)\right]_{\theta\in\Theta}, \tag{18}$$

$$-\varepsilon^\top(\pi_{\mathrm{PRMRE}}) + \lambda^\top \mathcal{A}_{\mathscr{L}} - \lambda_{\mathrm{p}}^\top + \lambda_{\mathrm{u}}\mathbf{1} = 0, \tag{19}$$

$$V(\pi_r^*(\cdot), \pi_{\mathrm{PRMRE}}; \theta) \geq V(\pi_r(\cdot), \pi_{\mathrm{PRMRE}}; \theta) \text{ for all } \pi_r(\cdot) \in \Pi_r \text{ and } \theta \in \Theta. \tag{20}$$

Illustrative Example

Consider an example normal-form game to illustrate the distinction between PR-MRE and the various equilibrium concepts discussed in Section 2. Consider the $1 \times 5 \times 3$ game from Table 3 with type-space $\Theta = \{\theta_0, \theta_1, \theta_2\}$, five Blue actions b_0 to b_4 on the columns, a single Red action r_0 as the row with three rows one for each type and a nominal prior on types $\bar{\mu} = (0.25, 0.70, 0.05)$. We use a single Red action in this illustrative example so that the informed Red best-response is unique, allowing differences between equilibrium fixed-points to be attributed solely to the Blue team's varied robust best-responses. A more complicated $2\times5\times3$ example with the added complexity of Red's best-response for determining the respective equilibrium fixed-points is provided in Appendix A.3 (see Table 4). From the example game shown below in Table 3, action b_0 is a θ_0-specialist such that $V^*(\theta_0) = V(b_0; \theta_0)$, b_1 is a θ_1-specialist and b_2 is a θ_2-specialist that is rare under the nominal ($\bar{\mu}(\theta_2) = 0.05$). Moreover, actions b_0 and b_1 are catastrophic on the rare type θ_2 and action b_2 is catastrophic on typical types θ_0 and θ_1. Actions b_3 and b_4 are generalists such that they are not too bad for any particular type. In other words, b_3 and b_4 have markedly better worst-case regret over types as compared to the first-three specialist columns, 1.1 and 1.6 against 2.1, 2.3 and 2.0 for b_0, b_1 and b_2 respectively. Moreover, b_3 is the best generalist with lowest worst-case regret of 1.1 across types (see column b_3 in Table 3).

From the taxonomy on robust Blue best-responses under asymmetric information and associated equilibria in Tables 1 and 2, each best-response corresponds to a degree of trust in Nature's nominal type distribution and an optimization objective resulting in a unique robustness property. We now contrast PR-MRE and the various equilibrium concepts with respect to robustness to shifts from the nominal on the example game and visualize their comparative advantage over different regions of the type-simplex (see Figure 4 and Table 3). The nominal type distribution $(0.25, 0.70, 0.05)$ represents a single point in the triangular barycentric representation of the three-dimensional simplex (Figure 4) and we evaluate the performance

Table 3: Example 1×5×3 game. Columns $\{b_0 \cdots b_4\}$ are Blue actions, row r_0 corresponds to the Red action and there are three total rows, one per type $\theta \in \{\theta_0, \theta_1, \theta_2\}$ with a nominal distribution $\bar{\mu} = (0.25, 0.70, 0.05)$. Each cell shows the Blue payoff $V(r, b; \theta)$, shorthand $V(b; \theta)$ for the unique Red choice r_0, the type-conditioned optimum values $V^*(\theta)$ read left to right per θ row are marked in bold (black), and the per-type regret for any Blue action (or column) given by $\varepsilon(b; \theta) = V^*(\theta) - V(b; \theta)$ is in red subscript. Worst-case regret over types $\sup_{\theta \in \Theta} \varepsilon(b; \theta)$ for any Blue action b is read top to bottom per column and marked in bold (red).

	b_0	b_1	b_2	b_3	b_4	
r_0	**1.5** $_{0.0}$	0.6 $_{0.9}$	−0.5 $_{\mathbf{2.0}}$	0.4 $_{\mathbf{1.1}}$	0.7 $_{0.8}$	$\}\ \theta_0, 0.25$
r_0	0.3 $_{0.7}$	**1.0** $_{0.0}$	−0.8 $_{1.8}$	0.8 $_{0.2}$	0.8 $_{0.2}$	$\}\ \theta_1, 0.70$
r_0	−0.6 $_{\mathbf{2.1}}$	−0.8 $_{\mathbf{2.3}}$	**1.5** $_{0.0}$	0.5 $_{1.0}$	−0.1 $_{\mathbf{1.6}}$	$\}\ \theta_2, 0.05$

(Blue payoff) of the Blue decision corresponding to different robust equilibrium fixed-points as the test type distribution probes over the entire simplex.

Evaluation Protocol: For our comparison and analysis (Figures 4 and 5), the performance of the Blue decision corresponding to different equilibria for any test type distribution is measured against an adaptive Red decision, specifically the Red BNE corresponding to the test type distribution (Red's decision corresponding to the BNE fixed-point for the test type distribution). Figure 4 shows the winner map and payoff difference heatmaps for PR-MRE and the different methods against such a Red opponent that adapts as the test type distribution probes over the simplex. For the single Red action example in Table 3, the Red decision is unique everywhere by construction and the robust Blue decisions are evaluated against variations in just the Nature's decision / type distribution (Red decision is same everywhere). For the $2 \times 5 \times 3$ example in Appendix A.3, evaluation and visualization is done against an adaptive Red opponent.

Remark: We note that we compare PR-MRE to two types of distributionally-robust equilibria, DRO-Ball based on an $L1$-ball around the nominal (local trust) and DRO-MaxMin based on the structured type ambiguity induced by the threat model in Definition 3.2 (coarse trust) that preserves the set of all high-probability covers under admissible distribution shifts. This provides a direct comparison of our regret-based approach PR-MRE to the payoff-based distributionally-robust approach DRO-MaxMin over the same structured ambiguity set.

BNE, DRO-Ball, MaxMin and DRO-MaxMin are all computed via linear-programs whereas MRE and PR-MRE ($\delta = 0.15$) are computed via semidefinite relaxations to the corresponding bilinear and robust bilinear programs as derived in Section 3.1 for normal form games under asymmetric information. For the example game in Table 3, the computed BNE Blue decision concentrates on the specialist b_1 for the highest-confidence type θ_1 under nominal $\bar{\mu} = (0.25, 0.70, 0.05)$ as $(0, 1, 0, 0, 0)$ whereas the computed DRO-Ball decision for an $L1$-ball of radius $\rho = 0.5$ around the nominal (dotted hexagon around the nominal star in Figure 4) hedges probability mass across generalist actions b_3, b_4 as $(0, 0, 0, 0.89, 0.11)$. MRE that discards probabilistic information from the nominal and minimizes worst-case regret across all types results in a decision $(0.39, 0, 0.39, 0.22, 0)$ with 0.22 mass to the best generalist b_3 and a significant probability mass of 0.39 to the rare-type specialist b_2. This comes at the expense of performance over typical types θ_0 and θ_1 as the poorest worst-case and average payoffs over the high-confidence under nominal structured type ambiguity set $\mu(\theta_2) \leq 0.15$ (the black dotted line and the trapezoidal band beneath it in Figure 4). The MaxMin decision that assumes an adversarial type distribution (Section 2.1.3) and optimizes for worst-case payoff across the type-simplex is the only other decision besides MRE that puts mass on the rare-type specialist b_2 as $(0.18, 0, 0.16, 0.66, 0)$. In contrast to distribution-free MRE, PR-MRE utilizes coarse information from the nominal and optimizes for worst-case regret across a structured type ambiguity set that respects typicality of types. For $\delta = 0.15$, the PR-MRE best-response discounts regret incurred on the rare-type θ_2 (nominal mass 0.05) by 0.15 at maximum (follows from Lemma 3) such that the PR-MRE decision for our example game is $(0.46, 0, 0, 0, 0.54)$.

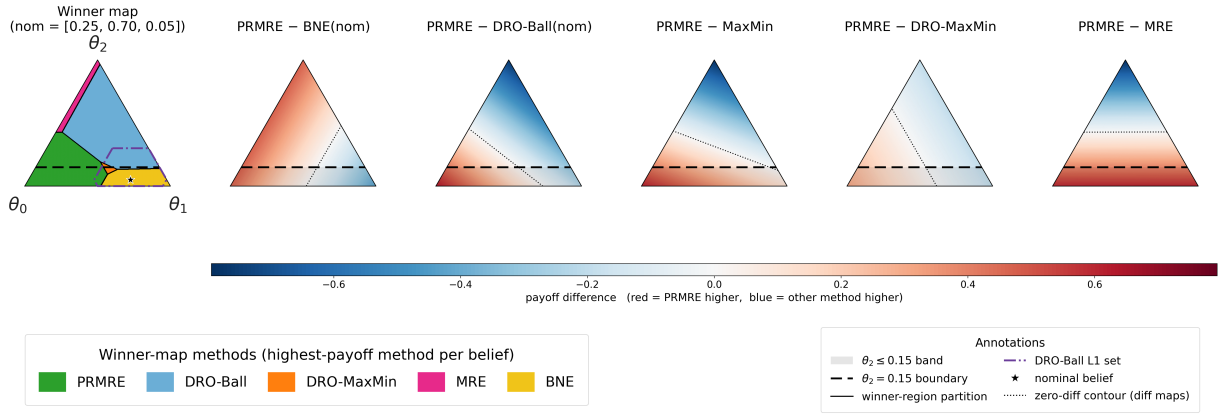

Figure 4: Comparison of PRMRE ($\delta = 0.15$) with BNE, DRO-Ball, MaxMin, DRO-MaxMin and MRE for the synthetic three-type game in Table 3: winner map (first triangle) and payoff difference heatmaps (remaining triangles). Triangles represent three-dimensional simplex in barycentric coordinates with vertices corresponding to full probability mass on the respective types. Black dotted line and the trapezoidal band beneath it represent the high-confidence under nominal structured type-ambiguity set $\mu(\theta_2) \leq 0.15$, the star in the winner map corresponds to the nominal distribution $\bar{\mu} = (0.25, 0.70, 0.05)$ and the dotted hexagon centered at the star corresponds to an $L1$-ball with radius $\rho = 0.5$ around it. For the trapezoidal band, PR-MRE dominates all other methods which can be seen from the green volume share in the winner-map and the red volume share in the payoff difference heatmaps. BNE and DRO-Ball perform better close to the nominal (θ_1-corner, mass 0.70 under nominal) but lose out on performance to PR-MRE for a type distribution chosen at random within the trapezoidal band. PR-MRE has the best average-case performance over the structured ambiguity set as compared to any other method (see Figure 5 for the numerical result, consistent with visual inspection of the winner-map and payoff difference heatmaps).

The PR-MRE decision favours b_4 over b_3 despite having a higher worst-case regret over the entire type-space (1.6 vs 1.1 for MRE) due to higher payoff and lower worst-case regret as compared to b_3 when restricted to typical types θ_0 and θ_1. Blue action b_3 is the best overall generalist and b_4 is a more specialized generalist that performs better over regions which matter probabilistically. The payoff-based DRO-MaxMin decision $(0.08, 0, 0, 0, 0.92)$ concentrates over an adversarial type distribution within the structured type ambiguity set as compared to the regret-based PR-MRE that hedges across the $\mu(\theta_2) \leq 0.15$ trapezoidal band by having roughly equal mass 0.46 to the θ_0-corner that constitutes the left side of the band in Figure 4. From Figure 5, PR-MRE has the best average-case performance over the structured ambiguity set as compared to any other method which is consistent with a visual inspection of the green share in the winner-map and the red share in the payoff difference heatmaps in Figure 4. Overall, the example game illustrates that the taxonomy in Table 1, defined by the optimization objective and degree of trust in Nature's nominal type distribution, induces qualitatively different robust Blue best-responses and corresponding equilibria, resulting in distinct robustness properties and complementary regions of advantage across the type simplex. We refer the reader to Appendix A.3 for results on a more complicated $2 \times 5 \times 3$ example game that extends the illustrative $1 \times 5 \times 3$ game by restoring the informed Red player's strategic choices when determining the corresponding robust equilibrium fixed points.

3.1 Computation of Probabilistically-Robust Minimax-Regret Equilibria via Robust Double Oracle

Computing equilibria in graph-structured adversarial team games is already intractable due to the size of the extensive-form representation, necessitating approximation methods such as CFR, PPO, and double-oracle approaches. In contrast, PR-MRE introduces a regret-based robustness objective over structured type uncertainty, which does not admit standard sequence-form linear programming formulations or straightforward extensions of CFR-style regret updates. To address this challenge, we derive a tractable optimization formu-

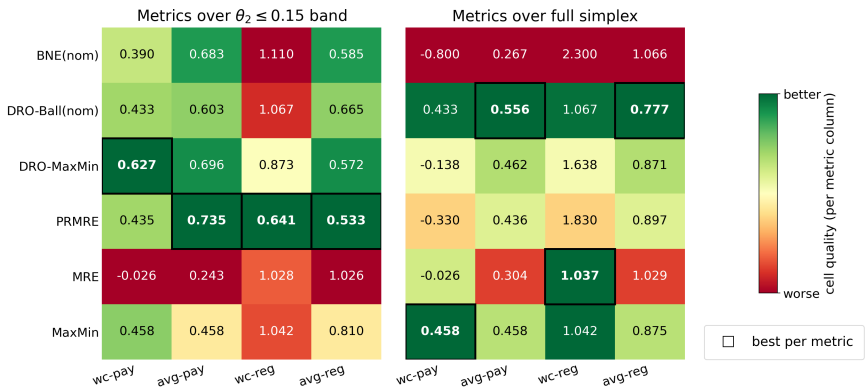

Figure 5: Comparison of PRMRE ($\delta = 0.15$) with BNE, DRO-Ball, MaxMin, DRO-MaxMin and MRE for the synthetic three-type game in Table 3 with respect to average and worst-case, payoff and regret over the structured ambiguity set $\mu(\theta_2) \leq 0.15$ and the entire type-simplex. Over the $\mu(\theta_2) \leq 0.15$ trapezoidal band (see Figure 4), PR-MRE has the best worst-case regret as compared to any other method. DRO-MaxMin on the other hand beats PR-MRE on worst-case payoff over the $\mu(\theta_2) \leq 0.15$ band which is consistent with its payoff-based objective (0.627 vs 0.435) but loses average-case performance for a type distribution chosen at random from the trapezoidal band such that PR-MRE attains the highest average payoff (0.735) as compared to any other method (0.696 for DRO-MaxMin). The regret-based objective of the PR-MRE best-response provides the best point-wise lower bound on payoff for type distributions belonging to the structured ambiguity set (as stated in Lemma 4) which in this example game manifests as a superior average-case performance for a type-distribution chosen at random from the $\mu(\theta_2) \leq 0.15$ structured ambiguity set.

lation of PR-MRE together with a semidefinite relaxation that can be embedded within a population-based learning framework.

For Finite Normal-Form Games

PR-MRE Best-Response Programming Formulation for Normal-Form Games

For a finite normal-form game defined by the payoff tensor $\{A(\cdot,\cdot;\theta)\}_{\theta\in\Theta} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{Y}|\times|\Theta|}$ and the omniscient Red strategy $x(\cdot) : \Theta \to \Delta(|\mathcal{X}|)$, the Blue PR-MRE best-response $y_{\text{PRMRE}} \in \Delta(|\mathcal{Y}|)$ denoted by $y_{\text{PRMRE}} \longleftarrow \text{PRMRE-BR}\{A, x(\cdot)\}$ can be formulated as the following program,

$$\inf_{\tilde{\varepsilon}, y_{\text{PRMRE}}} \quad \tilde{\varepsilon} \tag{PRMRE-BR-NF}$$

$$\delta\lambda^\top \mathbf{1}_{|\mathscr{L}|\times 1} + \lambda_{\text{u}} \leq \tilde{\varepsilon}, \tag{21}$$

$$x(\theta)A(\theta)e_j - x(\theta)A(\theta)y_{\text{PRMRE}} \leq \{\lambda^\top \mathcal{A}_{\mathscr{L}} - \lambda_{\text{p}}^\top + \lambda_{\text{u}}\mathbf{1}\}_\theta \text{ for all } j \in [|\mathcal{Y}|] \text{ and } \theta \in \Theta, \tag{22}$$

$$\lambda \in \mathbb{R}_+^{|\mathscr{L}|}, \lambda_{\text{p}} \in \mathbb{R}_+^{|\Theta|}, \lambda_{\text{u}} \in \mathbb{R}, y_{\text{PRMRE}} \in \Delta(|\mathcal{Y}|). \tag{23}$$

PR-MRE Equilibrium Bilinear Program for Normal-Form Games

For a finite normal-form game defined by the payoff tensor $\{A(\cdot,\cdot;\theta)\}_{\theta\in\Theta} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{Y}|\times|\Theta|}$, the PR-MRE fixed-point $\sigma_{\text{PRMRE}} = (x(\cdot), y)$ denoted by $\sigma_{\text{PRMRE}} \longleftarrow$ PRMRE-EQM$\{A\}$ is given by the following bilinear program,

$$\inf_{\tilde{\varepsilon},y,x(\cdot)} \quad \tilde{\varepsilon} \tag{PRMRE-EQM-NF}$$

$$\delta\lambda^\top \mathbf{1}_{|\mathscr{L}|\times 1} + \lambda_{\text{u}} \leq \tilde{\varepsilon}, \tag{24}$$

$$x(\theta)A(\theta)e_j - x(\theta)A(\theta)y \leq \{\lambda^\top \mathcal{A}_{\mathscr{L}} - \lambda_{\text{p}}^\top + \lambda_{\text{u}}\mathbf{1}\}_\theta \text{ for all } j \in [|\mathcal{Y}|] \text{ and } \theta \in \Theta, \tag{25}$$

$$-x(\theta)A(\theta)y \geq -e_k A(\theta)y \text{ for all } k \in [|\mathcal{X}|] \text{ and } \theta \in \Theta, \tag{26}$$

$$\lambda \in \mathbb{R}_+^{|\mathscr{L}|}, \lambda_{\text{p}} \in \mathbb{R}_+^{|\Theta|};\ \lambda_{\text{u}} \in \mathbb{R}, y \in \Delta(|\mathcal{Y}|), x(\theta) \in \Delta(|\mathcal{X}|) \text{ for all } \theta \in \Theta. \tag{27}$$

Rank-constrained Semidefinite Reformulation of PRMRE-EQM-NF: The details on the rank-constrained semidefinite formulation and convex relaxation are relegated to Section A.2.

Robust Double Oracle for Asymmetric Information Games

Double-oracle (DO) McMahan et al. (2003) is an algorithm for computing Nash Equilibrium (NE) in two-player zero-sum normal-form games. The algorithm proceeds by maintaining a population of strategies for each player, referred to as a metagame that abstracts away the full game, computes a NE over the metagame and iteratively adds strategies to the player populations by best-responding to the opponent meta-equilibria. The DO approach in principle recovers the NE for the full game. Policy Space Response Oracles (PSRO) Lanctot et al. (2017) extends DO to large games by employing reinforcement-learning to compute best-responses. The metagame NE is computed on the empirical game matrix derived from the matchup between each policy in the population set against all opponent policies and tracking the average payoff in a metagame payoff matrix.

As discussed in Section 2.1, NE (BNE) is risk-neutral and offers no robustness and performance guarantees under strategic distribution shifts for asymmetric information games such as Graph Capture-the-Flag where Nature's type is hidden from the Blue team and Red can condition it's play on the realized type and / or potentially control the type distribution. For the DO- and PSRO- policy expansion steps, best-responding to a NE (BNE) meta-equilibria over the opponent policy set might add brittle team policies to the population that suffer under distribution shift. Existing work on incorporating robustness at the metagame level and risk-aware policy expansion is limited Slumbers et al. (2023) and does not formally reason about exploitability of the population under hidden information and strategic shifts.

MRE and PR-MRE are regret-based notions of equilibria that provide robustness certificates in the form of uniform performance lower bounds (see Lemmas 1 and 2). We extend the double-oracle method to compute MRE and PR-MRE for large asymmetric information normal-form games via robust double-oracle (RDO). RDO (see Algorithm 1) is based on PRMRE-EQM-NF as the meta-equilibria and PRMRE-BR-NF as the best-response for policy set expansion of the uninformed player. PRMRE-BR-NF provides a principled best-response that adds strategies with minimum worst-case exploitability against the omniscient opponent type-conditioned meta-equilibria. The proposed RDO expansion results in a robust population that approximates the PR-MRE of the underlying normal-form game and is stable under distribution shifts of the hidden type. Termination of RDO is guaranteed by the fact that the meta-equilibrium PRMRE-EQM-NF is the fixed-point of the robust expansion operator based on PRMRE-BR-NF such that for the case of no new pure strategies added at an expansion iteration, the procedure terminates and the metagame approximates the PR-MRE support of the full normal-form game with enumeration of the entire strategy space (finite) in the worst-case.

Algorithm 1 PR-MRE Robust Double Oracle

Result: Probabilistically-Robust Minimax-Regret Equilibrium
Input: Θ, $\bar{\mu}$, δ, Initial population $\Pi^0 = (\mathcal{X}^0, \mathcal{Y}^0)$, Normal-form game payoff tensor $A_{\mathcal{X}\times\mathcal{Y}} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{Y}|\times|\Theta|}$
repeat
  $x^{\mathrm{k}}(\cdot), y^{\mathrm{k}} \leftarrow$ PRMRE-EQM-NF$\{A_{\Pi^{\mathrm{k}}}\}$
  for $i \in \{\mathrm{R}, \mathrm{B}\}$ **do**
    if i is B **then**
      $y^{\mathrm{k}+1} \leftarrow$ PRMRE-BR-NF$\{A_{\mathcal{X}^{\mathrm{k}}\times\mathcal{Y}}, x^{\mathrm{k}}(\cdot)\}$
      $\pi_i \leftarrow \mathrm{supp}(y^{\mathrm{k}+1}) \setminus \mathcal{Y}^{\mathrm{k}}$
    else
      $x^{\mathrm{k}+1}(\theta) \leftarrow \mathrm{BR}\{A_{\mathcal{X}\times\mathcal{Y}^{\mathrm{k}}}, y^{\mathrm{k}}\}$ for all $\theta \in \Theta$
      $\pi_i \leftarrow \left\{ \bigcup_{\theta\in\Theta} \mathrm{supp}(x^{\mathrm{k}+1}(\theta)) \right\} \setminus \mathcal{X}^{\mathrm{k}}$
    end if
    $\Pi_i^{\mathrm{k}+1} \leftarrow \Pi_i^{\mathrm{k}} \cup \{\pi_i\}$
  end for
until no novel best response exists for either player
return $x^{\mathrm{k}}(\cdot), y^{\mathrm{k}}$

4 PRMRE-PSRO: Learning Approximate Minimax-Regret Equilibria for Adversarial Team Games under Asymmetric Information

In Section 3, we proposed a novel equilibrium concept that combines the distribution-free robustness of minimax-regret reasoning with coarse probabilistic information about types from a nominal prior. The PR-MRE best-response operator for the uninformed Blue (see Definitions 2.5 and 2.4) computes a team best-response to the omniscient Red by minimizing worst-case regret of the Blue team strategy across subsets of the type-space weighted by coarse probabilistic information about their occurrence from the prior thus providing a robustness guarantee to strategic distribution shifts in terms of the best performance lower bound. For a finite normal-form game with asymmetric information, the PR-MRE best-response couples regrets across types resulting in a robust bilinear program for the PR-MRE equilibrium as compared to a bilinear program for the distribution-free MRE equilibrium and the linear-programming formulation for Bayesian Nash equilibria (see Section 2.1 for the various equilibrium concepts and their robustness properties).

PR-MRE (and MRE) does not admit a known decomposition to local CFR-style regret updates for an extensive form representation of a game with hidden types. Standard variants in the CFR literature recover approximately the Bayesian Nash equilibrium of the asymmetric information game and handle hidden information as known upto a prior (or chance). To compute robust team behavior in large adversarial team games with hidden information and potential strategic distribution shifts, we propose PRMRE-PSRO based on the robust double oracle procedure RDO for learning approximate PR-MRE for ATGs via reinforcement learning based best-responses.

Algorithm 2 details the pseudocode for PR-MRE Team PSRO. For the Red team (omniscient) and Blue team (uninformed) policy sets $\Pi_{\mathrm{R}}^{\mathrm{k}}$ and $\Pi_{\mathrm{B}}^{\mathrm{k}}$ at the kth iteration of the PRMRE-PSRO procedure, metagame payoff tensor $A_{\Pi_{\mathrm{R}}^{\mathrm{k}}\times\Pi_{\mathrm{B}}^{\mathrm{k}}} \in \mathbb{R}^{|\Pi_{\mathrm{R}}^{\mathrm{k}}|\times|\Pi_{\mathrm{B}}^{\mathrm{k}}|\times|\Theta|}$ is maintained with an explicit third dimension corresponding to the type-space. Maintaining a three-dimensional payoff tensor for the robust double oracle approach enables reasoning about the relative performance of policy sets across types as opposed to a two-dimensional empirical game matrix common in the PSRO literature Lanctot et al. (2017); Slumbers et al. (2023); McAleer et al. (2023b). For instance for Graph Capture-the-Flag, the three-dimensional payoff tensor measures performance of the Red defense and Blue attack-and-capture team policies across all possible flag hypotheses. The 3D payoff representation enables reasoning about regret across types for meta-equilibrium computation and policy set expansion. We now detail the robust population-based approach PRMRE-PSRO for learning approximate minimax-regret equilibria for adversarial team games under asymmetric information as follows,

Algorithm 2 PR-MRE Team PSRO

Result: Approximate PRMRE for the Adversarial Team Game under Asymmetric Information
Input: Θ, $\bar{\mu}$, δ, Initial population $\Pi_{\mathrm{R}}^{0}, \Pi_{\mathrm{B}}^{0}$

repeat {for $\mathrm{k} = 0, 1, \ldots$}
  Compute metagame payoff tensor $A_{\Pi_{\mathrm{R}}^{\mathrm{k}} \times \Pi_{\mathrm{B}}^{\mathrm{k}}} \in \mathbb{R}^{|\Pi_{\mathrm{R}}^{\mathrm{k}}| \times |\Pi_{\mathrm{B}}^{\mathrm{k}}| \times |\Theta|}$ with a third dimension for types
  $(x^{\mathrm{k}}(\cdot), y^{\mathrm{k}}) \leftarrow$ PRMRE-EQM-NF$\{A_{\Pi_{\mathrm{R}}^{\mathrm{k}} \times \Pi_{\mathrm{B}}^{\mathrm{k}}}\}$
  for m iterations **do**
    Update team best response π_{R} toward $\mathrm{BR}_{\mathrm{Val}}(y, \mathrm{Unif}(\Theta))$ via cooperative MARL
    Update team best response π_{B} toward $\mathrm{BR}_{\mathrm{MMR}}(x^{\mathrm{k}}(\cdot), \mathrm{Unif}(\Theta))$ via cooperative MARL
  $\Pi_{\mathrm{R}}^{\mathrm{k}+1} \leftarrow \Pi_{\mathrm{R}}^{\mathrm{k}} \cup \{\pi_{\mathrm{R}}\}$
  $\Pi_{\mathrm{B}}^{\mathrm{k}+1} \leftarrow \Pi_{\mathrm{B}}^{\mathrm{k}} \cup \{\pi_{\mathrm{B}}\}$
until max k number of iterations
Return: $(x^{\mathrm{k}}(\cdot), y^{\mathrm{k}})$

PRMRE PSRO (Algorithm 2): For PRMRE-PSRO, policy set expansion proceeds by best-responding to the opponent team meta-equilibrium as computed via the robust bilinear program PRMRE-BR-NF (described in Section 3.1) on the empirical metagame payoff tensor and a uniform type distribution. Team best-responses are computed via cooperative multi-agent reinforcement learning such as multi-agent proximal policy optimization (MAPPO) Yu et al. (2022). The omniscient Red team uses a value-based team best-response $\mathrm{BR}_{\mathrm{Val}}$ criterion that maximizes expected Red team payoff against an opponent distribution given by the Blue PR-MRE meta-equilibrium $y^{\mathrm{k}} \in \Delta(|\Pi_{\mathrm{B}}^{\mathrm{k}}|)$ at the latest PSRO iteration.

The uninformed Blue team best-responds to a uniform type distribution and the type-conditioned Red PR-MRE meta-equilibrium $x^{\mathrm{k}}(\cdot) \in \Delta(|\Pi_{\mathrm{R}}^{\mathrm{k}}|)$ with a minimax-regret stopping criteria $\mathrm{BR}_{\mathrm{MMR}}$ until the robust average regret of the training policy π_{B} (equivalent to worst-case regret over types for $\delta = 0$) falls below that of the Blue team meta-equilibrium y^{k} at the latest PSRO iteration. Type-wise regrets are tracked during Blue policy training as follows: $\varepsilon(x^{\mathrm{k}}(\cdot), \pi_{\mathrm{B}}; \theta) \approx \sup_{y^* \in \Delta(|\Pi_{\mathrm{B}}^{\mathrm{k}}|)} x^{\mathrm{k}}(\theta) A^{\mathrm{k}}(\theta) y^* - V(x^{\mathrm{k}}(\theta), \pi_{\mathrm{B}}; \theta)$ where $V(x^{\mathrm{k}}(\theta), \pi_{\mathrm{B}}; \theta)$ is the per-type Blue team payoff that is also tracked during training and the type-wise supremum against Red team meta-equilibrium $x^{\mathrm{k}}(\cdot)$ is approximated by restricting Blue team to the already discovered policies in the latest iteration of the metagame as $\sup_{\pi_{\mathrm{b}}^* \in \Pi_{\mathrm{b}}} V(x^{\mathrm{k}}(\cdot), \pi_{\mathrm{b}}^*; \theta) \approx \sup_{y^* \in \Delta(|\Pi_{\mathrm{B}}^{\mathrm{k}}|)} x^{\mathrm{k}}(\theta) A^{\mathrm{k}} y^*$. In summary, PRMRE-PSRO (i) uses a type-aware three-dimensional metagame representation (ii) to compute robust mixtures via the PR-MRE meta-equilibrium and (iii) a minimax-regret based stopping criteria for training Blue team policies robust to distribution shift.

BNE PSRO: BNE-PSRO for a nominal prior $\bar{\mu}(\cdot) \in \Delta(|\Theta|)$ is defined via the risk-neutral Bayesian Nash Equilibrium (BNE) as the meta-equilibrium for a metagame $\left\{\Pi_{\mathrm{R}}^{\mathrm{k}}, \Pi_{\mathrm{B}}^{\mathrm{k}}, A_{\Pi_{\mathrm{R}}^{\mathrm{k}} \times \Pi_{\mathrm{B}}^{\mathrm{k}} \times \Theta}\right\}$ along with risk-neutral $\mathrm{BR}_{\mathrm{Val}}(., \bar{\mu}(\cdot))$ best-response computation for both Red and Blue teams.

We now evaluate PRMRE-PSRO and BNE-PSRO on an example graph-structured ATG and compare the results. For our experiment, all Red and Blue team policies are parameterized as Graph Neural Networks (GNNs) that directly take in as input the graph-structured game state of the ATG. We note that the proposed PRMRE-PSRO is invariant to the choice of neural network architecture and provides a general methodology to train robust team policies to distribution shifts under hidden information.

4.1 Graph-structured Capture-the-Flag: Experiment and Results

Experiment 4.1. *Compare Blue team policies learned via* PRMRE-PSRO *and* BNE-PSRO *on an example graph-structured adversarial team game under asymmetric information, Graph Capture-the-Flag (Figure 2), in terms of robustness to distribution shift of the hidden flag hypotheses.*

For our experiment, we analyse one expansion iteration of the PSRO procedure to compare the robustness and behavioral properties of learned Blue team policies induced by the PRMRE and BNE meta-equilibria

$\mathrm{BR}_{\mathrm{MMR}}$ Team Best Response with Minimax-Regret Stopping

Track per-type regrets while training Blue team policy π_{B} as follows,

$$\varepsilon(x^{\mathrm{k}}(\cdot), \pi_{\mathrm{B}}; \theta) \approx \sup_{y^* \in \Delta(|\Pi_{\mathrm{B}}^{\mathrm{k}}|)} x^{\mathrm{k}}(\theta) A^{\mathrm{k}} y^* - V(x^{\mathrm{k}}(\theta), \pi_{\mathrm{B}}; \theta) \text{ for all } \theta \in \Theta. \tag{$\mathrm{BR}_{\mathrm{MMR}}$}$$

Proceed training until the robust average regret of π_{B} falls below that of the Blue team PR-MRE meta-equilibrium at the latest PSRO iteration i.e. $\inf_{\mu \in \mathcal{S}(\bar{\mu}, \delta)} \mu^\top \varepsilon(\pi_{\mathrm{B}}) < \inf_{\mu \in \mathcal{S}(\bar{\mu}, \delta)} \mu^\top \varepsilon(y^{\mathrm{k}})$.

starting from the same seed metagame on the example Graph Capture-the-Flag instance highlighted in Figure 2. The seed metagame for the PSRO runs is obtained as follows.

Seed Metagame: Blue team seed policies are learned by training against a Red team curriculum of progressively harder scripted flag defenders. The curriculum stages range from passive defenders to corridor sentry agents that patrol bottleneck nodes in the graph and chase any closing Blue attackers. Blue column 0 is a FH-1 (flag hypothesis 1, θ_1) specialist trained against the Red curriculum on flag hypothesis 1. Blue column 1 warmstarts from the column 0 checkpoint and is fine-tuned against uniformly sampled FH-0 and FH-1 flag hypotheses. Columns 2 and 3 are FH-0 (θ_0) specialists with a similar fine-tuning stage against uniformly sampled hypotheses. Rows correspond to learned Red team GNN policies trained against intermediate Blue checkpoints produced while training the columns.

Meta-equilibria and Expansion: Figures 6 and 7 highlight the evolution of the metagame for PR-MRE ($\delta = 0$) and BNE meta-equilibria respectively. Figure 6 describes one iteration of PRMRE-PSRO initialized from a 2×4 seed policy set for the teams as described above and the two panels correspond to the metagame states during PSRO iterations 0–1. For each panel, the heatmap on the top corresponds to θ_0 (Red flag on the left, $\mu(\theta_0) = 0.8$) and the bottom heatmap corresponds to θ_1 (Red flag on the right, $\mu(\theta_1) = 0.2$). Rows represent Red team (defender) policies and columns represent Blue team (attacker) policies whereas the cells indicate the Red team numeric score (negative of the Blue team score). Red team has knowledge of the true flag location whereas Blue team has knowledge only up to the nominal distribution before reaching the information frontier as marked in Figure 2. As shown in the graphic, the meta-equilibrium distribution over the Red team policy set conditions on the flag hypothesis such that $(\sigma(\pi_r^0|\theta_0), \sigma(\pi_r^1|\theta_0)) = (0.19, 0.81)$ and $(\sigma(\pi_r^0|\theta_1), \sigma(\pi_r^1|\theta_1)) = (0.90, 0.10)$. PR-MRE hedges the performance of the Blue team seed policy set across both flag hypothesis (types) and computes a robust mixture (minimum worst-case exploitability) over the four seed Blue team policies as $(\pi_b^0, \pi_b^1, \pi_b^2, \pi_b^3) = (0.23, 0, 0, 0.76)$ via the SDP relaxation provided in Section A.2. Note that even though π_b^0 has a drastic win-rate against Red for the majority type θ_0 under the nominal distribution (first panel, top heatmap, first column), PR-MRE favors π_b^3 over π_b^1 that has the best worst-case Blue win-rate across both types and lower regret. This is in contrast to BNE-PSRO (Figure 7) that concentrates all mass on π_b^0, the best Blue team policy on the majority type θ_0 (with nominal mass $\mu(\theta_0) = 0.8$) with an egregious win-rate against the minority type θ_1. Blue team policy trained against the robust Red team mixture results in a policy $\pi_{b,\mathrm{MRE}}^1$ (second panel, fourth column) with better worst-case Blue team win-rate (fourth column is Blue across both top and bottom heatmaps) as opposed to the BNE-trained Blue team best-response $\pi_{b,\mathrm{BNE}}^1$ (see Figure 7, second panel, fourth column) that loses poorly against the seed Red policy π_r^1 (second row) for θ_1. Figure 7 describes one iteration of BNE-PSRO initialized from the same 2×4 metagame as PRMRE-PSRO. Policies trained via BNE-PSRO are exploitable under distribution shifts as compared to the robust best-responses added in the PRMRE-PSRO iteration (see Figure 8 for the robustness experiment).

Robustness to Distribution Shift: The robust Blue team best-response trained against the PR-MRE mixture shows better robustness to type distribution shifts than the BNE-mixture trained best-response (see Figure 8 for the robustness experiment). In contrast, the BNE meta-equilibrium (Figure 7) concentrates probability mass on majority-hypothesis specialists, resulting in policies that are more vulnerable to type-distribution shifts.

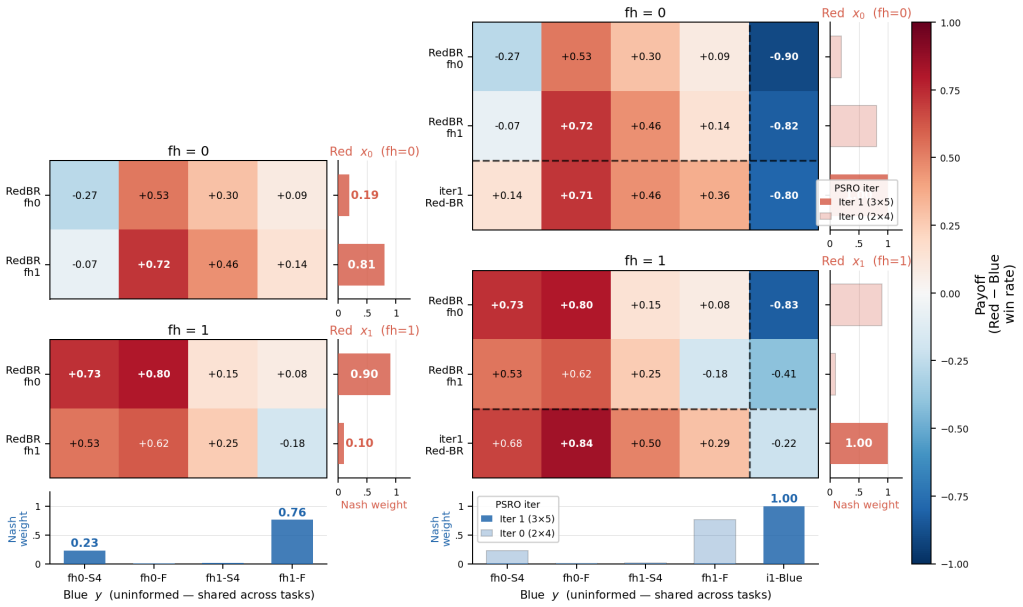

Figure 6: PRMRE-PSRO on the Graph CtF game under asymmetric information with nominal type distribution $(\mu(\theta_0), \mu(\theta_1)) = (0.8, 0.2)$. Rows correspond to Red-team policies and columns correspond to Blue-team policies. Cell values denote Red-team payoffs (negative values indicate Blue-team success). The three panels show the evolution of the metagame and PR-MRE meta-equilibrium over PSRO iterations. PR-MRE selects mixtures that hedge performance across both flag hypotheses rather than concentrating on policies that perform well only under the majority hypothesis. Consequently, the resulting Blue-team policies exhibit lower worst-case regret and improved robustness to type-distribution shifts.

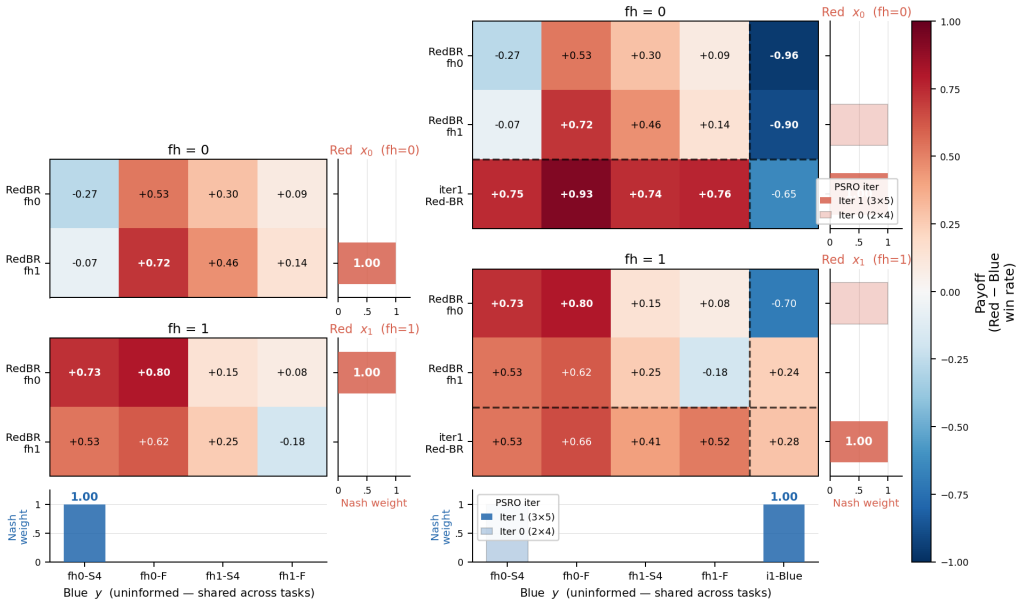

Figure 7: BNE-PSRO on the Graph CtF game under asymmetric information with nominal type distribution $(\mu(\theta_0), \mu(\theta_1)) = (0.8, 0.2)$. Rows, columns and cell values follow the same convention as Figure 6. The three panels show the evolution of the metagame and BNE meta-equilibrium over PSRO iterations. Unlike PRMRE-PSRO (Figure 6), the BNE meta-equilibrium concentrates probability mass on policies that perform well under the majority flag hypothesis, resulting in strategies that are more vulnerable to type-distribution shifts.

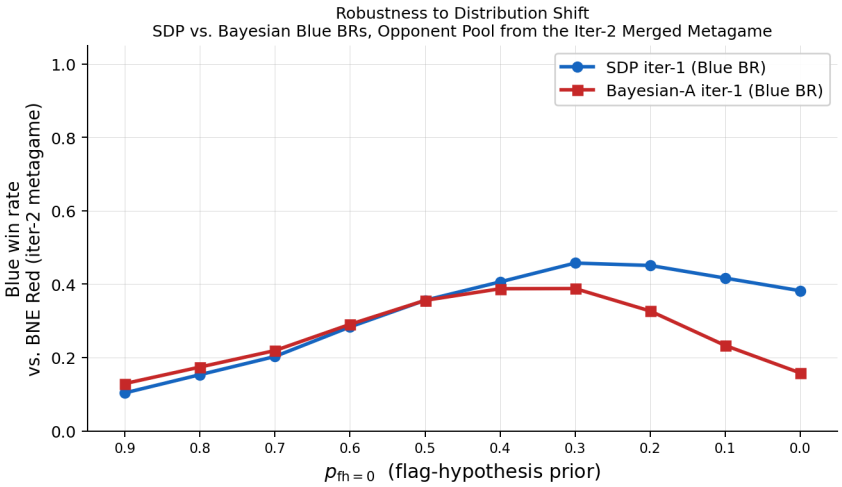

Figure 8: Performance-Robustness Curve for PRMRE − and BNE −trained Blue team policies for a nominal distribution of $(\mu(\theta_0), \mu(\theta_1)) = (0.8, 0.2)$. The x-axis on the above plot shows the probability mass for the hypothesis θ_0 under the test distribution, and the y-axis denotes the expected Blue team win-rate against flags sampled from the test distribution and learned Red team expert opponent policies conditioned on the sampled flag. For a sampled flag under a test-distribution, the conditioned Red opponent is the BNE under test-distribution over the concatenated Red team policy pool from the PSRO runs corresponding to Figures 6 and 7. From the above plot, PRMRE-trained Blue team policies overpower the BNE-trained team policies for $\mu(\theta_0) \leq 0.6$ demonstrating superior robustness to distribution shift across the type-simplex as the PRMRE-trained team policy win-rate flattens while the BNE-trained team policy win-rate deteriorates. Refer to Figure 9 for a behavioral interpretation of the learned Blue team policies.

Behavioral Analysis: Note the robust mixtures selected by PR-MRE in the metagame manifest behaviorally as scouting policies that avoid over-committing to the majority flag hypothesis. Figure 9 shows Blue team visitation density under learned team policies for the Graph Capture-the-Flag scenario. The PRMRE Blue team policy has a symmetric visitation behavior across the two corridors under the minority flag hypothesis unlike the BNE trained team policy which is biased towards the corridor corresponding to the majority flag hypothesis θ_0 suggesting that policies learned via PRMRE learned scouting tactics to resolve flag ambiguity before committing to a corridor leading to lower worst-case regret across the two hypothesis. A team policy that behaviorally learns to commit to a corridor (the majority flag hypothesis in this instance) will incur performance loss when the realized type is the minority type hypothesis. The lower regret behavior with symmetric visitation will be robust to distribution shifts because of learned scouting behavior as opposed to the team policy that has learned to commit to the majority flag hypothesis that will be brittle when the distribution shifts in mass to the formerly minority hypothesis (see Figure 8 for the performance-robustness curve of PRMRE and BNE under distribution shift). Moreover, in this instance of the Red opponent π_r^1 and minority type θ_1, the PRMRE trained team policy has a win-rate of **+0.41** versus a win-rate of **-0.24** for the BNE trained team policy suggesting favorable worst-case behavior due to lower regret across types.

5 Conclusions and Future Work

Adversarial team games with asymmetric information are susceptible to strategic type distribution shifts in settings where the omniscient opponent has knowledge of Nature's hidden type and potentially controls the type distribution or colludes with Nature such that it can condition its play on the Nature's realized type. A nominal prior distribution in such strategic multi-agent interactions is "trustable" only to a certain degree in the face of potential deception. Distributionally-robust methods provide a framework to address ambiguity in the type distribution via ambiguity sets but lack guarantees for shifts outside the ambiguity set

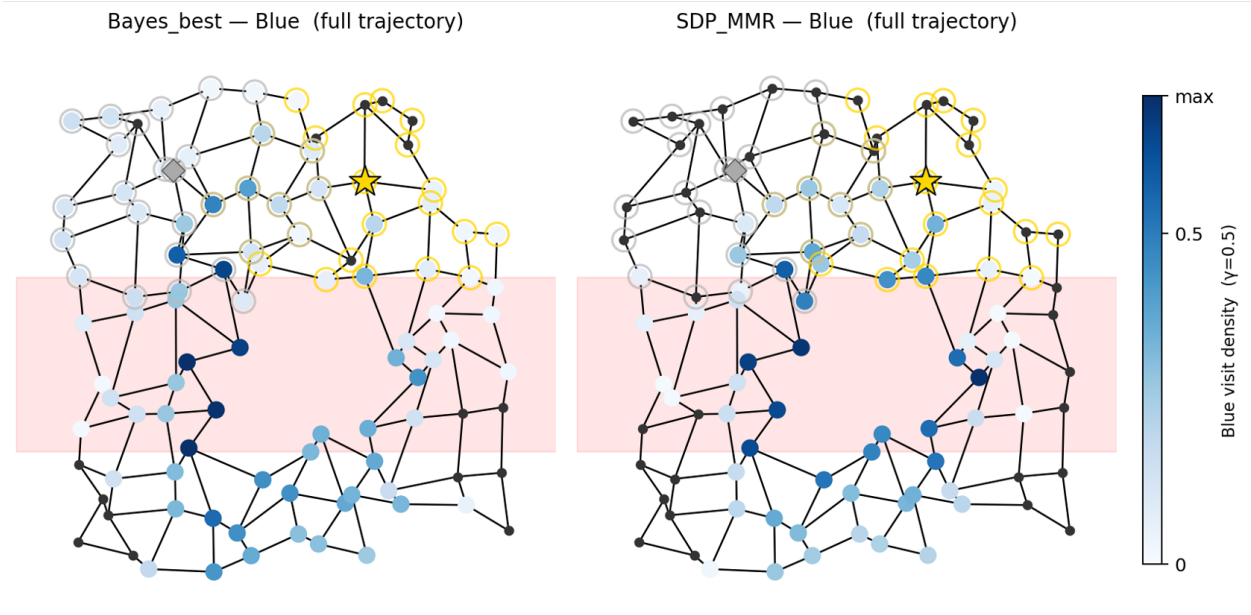

Figure 9: Behavioral comparison of BNE-PSRO (left) and PRMRE-PSRO (right) under the minority flag hypothesis. Node coloration indicates Blue-team visitation density. The BNE-trained policy is biased toward the corridor associated with the majority flag hypothesis, whereas the PRMRE-trained policy explores both corridors more symmetrically before committing to a search direction. This scouting behavior reduces exploitability across competing flag hypotheses and improves robustness to distribution shifts.

for instance strategic distribution shifts whereby Nature could be running an oracle procedure to compute a type distribution that places probability mass on high-regret types. Moreover, a worst-case approach that assumes an adversarial nature could be too conservative as it concentrates probability mass on the hardest type and loses performance under non-adversarial distribution shifts.

Motivated by the uniform lower bound interpretation of minimax reasoning, we propose probabilistically-robust minimax-regret equilibrium as a solution concept for adversarial team games under asymmetric information for robustness to strategic distribution shifts. PR-MRE best-response discounts regret over type-space subsets with coarse probabilistic information from the nominal prior distribution. This is in contrast to minimax-regret equilibrium that discards potentially useful probabilistic information about type occurrence by minimizing worst-case regret across all types. MRE is bound to lead to conservative performance on typical types when high-regret types are known to be rare for example rare flag hypothesis locations which require a remarkably different Blue team behavior for capture as compared to the typical flag hypotheses. PR-MRE is formulated as a robust bilinear program for finite normal-form games and is adapted as a meta-solver within a robust double-oracle based approach, PRMRE-PSRO to learn strategically robust team policies via reinforcement learning based best-responses. For an example graph capture-the-flag game, team strategies learned via PRMRE PSRO exhibit learned scouting behavior and enhanced robustness to distribution shift as compared to BNE PSRO.

Computation of sequential refinements of minimax-regret equilibria for treeplexes in extensive-form games (EFGs) remains an open question and a subject of future research. The nested optimization structure of MRE that characterizes deviation from type-conditioned optima and the bilinear program for MRE in normal-form games as opposed to a linear program for BNE hints at non-trivial computational complexity such that a naive approach to compute the MRE would first solve an EFG corresponding to each type. It remains an open question whether regret decomposition across infosets holds for MRE in the EFG representation as it does for the computation of BNE in the established counterfactual regret minimization literature.

References

Amir Ali Ahmadi and Georgina Hall. On the construction of converging hierarchies for polynomial optimization based on certificates of global positivity. *Math. Oper. Res.*, 44(4):1192–1207, November 2019. ISSN 0364-765X. doi: 10.1287/moor.2018.0962. URL https://doi.org/10.1287/moor.2018.0962.

Amir Ali Ahmadi and Jeffrey Zhang. Semidefinite programming and nash equilibria in bimatrix games. *INFORMS J. on Computing*, 33(2):607–628, May 2021. ISSN 1526-5528. doi: 10.1287/ijoc.2020.0960. URL https://doi.org/10.1287/ijoc.2020.0960.

Ioannis Anagnostides, Vaggos Chatziafratis, Fivos Kalogiannis, Manolis Vlatakis, and Stelios Stavroulakis. Efficiently computing nash equilibria in adversarial team markov games. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://openreview.net/forum?id=mjzm6btqgV.

Ioannis Anagnostides, Fivos Kalogiannis, Ioannis Panageas, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Stephen Mcaleer. Algorithms and complexity for computing nash equilibria in adversarial team games. *Games and Economic Behavior*, 157:138–152, 2026. doi: 10.1016/j.geb.2026.01.006.

Luca Carminati, Federico Cacciamani, Marco Ciccone, and Nicola Gatti. A marriage between adversarial team games and 2-player games: Enabling abstractions, no-regret learning, and subgame solving, 2022a. URL https://arxiv.org/abs/2206.09161.

Luca Carminati, Federico Cacciamani, Marco Ciccone, and Nicola Gatti. A marriage between adversarial team games and 2-player games: Enabling abstractions, no-regret learning, and subgame solving. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 2638–2657. PMLR, 17–23 Jul 2022b. URL https://proceedings.mlr.press/v162/carminati22a.html.

Andrea Celli and Nicola Gatti. Computational results for extensive-form adversarial team games. *Proc. Conf. AAAI Artif. Intell.*, 32(1), April 2018.

Darshan Chakrabarti, Julien Grand-Clément, and Christian Kroer. Extensive-form game solving via blackwell approachability on treeplexes. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9798331314385.

Tong Chen, Jean-Bernard Lasserre, Victor Magron, and Edouard Pauwels. A sublevel moment-sos hierarchy for polynomial optimization. *Comput. Optim. Appl.*, 81(1):31–66, January 2022. ISSN 0926-6003. doi: 10.1007/s10589-021-00325-z. URL https://doi.org/10.1007/s10589-021-00325-z.

John C. Harsanyi. Games with incomplete information played by "bayesian" players, i–iii part i. the basic model. *Management Science*, 14(3):159–182, 1967. doi: 10.1287/mnsc.14.3.159. URL https://doi.org/10.1287/mnsc.14.3.159.

John C. Harsanyi. Games with incomplete information played by "bayesian" players, i-iii. part ii. bayesian equilibrium points. *Management Science*, 14(5):320–334, 1968. ISSN 00251909, 15265501. URL http://www.jstor.org/stable/2628673.

Nathanael Hyafil and Craig Boutilier. Regret minimizing equilibria and mechanisms for games with strict type uncertainty. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pp. 268–277, Arlington, Virginia, USA, 2004. AUAI Press. ISBN 0974903906.

Fivos Kalogiannis, Jingming Yan, and Ioannis Panageas. Learning equilibria in adversarial team markov games: a nonconvex-hidden-concave min-max optimization problem. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9798331314385.

Christine Konicki, Mithun Chakraborty, and Michael P. Wellman. Computing perfect bayesian equilibria, with application to empirical game-theoretic analysis, 2026. URL https://arxiv.org/abs/2602.15233.

Marc Lanctot, Vinicius Zambaldi, Audrūnas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Pérolat, David Silver, and Thore Graepel. A unified game-theoretic approach to multiagent reinforcement learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 4193–4206, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Stephen McAleer, Gabriele Farina, Gaoyue Zhou, Mingzhi Wang, Yaodong Yang, and Tuomas Sandholm. Team-psro for learning approximate tmecor in large team games via cooperative reinforcement learning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023a. Curran Associates Inc.

Stephen McAleer, Gabriele Farina, Guanzhi Zhou, Mingzhi Wang, Yining Yang, and Tuomas Sandholm. Team-psro for learning approximate tmecor in large team games via cooperative reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 56872–56897, 2023b.

H Brendan McMahan, Geoffrey J Gordon, and Avrim Blum. Planning in the presence of cost functions controlled by an adversary. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pp. 536–543, 2003.

Yi Ouyang, Hamidreza Tavafoghi, and Demosthenis Teneketzis. Dynamic games with asymmetric information: Common information based perfect bayesian equilibria and sequential decomposition. *IEEE Transactions on Automatic Control*, 62(1):222–237, 2016.

Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

Ludovic Renou and Karl H Schlag. Minimax regret and strategic uncertainty. *Journal of Economic Theory*, 145(1):264–286, 2010.

Ludovic Renou and Karl H. Schlag. Implementation in minimax regret equilibrium. *Games and Economic Behavior*, 71(2):527–533, 2011. ISSN 0899-8256. doi: 10.1016/j.geb.2010.05.010. URL `https://doi.org/10.1016/j.geb.2010.05.010`.

Max Rudolph, Nathan Lichtlé, Sobhan Mohammadpour, Alexandre M. Bayen, J. Zico Kolter, Amy Zhang, Gabriele Farina, Eugene Vinitsky, and Samuel Sokota. Reevaluating policy gradient methods for imperfect-information games. In *International Conference on Learning Representations (ICLR)*, 2026.

Oliver Slumbers, David Henry Mguni, Stefano B. Blumberg, Stephen McAleer, Yaodong Yang, and Jun Wang. A game-theoretic framework for managing risk in multi-agent systems. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Jinheng Xiao, Chen Qiu, Yingying Xu, Jiajia Zhang, Shuhan Qi, and Xuan Wang. Solving equilibrium for adversarial team games utilizing fictitious team play with refined team plans. *Expert Systems with Applications*, 298:129496, 2026. ISSN 0957-4174. doi: 10.1016/j.eswa.2025.129496. URL `https://www.sciencedirect.com/science/article/pii/S0957417425031112`.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems*, 35:24611–24624, 2022.

Youzhi Zhang, Bo An, and Daniel Dajun Zeng. DAG-based column generation for adversarial team games. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 58563–58578. PMLR, 21–27 Jul 2024. URL `https://proceedings.mlr.press/v235/zhang24b.html`.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20, 2007.

A Appendix

A.1 Robustness Certificates for MRE and PR-MRE Best-response Operators

Proof of Lemma 1.

$$\mathbb{E}_{\theta\sim\mu}[V(\pi_r(\cdot),\pi_b;\theta] = \sum_{\theta\in\Theta}\mu(\theta)V(\pi_r(\cdot),\pi_b;\theta) = \sum_{\theta\in\Theta}\mu(\theta)\left[\sup_{\pi^*\in\Pi_b} V(\pi_r(\cdot),\pi^*;\theta) - \mathcal{E}(\pi_r(\cdot),\pi_b;\theta)\right], \tag{28}$$

$$= \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_b} V(\pi_r(\cdot),\pi^*;\theta) - \sum_{\theta\in\Theta}\mu(\theta)\mathcal{E}(\pi_r(\cdot),\pi_b;\theta), \tag{29}$$

$$\geq \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_b} V(\pi_r(\cdot),\pi^*;\theta) - \sup_{q\in\Delta(|\Theta|)}\underbrace{\left(\sum_{\theta\in\Theta} q(\theta)\mathcal{E}(\pi_r(\cdot),\pi_b;\theta)\right)}_{\bar{\mathcal{E}}(\pi_b,q)}, \tag{30}$$

$$= \underbrace{\sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_b} V(\pi_r(\cdot),\pi^*;\theta) - \sup_{\theta\in\Theta}\mathcal{E}(\pi_r(\cdot),\pi_b;\theta)}_{\mathcal{V}^{(-)}(\pi_b,\mu)}. \tag{31}$$

Therefore, $\mathbb{E}_{\theta\sim\mu}[V(\pi_r(\cdot),\pi_b;\theta] \geq \mathcal{V}^{(-)}(\pi_b,\mu) = \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_\mathrm{b}} V(\pi_r,\pi^*;\theta) - \mathcal{E}(\pi_r(\cdot),\pi_b,\Theta)$ for all $\pi_b\in\Pi_b$ and $\mu(\cdot)\in\Delta(|\Theta|)$. Moreover, the lower bound is tight such that equality holds for $q = e_{\mathrm{argsup}_{\theta\in\Theta}\mathcal{E}(\pi_r(\cdot),\pi_b;\theta)}$ where e_k is the k-th canonical vector in $\mathbb{R}^{|\Theta|}$. □

Proof of Lemma 4. We recall that,

$$\mathcal{V}_{\mathcal{S}}^{(-)}(\pi,\mu) = \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_\mathrm{b}} V(\pi_r,\pi^*;\theta) - \sup_{q\in\mathcal{S}(\bar{\mu},\delta)}\bar{\mathcal{E}}(\pi,q), \tag{32}$$

where $\bar{\mathcal{E}}(\pi,q) = \sum_{\theta\in\Theta} q(\theta)\mathcal{E}(\pi_r(\cdot),\pi;\theta)$. Now,

$$\mathcal{V}_{\mathcal{S}}^{(-)}(\pi_{\mathrm{PRMRE}},\mu) = \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_\mathrm{b}} V(\pi_r,\pi^*;\theta) - \sup_{q\in\mathcal{S}(\bar{\mu},\delta)}\bar{\mathcal{E}}(\pi_{\mathrm{PRMRE}},q), \tag{33}$$

$$= \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_\mathrm{b}} V(\pi_r,\pi^*;\theta) - \inf_{\pi'\in\Pi_b}\sup_{q\in\mathcal{S}(\bar{\mu},\delta)}\bar{\mathcal{E}}(\pi',q), \tag{34}$$

$$\geq \sum_{\theta\in\Theta}\mu(\theta)\sup_{\pi^*\in\Pi_\mathrm{b}} V(\pi_r,\pi^*;\theta) - \sup_{q\in\mathcal{S}(\bar{\mu},\delta)}\bar{\mathcal{E}}(\pi'',q), \tag{35}$$

$$= \mathcal{V}_{\mathcal{S}}^{(-)}(\pi'',\mu), \tag{36}$$

for all $\pi''\in\Pi_b$. The same holds for $\pi_{\mathrm{MRE}}\in\operatorname{arginf}_{\pi\in\Pi_b}\sup_{\theta\in\Theta}\mathcal{E}(\pi_r(\cdot),\pi;\theta)$ such that $\mathcal{V}_{\mathcal{S}}^{(-)}(\pi_{\mathrm{PRMRE}},\mu) \geq \mathcal{V}_{\mathcal{S}}^{(-)}(\pi_{\mathrm{MRE}},\mu)$. Hence proved that π_{PRMRE} provides the tightest lower bound on expected performance for type distributions restricted to the threat model. □

A.2 Rank-constrained Semidefinite Reformulation and Convex Relaxation

In this section, we lift the robust bilinear program PRMRE-EQM-NF to a semidefinite program (SDP) by reformulating the constraints via an introduced matrix positive semidefinite (psd) variable. Consider a symmetric psd matrix $Z(\theta)\in\mathbb{S}^+_{|\mathcal{X}|+|\mathcal{Y}|+1}$ such that,

$$Z(\theta) = \begin{bmatrix} 1 & x(\theta)^\top & y^\top \\ x(\theta) & X(\theta) & W(\theta)^\top \\ y & W(\theta) & Y \end{bmatrix} \succeq 0, \tag{37}$$

with the first row and column equal to the concatenated x and y vectors lying in the respective probability simplices and block matrices X and Y at the diagonals and an off-diagonal block matrix W. We follow a zero-based indexing convention to refer to the rows and columns of $Z(\theta)$ such that $Z_{1,1}(\theta) = X(\theta)$, $Z_{2,1}(\theta) = W$ and so on. Consider a reformulation of constraints (25)–(26) with respect to the block matrices of the psd matrix $Z(\theta)$,

$$x(\theta)A(\theta)e_j - \langle A(\theta), Z_{2,1}(\theta)\rangle \leq \{\lambda^\top \mathcal{A}_{\mathscr{L}} - \lambda_{\mathrm{p}}^\top + \lambda_{\mathrm{u}}\mathbf{1}\}_\theta \text{ for all } j \in [|\mathcal{Y}|] \text{ and } \theta \in \Theta, \tag{38}$$

and $-\langle A(\theta), Z_{2,1}(\theta)\rangle \geq -e_k A(\theta)y$ for all $k \in [|\mathcal{X}|]$ and $\theta \in \Theta$ where $\langle \cdot, \cdot \rangle$ is the trace operator such that $\langle A, B\rangle = \mathrm{trace}(AB)$.

The off-diagonal block matrix $Z_{2,1}(\theta) = W(\theta)$ has the interpretation of capturing the cross-terms between the $x(\theta)$ and y vectors such that $W(\theta) \approx yx(\theta)^\top$. This enables us to lift the bilinear term $x(\theta)^\top A(\theta)y$ and express it in terms of a block matrix of a positive semidefinite matrix variable as, $x(\theta)^\top A(\theta)y = \langle A(\theta), yx(\theta)^\top\rangle \approx \langle A(\theta), W(\theta)\rangle$. This is an instance of lifting a polynomial program into an SDP by the introduction of psd matrix variables (see Ahmadi & Hall (2019); Chen et al. (2022) for a survey on Moment-SoS hierarchy for polynomial optimization). Any rank-1 solution $\{Z(\theta_i)\}$ to the lifted program recovers the exact PR-MRE fixed-point solution $\{\mathbf{x}(\cdot), y\}$ to PRMRE-BR-NF. This follows from the eigen decomposition of a rank-1 positive semidefinite matrix and we omit a formal proof. See Proposition 2.1 in Ref. Ahmadi & Zhang (2021) for a related result on Nash equilibria in a bimatrix game with perfect information.

Remark on Asymmetric Information between Red and Blue Teams: Note that $Z_{2,0}(\theta) = y$ for all $\theta \in \Theta$ from the definition (37) of $Z(\theta)$. This is the robust Blue decision y under asymmetric information whereas Red can condition its play on the hidden type as reflected by $x(\theta)$ in $Z(\theta)$. Therefore, psd matrices $Z(\theta)$ are constrained to be consistent in their y decision for all types $\theta \in \Theta$ and this is added as an additional consistency constraint in the lifted program.

Convex Relaxation: Rank-constrained SDPs are an important class of non-convex problems that are generally NP-hard to solve Recht et al. (2010). The rank constraint appears because of the cross-terms in the bilinear program that was lifted via the use of matrix PSD variables. We relax the rank constraint into a convex semidefinite program for our application using McCormick cuts. *McCormick Cuts and Valid Inequalities*: We begin our exposition by stating the following proposition that characterizes the desired rank-one correlations between the lifted matrix variables and the original decision variables.

Proposition 1. *If* $\mathrm{rank}(Z(\theta)) = 1$, $X(\theta) = x(\theta)x^\top(\theta)$, $Y = yy^\top$ *and* $W(\theta) = yx^\top(\theta)$.

Since we relax the rank-constraint, we add valid inequalities in the form of McCormick cuts to the block matrices $X(\theta)$, Y and $W(\theta)$ to encode the correlation between the cross-terms as stated in Proposition 1. This is a valid approach since all feasible rank-1 solutions to the rank-constrained program satisfy these inequalities and is a way of pruning out high-rank solutions that do not obey the correlational structure possessed by low-rank solutions. For example, we add the following set of inequalities for $W_{i,j}(\theta)$, since it is supposed to be a proxy for $y_i x_j$ such that $W_{i,j} \approx y_i x_j$,

$$W_{i,j}(\theta) \leq x_j(\theta), W_{i,j}(\theta) \leq y_i, \tag{39}$$

$$W_{i,j}(\theta) \geq x_j(\theta) + y_i - 1 \tag{40}$$

for all $i \in [|m|]$ and $j \in [|n|]$. We add similar cuts for the i, jth terms ($i \neq j$) of the block-matrices $X(\theta)$ and Y which are supposed to be proxies for $x_i x_j$ and $y_i y_j$ respectively. For the diagonal terms of the $X(\theta)$ and Y blocks, we enforce constraints of the form $0 \leq X_{ii}(\theta) \leq x_i(\theta)$ and $0 \leq Y_{ii} \leq y_i(\theta)$ which follow from $x(\theta)$ and y belonging to probability simplices.

The solution $\{Z(\theta_i)\}$ obtained from the relaxed program via the valid inequalities provided by McCormick cuts is not guaranteed to be rank-1 and the relaxation might not be lossless. Therefore, as part of our method, we project the obtained solution for $Z(\theta)$'s on the manifold of rank-1 matrices. From the classical Eckart–Young–Mirsky theorem, the optimal rank-r approximant of a $m \times n$ rectangular matrix M is given by the largest r singular values of M and the corresponding eigenvectors. In our method, we obtain rank-1 approximations for $\{Z(\theta_i)\}$ via the largest eigenvector for each θ_i. Let $v_{\max}(Z(\theta_i)) \in \mathbb{R}^{|\mathcal{X}|+|\mathcal{Y}|+1}$ denote the eigenvector corresponding to the largest singular value of $Z(\theta_i)$. We re-normalize $v_{\max}(\theta_i)$ such that

the first-entry becomes unity as $v_{\max}(\theta_i) = [1, x(\theta_i), y]^\top$ and obtain $x(\theta_i)$ and y. Note that the $x(\theta)$ and y obtained in this fashion might not lie in their respective probability simplices (due to negative entries for example) and will require an additional projection step. In our implementation, we clip the negative entries at 0 and re-normalize such that the sum of all entries equals 1.

A.3 Additional Illustrative Example

Consider the following $2 \times 5 \times 3$ normal-form game such that rows corresponding to red action r_0 are the same as the $1 \times 5 \times 3$ game from Table 3. Red action r_1 is an additional *punisher* action that punishes Blue specialists b_0, b_1 and b_2 on their specialized types respectively and also the generalists b_3, b_4 across all types hence modifying the regret and payoff landscape for the game. We note that b_0, b_1 and b_2 retain their specialist roles for θ_0, θ_1 and θ_2 respectively as before as can be verified from the payoff values in the game below. Unlike the main-text example, the informed Red player now has multiple strategic choices, so that both the robust Blue best-responses and the corresponding informed Red best-responses differ across equilibrium fixed-points.

Table 4: Example 2×5×3 game. Columns $\{b_0 \cdots b_4\}$ are Blue actions, rows $\{r_0, r_1\}$ correspond to Red actions and a total of three 2×5 matrices stacked one on top of another along a third type-axis for $\theta \in \{\theta_0, \theta_1, \theta_2\}$ with a nominal distribution $\bar{\mu} = (0.25, 0.70, 0.05)$ such that each cell shows the Blue payoff $V(r, b; \theta)$ for the (r, b, θ) matchup.

	b_0	b_1	b_2	b_3	b_4	
r_0	1.5	0.6	-0.5	0.4	0.7	$\theta_0, 0.25$
r_1	0.7	0.1	-0.5	0.1	0.1	
r_0	0.3	1.0	-0.8	0.8	0.8	$\theta_1, 0.70$
r_1	-0.1	0.3	-0.8	0.2	-0.1	
r_0	-0.6	-0.8	1.5	0.5	-0.1	$\theta_2, 0.05$
r_1	-0.4	-0.8	0.5	0.0	-0.3	

BNE, DRO-Ball, MaxMin and DRO-MaxMin are all computed via linear-programs as before whereas MRE and PR-MRE ($\delta = 0.15$) are computed via semidefinite relaxations to the corresponding bilinear and robust bilinear programs as derived in Section 3.1. The evaluation protocol for the winner-map, payoff difference heatmaps and the metrics table is the same as before such that the computed Blue decisions are evaluated against the Red BNE corresponding to the test type distribution that probes over the entire type-simplex. The robust Blue decisions corresponding to the above computed fixed-points are: BNE $(0, 1, 0, 0, 0)$, DRO-Ball $(0, 0, 0, 1, 0)$, MaxMin $(0.05, 0, 0.14, 0.81, 0)$, DRO-MaxMin $(0.11, 0, 0, 0.89, 0)$, MRE $(0.42, 0, 0.29, 0.29, 0)$ and PR-MRE $(0.56, 0, 0, 0.45, 0)$. We note that no Blue decision has b_4 in its support including PR-MRE in contrast to the $1 \times 5 \times 3$ game from before. This is due to the additional punisher row that penalizes b_4 more drastically than b_3 over typical types θ_0 and θ_1 as seen in Table 4. The computed MRE Blue decision allocates meaningful probability mass on the rare-type specialist b_2 same as before, at the expense of catastrophic performance on the typical set $\mu(\theta_2) \leq 0.15$ (-0.222 worst-case and -0.022 average payoff from Figure 11), as opposed to the computed PR-MRE Blue that discounts regret incurred on the rare-type and hedges across the trapezoidal band representing the typical subset of the type-simplex under confidence-level $\delta = 0.15$ (see Figure 10). From the winner-map in Figure 10, PR-MRE dominates the trapezoidal band and has the lowest worst-case regret and the highest average-case payoff (0.200) for type distributions chosen at random from the structured type ambiguity set as opposed to DRO-MaxMin which has a higher worst-case payoff (0.135) as compared PR-MRE (-0.005) but loses out in the average-case. The regret-based objective of PR-MRE provides a performance guarantee in the form of a point-wise uniform lower bound (see Lemma 4)

for all type distributions belonging to the structured ambiguity set that results in a favorable average-case performance over the trapezoidal band in the example game.

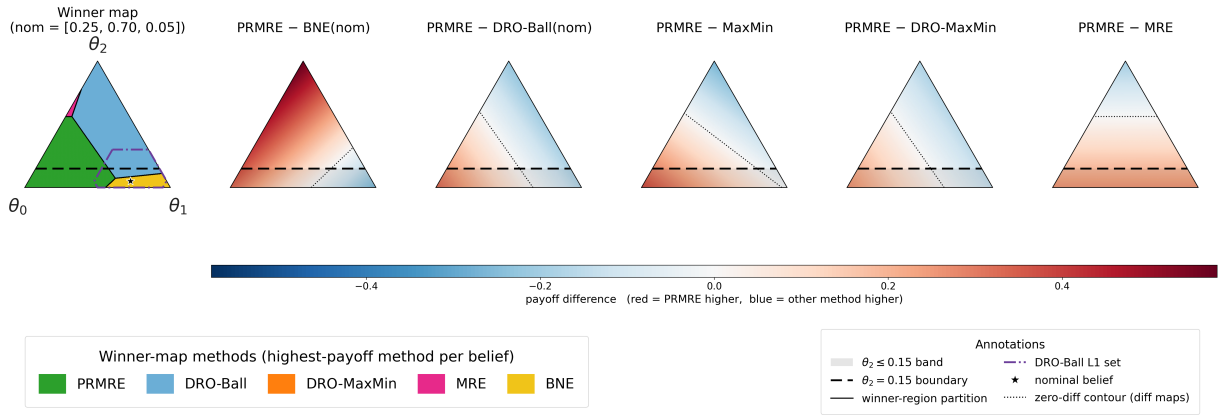

Figure 10: Comparison of PRMRE ($\delta = 0.15$) with BNE, DRO-Ball, MaxMin, DRO-MaxMin and MRE for the synthetic three-type game in Table 4: winner map (first triangle) and payoff difference heatmaps (remaining triangles). Triangles represent three-dimensional simplex in barycentric coordinates with vertices corresponding to full probability mass on the respective types. Black dotted line and the trapezoidal band beneath it represent the high-confidence under nominal structured type-ambiguity set $\mu(\theta_2) \leq 0.15$, the star in the winner map corresponds to the nominal distribution $\bar{\mu} = (0.25, 0.70, 0.05)$ and the dotted hexagon centered at the star corresponds to an $L1$-ball with radius $\rho = 0.5$ around it. For the trapezoidal band, PR-MRE dominates all other methods which can be seen from the green volume share in the winner-map and the red volume share in the payoff difference heatmaps. BNE and DRO-Ball perform better close to the nominal (θ_1-corner, mass 0.70 under nominal) but lose out on performance to PR-MRE for a type distribution chosen at random within the trapezoidal band. PR-MRE has the best average

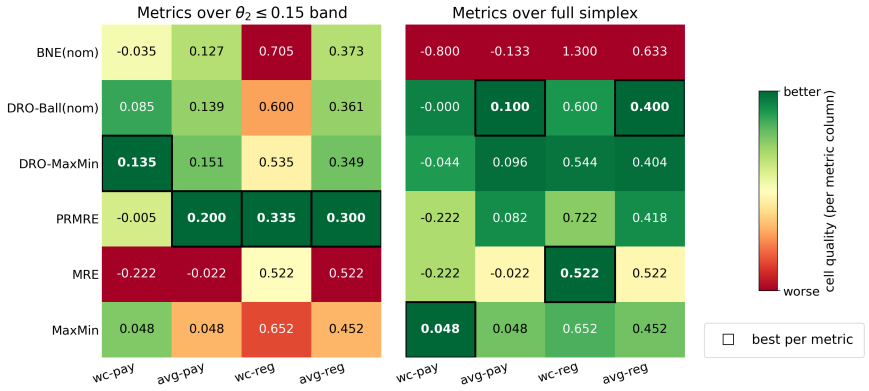

Figure 11: Comparison of PRMRE ($\delta = 0.15$) with BNE, DRO-Ball, MaxMin, DRO-MaxMin and MRE for the synthetic three-type game in Table 4 with respect to average and worst-case, payoff and regret over the structured ambiguity set $\mu(\theta_2) \leq 0.15$ and the entire type-simplex. Over the $\mu(\theta_2) \leq 0.15$ trapezoidal band (see Figure 10), PR-MRE has the best worst-case regret as compared to any other method. DRO-MaxMin on the other hand beats PR-MRE on worst-case payoff over the $\mu(\theta_2) \leq 0.15$ band which is consistent with its payoff-based objective (0.135 vs -0.005) but loses average-case performance for a type distribution chosen at random from the trapezoidal band such that PR-MRE attains the highest average payoff (0.200) as compared to any other method (0.151 for DRO-MaxMin). The regret-based objective of the PR-MRE best-response provides the best point-wise lower bound on payoff for type distributions belonging to the structured ambiguity set (as stated in Lemma 4) which in this example game manifests as a superior average-case performance for a type-distribution chosen at random from the $\mu(\theta_2) \leq 0.15$ structured ambiguity set.

A.4 GNN-MAPPO PSRO Training Details

A.4.1 Graph Neural Network Architecture

Table 5: Node feature vector $x_v \in \mathbb{R}^{13}$ (`obs_version`=3). Indices 6–10 are populated only on the acting agent's node.

#	Feature	#	Feature
0	`is_ego`	7	dist. to opponent flag (ego)
1	`is_frontier`	8	`enemy_flag_known` (ego)
2	frontier resolution mass	9	min. enemy dist. to ego (ego)
3	`is_regular`	10	min. teammate dist. to ego (ego)
4	opponent flag present here	11	structural (graph) degree
5	hop distance to ego node	12	visibility bit
6	dist. to own flag (ego)		

A.4.2 PPO Configuration

Table 6: Policy and PPO/MAPPO training configuration. Best responses are warm-started from a v4-hybrid checkpoint (shared MPNN transfers; per-agent heads re-initialized).

Hidden width / embedding D	64	Optimizer	Adam
MPNN layers L	3	Learning rate	2.5×10^{-4}
Aggregation	sum (+self-loop)	PPO clip range	0.2
Ego-subgraph radius K	6 hops	GAE λ / γ	0.95 / 0.99
Node feature dim	13	Rollout length n_{steps}	$2048^{\dagger}$
Team size	2v2	Minibatch size	128
Attention	disabled	PPO epochs	$2^{\dagger}$
Algorithm	MAPPO (CTDE)	Value coef. c_v	$0.1^{\dagger}$
Steps per best response	4×10^6	Entropy coef.	0.0