COMPUTING EQUILIBRIA IN GAMES WITH STOCHASTIC ACTION SETS

Anonymous authors
Paper under double-blind review

ABSTRACT

The study of learning in games typically assumes that each player always has access to all of their actions. However, in many practical scenarios, players' available actions might be restricted due to exogenous stochasticity. To model this setting, for a game $\mathcal{G}_{\text{orig}}$ with action set A_i for each player i, we introduce the corresponding Game with Stochastic Action Sets (GSAS) which is parametrized by a probability distribution over the players' set of possible action subsets $\mathcal{S}_i \subseteq 2^{A_i} \setminus \{\varnothing\}$. In a GSAS, players' strategies and Nash equilibria (NE) admit prohibitively large representations, and existing algorithms for NE computation scale poorly. Under the assumption that action availabilities are independent between players, we show that NE in two-player zero-sum (2p0s) GSAS can be approximately represented by a compact vector of size $|A_i|$, overcoming the naïve exponential-sized representation. Computationally, we introduce an algorithm that minimizes ranking regret, converging to NE with high probability in 2p0s-GSAS with rate $O(\sqrt{|A_i| \log |A_i| / T})$ for time horizon T. Finally, using the iterates of our algorithm, we develop a stochastic approximation procedure to recover compactly represented NE.

1 INTRODUCTION

A common assumption in game theory is that the players always have access to all of their actions. However, in many multi-agent systems, a player's actions might be randomly restricted at certain times. For instance, a trader might only have access to a subset of options on any given day due to exogenous constraints outside their control. A naval ship might be unable to enter or patrol part of its usual operating area due to the presence of civilians or inclement weather. In these scenarios, a player has access to only a subset of their action set and must choose among the available actions, which could change each time the game is played.

To formally model this phenomenon, we introduce the class of Games with Stochastic Action Sets (GSAS). A GSAS $\mathcal{G}$ begins with a 'base' normal-form game $\mathcal{G}_{\text{orig}}$ with finite action sets A_i for each player i. We further equip $\mathcal{G}$ with a probability distribution $\rho \in \Delta(\mathcal{S})$, where $\mathcal{S} \coloneqq \bigtimes_i \mathcal{S}_i$ and each $\mathcal{S}_i$ is the set of all possible action subsets $\mathcal{S}_i \subseteq 2^{A_i} \setminus \{\varnothing\}$ for each player i. At each round, Nature draws a joint availability set $S \in \mathcal{S}$ from ρ. Each player observes only their own availability set $S_i \subseteq A_i$ and is restricted to selecting an action $a_i \in S_i$.

In this paper, we are interested in obtaining Nash equilibria (NE) of GSAS, which are given by strategy profiles from which no player has an incentive to unilaterally deviate. However, existing strategy formalisms require that players select, for each possible action subset, a distribution over available actions to play. For instance, formulating a GSAS as a Bayesian game (Harsanyi, 1968) requires modeling action availabilities as types, where strategies are mappings from types to action distributions. Furthermore, since some actions may be unavailable, the NE structure of a GSAS $\mathcal{G}$ can differ significantly from $\mathcal{G}_{\text{orig}}$ (cf. Example 3.5).

This raises a key challenge in understanding the properties of NE in GSAS. In particular, the naïve specification of a player's strategy might require a mapping from action availabilities to distributions over actions, whose domain could be exponentially large in the worst case. Thus, even *specifying* a NE could be challenging, raising concerns about NE as an implementable solution concept and motivating a study of conditions under which NE can be *compactly represented* in GSAS. This

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

type of result has several seminal precedents: Kuhn's theorem (Kuhn, 1953) showed that behavioral strategies (as opposed to exponentially sized mixed strategies) suffice to capture NE in extensive-form games of perfect recall; Shapley (1953) showed that every two-player zero-sum Markov game admits a Markovian equilibrium.

We next turn our attention to the complementary task of *computing* NE in the special but important case of two-player zero-sum (2p0s) GSAS. A standard paradigm for equilibrium computation in normal-form games utilizes a canonical connection between no-regret learning and game-theoretic equilibria, which states that no-regret learners converge in time-average to approximate equilibria (Cesa-Bianchi & Lugosi, 2006). This connection has led to the design of decentralized algorithms that can efficiently compute NE in standard two-player zero-sum games; see, e.g., (Freund & Schapire, 1999; Rakhlin & Sridharan, 2013; Daskalakis et al., 2011; Syrgkanis et al., 2015; Daskalakis et al., 2021). However, when attempting to solve a 2p0s-GSAS directly, the connection between 'standard' no-external-regret learning and equilibria breaks down since players' action sets can vary between rounds and the competing action may not be available. To circumvent this, one could also consider special cases of GSAS that allow for the application of existing techniques. For instance, when ρ is fully known to the solver, the game can be expanded into a Bayesian game. Nevertheless, even in this special case, existing solvers do not scale well since the number of possible action sets could be large. This motivates the design of computational techniques that can efficiently compute *compact representations* of NE, even without specifying ρ explicitly.

Our contributions. The primary contribution of our work is the design and analysis of a complete procedure that computes compact representations of approximate NE in 2p0s-GSAS. This requires three key aspects. First, we **formalize and study properties of equilibria** in 2p0s-GSAS under the assumption that players' action availabilities are independent. We show that players' strategy sets can be restricted to the set of 'implementable' marginals which contains all equilibria of the GSAS. We also establish that any implementable equilibrium marginal can be approximately represented as a vector w of size $|A_i|$. Second, we introduce an algorithm called CAT-FTPL that **minimizes a suitable notion of regret** called *ranking regret* in GSAS, and converges to approximate NE in 2p0s-GSAS. We prove a sublinear bound on expected pathwise ranking regret and derive a corresponding high-probability upper-tail bound. Third, using *only* the CAT-FTPL iterates, we introduce a stochastic approximation-based method to **extract a compact vector** w that represents an approximate NE. Finally, we investigate the empirical efficacy of our proposed method in 2p0s-GSAS. Our experimental results demonstrate the scalability of our method compared to standard game solvers and show that it exhibits approximate convergence to compact equilibria in large games.

2 Related Work

Due to space constraints, in this section we discuss related work that directly concerns our techniques and analysis. Additional discussions on related work and on the connection to Bayesian games are deferred to Appendix A.

Sleeping bandits and stochastic actions. Our work is closely related to the *sleeping bandit* setting in the multi-armed bandit literature, which studies regret minimization when arms are available stochastically or adversarially (Auer et al., 2002; Blum & Mansour, 2007; Kanade et al., 2009; Kleinberg et al., 2010; Kanade & Steinke, 2014; Saha et al., 2020; Nguyen & Mehta, 2024; Schneider & Zimmert, 2023). These techniques have been applied to online combinatorial optimization (Neu & Valko, 2014; Kale et al., 2016) and reinforcement learning (Drago et al., 2025). Finally, Boutilier et al. (2018) studied MDPs with stochastic action availabilities, obtaining efficient algorithms. In our work, we focus on computing compact Nash equilibria in a multi-agent setting, which requires additional game-theoretic insights beyond those studied in prior works.

Games with exponential action sets. Panageas et al. (2023) and Dong et al. (2023) studied semi-bandit learning in congestion games which admit exponentially large action sets, while Farina et al. (2022) and Kontogiannis et al. (2025) studied regret minimization in *polyhedral* games (with underlying combinatorial structure). The emphasis of these works is on deriving strong regret guarantees, whereas we additionally analyze efficient equilibrium *representation*.

Connection to Bayesian games. By interpreting each player's availability set S_i as their private type drawn from the marginal distribution induced by ρ, a GSAS can be viewed as a variant of Bayesian games (Harsanyi, 1968). However, while a GSAS shares some structure with Bayesian games, they are distinct in several ways, introducing new challenges and motivating specialized algorithms: (i) classical Bayesian game solvers assume common knowledge of payoff functions, possible types, and a prior distribution over types, whereas a GSAS allows for ρ to be unknown to the solver; and (ii) algorithms for computing Bayesian NE have rates that depend on the number of types (Fujii, 2025; Dagan et al., 2024; Peng & Rubinstein, 2024).

3 Games with Stochastic Action Sets

Notation. Denote the n-dimensional nonnegative orthant by $\mathbb{R}^n_{\geq 0}$. For a finite set S, let $\Delta(S)$ be the probability simplex $\{x \in \mathbb{R}^{|S|}_{\geq 0} \mid \sum_{s\in S} x(s) = 1\}$, such that if $y \in \Delta(S), s \in S$, $y(s)$ is the probability that item s is selected.

Consider an n-player normal/strategic-form game $\mathcal{G}_{\text{orig}}$ with finite action set A_i, joint action space $A = A_1 \times \cdots \times A_n$, and payoff functions scaled as $u_i : A \to [-1, 1]$ for player i. In line with prevailing conventions, we let $a = (a_1, \ldots, a_n)$ be an action profile, and as shorthand $u_i(a) = u_i(a_1, \ldots, a_n)$. We also denote by $-i$ the set of players other than i, such that $u_j(a'_i, a_{-i}) = u_j(a_1, \ldots, a'_i, \ldots, a_n)$.

Definition 3.1 (GSAS)**.** *Given a game $\mathcal{G}_{\text{orig}} = (A_1, \ldots, A_n, u_1, \ldots, u_n)$, let $\mathcal{S}_i \subseteq 2^{A_i} \setminus \{\varnothing\}$, let $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$, and let $\rho \in \Delta(\mathcal{S})$ be a distribution over elements $\mathcal{S}$, such that $\rho(S)$ gives the probability that the availability set S is observed for $S \in \mathcal{S}$. Then, a normal-form **G**ame with **S**tochastic **A**ction **S**ets (*GSAS*) is given by the tuple $\mathcal{G} = (\mathcal{G}_{\text{orig}}, \mathcal{S}, \rho)$.*

Definition 3.2 (2p0s-GSAS)**.** *A two-player zero-sum GSAS (2p0s-GSAS) $\mathcal{G}$ is one where $\mathcal{G}_{\text{orig}}$ is two-player zero-sum, i.e., $n = 2, u_1(a) = -u_2(a)$ for all action profiles $a \in A$.*

Informally, a GSAS proceeds as follows. At the start of the game, each player privately receives their availability set $S_i \in \mathcal{S}_i$ from Nature based on ρ. Each then plays an action $a_i \in S_i$ simultaneously and receives a payoff $u_i(a)$ based on the action profile $a \in A$. While each player i observes their availability set S_i and might have full knowledge of ρ, they do not observe their opponents' availability sets S_{-i}. In the remainder of the paper, unless otherwise stated, we make a technical assumption that facilitates our analysis going forward. A similar assumption is often made to facilitate price-of-anarchy analyses in Bayesian games (Fujii, 2025; Roughgarden, 2015; Syrgkanis & Tardos, 2013; Syrgkanis, 2012).

Assumption 3.3. *$\rho(S) = \prod_{i=1}^n \rho_i(S_i)$ for some probability distributions $\rho_i : \mathcal{S}_i \to [0, 1]$, i.e., the availability of actions is independent across players.*

A *pure strategy* in a GSAS is a deterministic mapping $\pi_i : \mathcal{S}_i \to A_i$ where $\pi_i(S_i) \in S_i$. More generally, we define a *mixed strategy* (or simply *strategy*) for player i as a mapping $\pi_i : \mathcal{S}_i \to \Delta(A_i)$ such that $\text{supp}(\pi_i(S_i)) \subseteq S_i$. A player's strategy gives, for every possible observed subset of actions, a distribution of actions corresponding to the observed action subset[1]. Note that even under Assumption 3.3, the explicit representation of the π_i's could be exponentially large, and dealing with this is a key contribution of our work.

Given a joint availability set $S \in \mathcal{S}$, we denote by π the joint strategy of all players and by $\pi(a \mid S) = \prod_{i\in[n]} \pi_i(a_i \mid S_i)$ the probability of an action profile a for every $a \in A$. Player i's expected payoff is

$$U_i(\pi) = \mathbb{E}_{S\sim\rho}\left[\mathbb{E}_{a\sim\pi(S)}[u_i(a)]\right], \tag{1}$$

where the inner expectation is over the actions sampled independently according to each player's strategy given their action sets, and the outer expectation is over the availability set S drawn from ρ. We also define the expected payoff of a specific action a_i of player i with respect to the ensemble of the opponents' strategies π_{-i} by

$$U_i(a_i; \pi_{-i}) = \mathbb{E}_{S\sim\rho}\left[\mathbb{E}_{a_{-i}\sim\pi_{-i}(S_{-i})}[u_i(a_i, a_{-i})]\right]. \tag{2}$$

[1]Mixed strategies in Bayesian games are classically defined by *distributions over pure strategies*. It follows from classical results (Harsanyi, 1968) that this is strategically equivalent (for most equilibrium computation purposes, including ours) to our simpler definition of per-type conditional distribution over actions.

Definition 3.4 (ϵ-Nash equilibrium (ϵ-NE))**.** *For $\epsilon \geq 0$, a strategy profile $\pi = (\pi_1, \ldots, \pi_n)$ is an* ϵ-Nash equilibrium *if no player can improve their expected payoff by more than ϵ by unilaterally deviating from π, i.e., for every player $i \in [n]$ and for every alternative strategy π_i',*

$$U_i(\pi_i, \pi_{-i}) \geq U_i(\pi_i', \pi_{-i}) - \epsilon. \tag{3}$$

A 0-NE is a Nash equilibrium*, i.e., no player can strictly benefit by unilaterally deviating.*

These definitions extend the classical normal-form Nash equilibrium to GSAS and are consistent with Bayesian games and Bayesian Nash equilibria (Harsanyi, 1968). While it is tempting to assume that solutions to GSAS may be easily extracted from the underlying $\mathcal{G}_{\text{orig}}$, we show in the following example that this is generally *untrue*. Indeed, the presence of stochastic action availabilities for a player can significantly change the set of NE, even if other players do not face stochastic action sets.

Example 3.5. *Let $\mathcal{G}_{\text{orig}}$ be a matching pennies game with actions H and T. Define a GSAS by setting $u_1(H,H) = u_1(T,T) = 1$, $u_1(H,T) = u_1(T,H) = -1$, and $u_2 = -u_1$. Suppose that $\mathcal{S}_2 = \{\{H,T\}\}$, i.e., all actions are always available for player 2, but $\mathcal{S}_1 = \{S_{11} = \{H,T\}, S_{12} = \{H\}\}$ with $\rho_1(S_{11}) = \lambda$ and $\rho_1(S_{12}) = 1 - \lambda$. If $\lambda = 1$, we have the regular matching pennies game, while for $\lambda = 0$, player 1's only available action is H. The set of NE is as follows:*

- *(i) $\lambda > 0.5$: player 1 sets $\pi_1(T \mid S_{11}) = 0.5/\lambda$ which "effectively" plays T with probability 0.5, while player 2 plays uniformly at random: this is essentially the standard matching pennies;*
- *(ii) $0 < \lambda < 0.5$: player 1 sets $\pi_1(T \mid S_{11}) = 1$ and player 2 plays T deterministically; and*
- *(iii) $\lambda = 0.5$: player 1 sets $\pi_1(T \mid S_{11}) = 1$ and player 2 plays T with probability in $[0.5, 1]$.*

Intuitively, when $0 < \lambda < 0.5$, player 1 can never play T frequently enough and player 2 takes advantage of this by playing T deterministically, while player 1 plays T whenever possible. When $\lambda > 0.5$, player 1 compensates by playing T with higher probability when S_{11} is offered, i.e., T is available, since they must play H all the time if S_{12} is drawn.

4 Properties of Equilibria in GSAS

For a player i, the marginal distribution over actions $a_i \in A_i$ induced by any of their strategies π_i is $\mathbb{P}[a_i; \rho_i, \pi_i] = \sum_{S_i \in \mathcal{S}_i} (\rho_i(S_i)\pi_i(a_i \mid S_i)\mathbb{1}\{a_i \in S_i\})$. A crucial observation is that since action availabilities are independent by Assumption 3.3, player i's payoff can be expressed in terms of $\mathbb{P}[a_j; \rho_j, \pi_j]$ for all players j, rather than the much larger π_j's (derivation in Appendix B.1):

$$U_i(\pi) = \sum_{a \in A} u_i(a) \prod_{j \in [n]} \mathbb{P}[a_j; \rho_j, \pi_j]. \tag{4}$$

Definition 4.1. *Let $\mu_i \in \Delta(A_i)$ be a probability distribution over player i's possible actions. We say that μ_i is an* implementable marginal *if there exists a strategy $\pi_i : \mathcal{S}_i \to \Delta(A_i)$ such that $\forall a_i \in A_i$, $\mu_i(a_i) = \mathbb{P}[a_i; \rho_i, \pi_i]$. In this case, we also say that π_i* implements *μ_i, and that $\pi = (\pi_1, \ldots, \pi_n)$ implements $\mu = (\mu_1, \ldots, \mu_n)$ if π_i implements μ_i for all i. The set of implementable marginals for player i is denoted $M_i \subseteq \Delta(A_i)$.*

Given an implementable marginal μ, we abuse notation to define the expected payoff to player i following μ as $U_i(\mu) = \mathbb{E}_{S \sim \rho}[\mathbb{E}_{a \sim \mu}[u_i(a)]]$. Note that by definition of μ, $\mathbb{E}_{S \sim \rho}\left[\mathbb{E}_{a \sim \pi(S)}[u_i(a)]\right] = \mathbb{E}_{S \sim \rho}[\mathbb{E}_{a \sim \mu}[u_i(a)]]$. Similarly, we let $U_i(a_i; \mu_{-i}) = \mathbb{E}_{S \sim \rho}\left[\mathbb{E}_{a_{-i} \sim \mu_{-i}}[u_i(a_i, a_{-i})]\right]$. The original strategy space is the finite product $\prod_{S_i \in \mathcal{S}_i} \Delta(S_i)$, which is compact and convex. Since the map $\pi_i \mapsto \mathbb{E}_{S_i \sim \rho_i}[\pi_i(\cdot \mid S_i)]$ is linear and continuous, its image M_i is also compact and convex. Consequently, we can take each player's "effective" strategy space to be M_i, implying an equivalence between Nash equilibria in the sense of Definition 3.4 and implementable marginal distributions that disincentivize unilateral deviations, as stated below.

Proposition 4.2. *Consider a GSAS where $\pi = (\pi_1, \ldots, \pi_n)$ is a strategy profile that implements $\mu = (\mu_1, \ldots, \mu_n)$. Then π is an ϵ-Nash equilibrium if and only if for all $i \in [n]$, $U_i(\mu) \geq \max_{\mu_i' \in M_i} U_i(\mu_i', \mu_{-i}) - \epsilon$.*

Notice that multiple π may implement the same μ, but not vice versa. Proposition 4.2 implies that if π and π' both implement μ *and* π is a NE, then so is π'. In 2p0s-GSAS, a NE corresponds to a bilinear saddle point problem over M_1 and M_2, and a minimax theorem over implementable strategies holds.

Proposition 4.3 (Minimax theorem for 2p0s-GSAS)**.** *For a 2p0s-GSAS, a strategy profile (π_1^*, π_2^*) is a NE if and only if the associated (μ_1^*, μ_2^*) is a saddle point of the function $U = U_1 = -U_2$, i.e., $\mu_1^* \in \operatorname{argmax}_{\mu_1 \in M_1} \min_{\mu_2 \in M_2} U(\mu_1, \mu_2)$ and $\mu_2^* \in \operatorname{argmin}_{\mu_2 \in M_2} \max_{\mu_1 \in M_1} U(\mu_1, \mu_2)$, where $\max_{\mu_1 \in M_1} \min_{\mu_2 \in M_2} U(\mu_1, \mu_2) = \min_{\mu_2 \in M_2} \max_{\mu_1 \in M_1} U(\mu_1, \mu_2)$.*

The following shows that NE in $\mathcal{G}_{\text{orig}}$ correspond to NE in $\mathcal{G}$, if the original NE are implementable.

Proposition 4.4. *Consider a GSAS $\mathcal{G} = (\mathcal{G}_{\text{orig}}, \mathcal{S}, \rho)$. Let $x^* = (x_1^*, \ldots, x_n^*)$ be an ϵ-NE of $\mathcal{G}_{\text{orig}}$, $x_i^* \in \Delta(A_i)$. If $\mu^* = x^*$ is implementable in $\mathcal{G}$ by $\pi^* = (\pi_1^*, \ldots, \pi_n^*)$, then π^* is an ϵ-NE in $\mathcal{G}$.*

Note that even in the two-player zero-sum case, Proposition 4.4 requires the NE in $\mathcal{G}_{\text{orig}}$, x^*, to be implementable for both players, i.e., x_i^* alone being implementable does not imply a solution to the maximin problem or optimal strategy for player i (see Example 3.5 (ii)). The above discussion indicates that for the purposes of equilibrium representation, we can work with M_i rather than the larger set of possible π_i. However, it is still unclear how to recover π_i from μ_i efficiently. We next show that every $\mu_i \in M_i$ can be *approximately* implemented by a compact, polynomially sized weight vector w_i.

Theorem 4.5. *Let π_i implement $\mu_i \in M_i$. Then, $\forall \delta \in (0, 1)$, there exists $w_i \in \Delta(A_i)$ prescribing a strategy π_i' where $\pi_i'(a_i \mid S_i) := \frac{w_i(a_i)\mathbb{1}\{a_i \in S_i\}}{\sum_{a_i' \in A_i} w_i(a_i')\mathbb{1}\{a_i' \in S_i\}}$. Furthermore, π_i' induces a marginal μ_i' with total variation distance $d_{\text{TV}}(\mu_i, \mu_i') < \delta$.*

Intuitively, any implementable μ_i can be δ-approximated in terms of total variation distance using a compact representation by an $|A_i|$-dimensional vector w_i. Given w_i, the corresponding $\pi_i'(a_i \mid S_i)$ for a fixed S_i can be computed in time linear in $|A_i|$, playing proportionally to w_i but restricted to the available actions S_i. We show an example of this representation in Example 4.6.

Example 4.6. *Consider a 2p0s-GSAS $\mathcal{G}$ where $\mathcal{G}_{\text{orig}}$ is a variant of the standard rock-paper-scissors (RPS) game, where the payoff is halved whenever player 1 plays paper, i.e., the payoff is 0.5 (resp. -0.5) instead of 1 (resp. -1) for winning (resp. losing). The full payoff matrix of $\mathcal{G}_{\text{orig}}$ is given in Definition C.7. For simplicity, denote the actions **R**ock, **P**aper, **S**cissors as **R**, **P**, **S**, respectively.*

Let $\mathcal{S}_1 = \{S_{11} = \{\boldsymbol{R}, \boldsymbol{P}, \boldsymbol{S}\}, S_{12} = \{\boldsymbol{P}, \boldsymbol{S}\}\}$ and $\mathcal{S}_2 = \{S_{21} = \{\boldsymbol{R}, \boldsymbol{S}\}, S_{22} = \{\boldsymbol{P}, \boldsymbol{S}\}\}$ and $\rho_1(S_{11}) = \rho_1(S_{12}) = \rho_2(S_{21}) = \rho_2(S_{22}) = 0.5$. It is easy to verify that a possible NE π^* is $\pi_1^*(S_{11}) = \left(\boldsymbol{R} : \frac{1}{2}, \boldsymbol{P} : 0, \boldsymbol{S} : \frac{1}{2}\right)$, $\pi_1^*(S_{12}) = (\boldsymbol{R} : 0, \boldsymbol{P} : 1, \boldsymbol{S} : 0)$, $\pi_2^*(S_{21}) = \left(\boldsymbol{R} : \frac{2}{3}, \boldsymbol{P} : 0, \boldsymbol{S} : \frac{1}{3}\right)$, and $\pi_2^*(S_{22}) = \left(\boldsymbol{R} : 0, \boldsymbol{P} : \frac{2}{3}, \boldsymbol{S} : \frac{1}{3}\right)$. Both players obtain an expected payoff of 0. The corresponding marginal distributions of play are $\mu_1^* = \left(\boldsymbol{R} : \frac{1}{4}, \boldsymbol{P} : \frac{1}{2}, \boldsymbol{S} : \frac{1}{4}\right)$ and $\mu_2^* = \left(\boldsymbol{R} : \frac{1}{3}, \boldsymbol{P} : \frac{1}{3}, \boldsymbol{S} : \frac{1}{3}\right)$. It can be shown that compact vectors $w_1^* = \left(\boldsymbol{R} : \frac{1}{2}, \boldsymbol{P} : \frac{1}{3}, \boldsymbol{S} : \frac{1}{6}\right)$ and $w_2^* = \left(\boldsymbol{R} : \frac{2}{5}, \boldsymbol{P} : \frac{2}{5}, \boldsymbol{S} : \frac{1}{5}\right)$ are consistent with the marginal distribution and yield a NE of $\mathcal{G}$ (see Appendix B.6 for a derivation).

Utilizing a compact vector to recover appropriate choices from a stochastic set of actions has been studied in the literature of the Luce model of stochastic choice (see e.g., Azrieli & Rehbeck (2026)). Nevertheless, our techniques further allow us to establish novel approximate bit-complexity results for the GSAS setting (cf. Remark 4.7). Moreover, the compactness result can also be viewed as an analogue of Kuhn's theorem (Kuhn, 1953) for extensive-form games, which shows that behavioral strategies are *outcome-equivalent* to the larger set of normal-form strategies under the assumption of perfect recall. Similarly, Theorem 4.5 ensures that compact vectors suffice to *approximately* capture implementable NE under Assumption 3.3.

Remark 4.7. *Thus far, we have used the term "compact representation" rather loosely to refer to a simplex vector of size $|A_i|$. The astute reader might be concerned about their precise* bit complexity*, i.e., whether the entries of w_i^* can be written using a polynomial number of bits with respect to $\mathcal{G}_{\text{orig}}$. In Appendix D, we derive several results in this direction. First, we formalize the notion of a "bit-compact" representation in the context of GSAS strategies. We show in Theorem D.2 and Corollary D.4 that under Assumption 3.3, 2p0s-GSAS admit* approximate *bit-compact representations. Conversely, we show in Theorem D.5 that without Assumption 3.3, there exist families of 2p0s-GSAS that do* not *admit bit-compact representations. This further justifies the importance of Assumption 3.3 in our analysis, and motivates future work on finding subclasses of GSAS that always admit bit-compact representations, even after relaxing the assumption.*

5 COMPUTING EQUILIBRIA IN GSAS

In this section, we focus on the problem of NE computation in 2p0s-GSAS. There are three main regimes of interest. (i) The **small-support** regime, where $\mathcal{S}$ is small and ρ is known exactly (cf. Example 4.6). (ii) The **oracle-access** regime, where $\mathcal{S}$ is exponentially large in $|A_i|$ so that ρ cannot be explicitly enumerated, but can be queried in constant time for every $S \in \mathcal{S}$. (iii) The **sample-access** regime, where ρ is unknown and we only have access to a simulator that samples $S \sim \rho$. In regime (i), one could represent the game as a Bayesian game and apply off-the-shelf Bayesian game solvers. However, this does not take advantage of the additional structure afforded by GSAS, and incurs runtime costs linear in the number of types, which can be exponential in $|A_i|$ (cf. Appendix A.1). Our goal is to design a broad approach that applies to any GSAS, even in regime (iii). All proofs of the theoretical results in this section are deferred to Appendix B.

5.1 RANKING REGRET MINIMIZATION IN GSAS

Going forward, we utilize the *online learning* or *repeated* game paradigm. In this setting, for each player $i \in [n]$, the sequence $\{S_i^t\}_{t=1,\dots,T}$ is sampled i.i.d. from ρ_i. In each iteration t, player i selects a strategy π_i^t, observes S_i^t, samples $a_i^t \sim \pi_i^t(\cdot \mid S_i^t)$, and then observes $u_i(a_i, a_{-i}^t)$ for every $a_i \in S_i^t$. Notice that this setting applies to all regimes (i)-(iii) as described above. A standard performance metric for learning in games is (cumulative) *external regret*, formally defined as $\max_{a_i' \in A_i} \sum_{t=1}^T \left[u_i(a_i', a_{-i}^t) - u_i(a_i^t, a_{-i}^t)\right]$, with the folk result that no-external-regret algorithms converge in time-average to the set of Nash equilibria in 2p0s games. In GSAS, standard notions of regret are unsuitable since the competing action may be unavailable in certain rounds. This motivates the adoption of a regret variant called *ranking regret* that was first studied in the sleeping bandit literature (Kleinberg et al., 2010; Kanade et al., 2009).

Some additional notation is in order: a ranking τ_i is a total order on A_i. The corresponding *ranking strategy* $\pi_i^{\tau_i}$ selects $\mathrm{top}_{\tau_i}(S_i)$, the highest-ranked action in the realized availability set S_i. Let $\mathcal{T}_i$ denote the set of rankings of A_i, and let $\mu_i^{\tau_i} \in M_i$ be the marginal induced by $\pi_i^{\tau_i}$. In the GSAS setting, the regret definition is given by comparing the players' *expected* payoffs U_i for playing strategy π_i^t with the expected payoff for playing according to the ranking strategy $\pi_i^{\tau_i}$, which is selected in hindsight.

Definition 5.1 (Ranking Regret)**.** *Consider a sequence of strategy profiles $\pi^t = (\pi_1^t, \pi_2^t, \dots, \pi_n^t)$, $t = 1, \dots, T$. Player i's* ranking regret *is $R_{T,i}^{\mathrm{rank}} := \max_{\tau_i \in \mathcal{T}_i} \sum_{t=1}^T \left[U_i(\pi_i^{\tau_i}, \pi_{-i}^t) - U_i(\pi^t)\right]$.*

If a player's ranking regret grows sublinearly, i.e., $R_{T,i}^{\mathrm{rank}} = o(T)$, they are said to have *no-ranking-regret*. Intuitively, the player does not regret playing their strategy π_i^t against the best ranking over actions in expectation. Next, while ranking regret is defined for arbitrary GSAS, in Proposition 5.3 we establish a connection between algorithms that achieve no-ranking-regret and Nash equilibria in the special case of 2p0s-GSAS. Let μ_i^t be the marginal distribution induced by π_i^t, and define the time-averaged induced marginal distribution $\bar{\mu}_i^T := \frac{1}{T}\sum_{t=1}^T \mu_i^t$. This marginal is implementable by definition, since the pointwise average strategy $\bar{\pi}_i^T(\cdot \mid S_i) := \frac{1}{T}\sum_{t=1}^T \pi_i^t(\cdot \mid S_i)$ induces $\bar{\mu}_i^T$. In 2p0s-GSAS, let $U := U_1 = -U_2$. We consider the standard NE convergence metric of *saddle point residual* (SPR). By Proposition 4.2, it suffices to consider the implementable strategies $\mu_i \in M_i$ for each player i.

Definition 5.2 (SPR in implementable strategies)**.** *The saddle point residual (*SPR*) in implementable strategies is defined as*

$$\mathrm{SPR}(\mu_1, \mu_2) := \max_{\mu_1' \in M_1} U(\mu_1', \mu_2) - \min_{\mu_2' \in M_2} U(\mu_1, \mu_2'). \tag{5}$$

Proposition 5.3. *Consider a 2p0s-GSAS $\mathcal{G}$ satisfying Assumption 3.3. For every sequence of strategies, $\mathrm{SPR}(\bar{\mu}_1^T, \bar{\mu}_2^T) = \frac{R_{T,1}^{\mathrm{rank}} + R_{T,2}^{\mathrm{rank}}}{T} =: \epsilon_T$. Consequently, every strategy profile implementing $(\bar{\mu}_1^T, \bar{\mu}_2^T)$ is an ϵ_T-Nash equilibrium. In particular, if $R_{T,i}^{\mathrm{rank}} = o(T)$ for both players, then the time-averaged marginals converge to the set of NE of $\mathcal{G}$ in terms of SPR.*

In practice, we use the empirical marginals $\tilde{\mu}_i^T := \frac{1}{T}\sum_{t=1}^T \pi_i^t(\cdot \mid S_i^t)$ in place of $\bar{\mu}_i^T$; a standard concentration argument shows that the same SPR guarantee, up to an additional error term, holds with high probability.

Remark 5.4 (Necessity of Ranking Regret)**.** *While sublinear external regret suffices to show time-averaged convergence to the set of NE in standard 2p0s games, this implication does not hold in GSAS. In the sleeping bandits literature, other natural analogues of regret compatible with action restrictions are sleeping external and internal regret. However, we show in Appendix B.8 that neither notion is sufficient to guarantee convergence to Nash equilibrium in GSAS. Due to this, many prior algorithms which were designed for the sleeping bandit setting do not apply to equilibrium computation in our setting.*

Proposition 5.3 reduces equilibrium computation to the problem of achieving sublinear *expected pathwise* ranking regret. Here "pathwise" means that the best fixed ranking is chosen after the randomly sampled trajectory is realized. We address this problem in the *sample-access* setting (regime (iii)), where we assume that player i can draw independent samples from ρ_i in addition to observing the realized set S_i^t.

Our proposed algorithm, CAT-FTPL, uses some ideas from prior work. Follow-the-Perturbed-Leader (FTPL) (Hannan, 1957) selects an available action by perturbing cumulative estimated losses; SLEEPINGCAT (Neu & Valko, 2014) uses Counting Asleep Times (CAT) to estimate the inverse of action-availability probabilities; and Geometric Resampling (GR) (Neu & Bartók, 2016) provides a way to estimate these probabilities by repeatedly sampling from ρ (which is possible since we are in regime (iii)). CAT-FTPL combines FTPL with a GR implementation of a CAT-style estimator, together with a small implicit-exploration offset (Kocák et al., 2014) which yields the exponential-moment control needed to obtain sublinear regret. Full details of CAT-FTPL are given in Appendix B.9 as Algorithm 1.

Theorem 5.5. *A player using CAT-FTPL has expected pathwise ranking regret bounded as*

$$\mathbb{E}\left[R_{T,i}^{\mathrm{rank}}\right] \le 2\left(\log|A_i|\left(\frac{1}{\eta_i}+\frac{1}{\gamma_i}\right)+|A_i|T(\eta_i+\gamma_i)\right). \tag{6}$$

In particular, with parameters $\eta_i=\gamma_i=\sqrt{\frac{\log|A_i|}{|A_i|T}}$*, we obtain* $\mathbb{E}\left[R_{T,i}^{\mathrm{rank}}\right] \le 8\sqrt{|A_i|T\log|A_i|}$.

Going forward, unless otherwise stated, we utilize CAT-FTPL with $\eta_i=\gamma_i=\sqrt{\log|A_i|/|A_i|T}$. We additionally derive a high-probability bound beyond expected ranking regret, establishing that every player using CAT-FTPL has sublinear ranking regret with high probability.

Proposition 5.6. *Suppose players* $i\in[n]$ *run CAT-FTPL in a GSAS. Then for any* $p_1,\dots,p_n\in(0,1)$ *with* $\sum_i p_i<1$, $\Pr\left(\bigcap_{i=1}^n\left\{R_{T,i}^{\mathrm{rank}}\le\frac{16}{p_i}\sqrt{|A_i|T\log|A_i|}\right\}\right)\ge 1-\sum_{i=1}^n p_i$.

Consequently, if the GSAS is 2p0s, Proposition 5.3 directly implies an approximate-NE guarantee for the time-averaged induced marginals produced by CAT-FTPL.

5.2 COMPUTING COMPACT EQUILIBRIA IN GSAS

While the sublinear ranking regret guarantee of CAT-FTPL ensures that the time-averaged marginals converge to the set of NE in probability, these iterates alone do not constitute an explicit solution to the GSAS. In light of Theorem 4.5, we seek a method to compute approximate *compact* vectors associated with these averaged strategies. For the following asymptotic result, suppose that a sequence of strategies $\{\pi_i^t\}_{t=1,\dots,T}$ satisfies $\frac{1}{T}\sum_{t=1}^T\pi_i^t(S_i^t)\to\mu_i^*$ almost surely as $T\to\infty$, where μ_i^* is a marginal distribution induced by some (approximate) Nash equilibrium π^*. Specifically, for a vector $w_i\in\mathbb{R}_{>0}^{|A_i|}$, let $\hat{\pi}_i(a_i\mid S_i,w_i)$ be the probability of playing action a_i given availability set S_i. Then, by Theorem 4.5 there exists a δ-approximate w_i so that $\forall S_i\in\mathcal{S}_i,\ a_i\in A_i$, $\hat{\pi}_i(a_i\mid S_i,w_i)=\frac{w_i(a_i)\mathbb{1}\{a_i\in S_i\}}{\sum_{a_i'\in A_i}w_i(a_i')\mathbb{1}\{a_i'\in S_i\}}$. Let $\hat{\mu}_i(w_i)$ be the corresponding marginal distribution where $\hat{\mu}_i(a_i|w_i)=\mathbb{E}_{S_i\sim\rho_i}[\hat{\pi}_i(a_i\mid S_i,w_i)]$ for all $a_i\in A_i$. Our objective is to find a w_i that approximately solves $\hat{\mu}_i(w_i)=\mu_i^*$. This can be done through a stochastic approximation (SA) procedure as outlined in Appendix B.12 as Algorithm 2.

Proposition 5.7. *Let* w_i^T *be the weight vector produced by Algorithm 2. If* $\frac{1}{T}\sum_{t=1}^T\pi_i^t(S_i^t)\to\mu_i^*$ *almost surely as* $T\to\infty$*, and that* $\sum_{t=1}^\infty\eta_t=\infty$ *and* $\sum_{t=1}^\infty\eta_t^2<\infty$*, then the marginal induced by* w_i^T *converges to* μ_i^* *almost surely. In particular, almost surely, for every* $\delta>0$*,* w_i^T *is a* δ*-approximate compact representation of* μ_i^* *for all sufficiently large* T.

Indeed, in our setting, Algorithm 2 uses the time-averaged, empirical marginal distributions $\tilde{\mu}_i^t := \frac{1}{t}\sum_{\tau=1}^{t} \pi_i^\tau(S_i^\tau)$, where π_i^τ are the strategies generated by CAT-FTPL. In addition to the result above, we also utilize a modified version of the *robust stochastic approximation* (RSA) approach introduced by Nemirovski & Yudin (1978) and further developed by Nemirovski et al. (2009) to obtain finite-time convergence bounds by running Algorithm 3 and taking robust time-averages over the iterates (see Appendix B.13 for details). These results imply a simple procedure for computing *compact* approximate equilibria in GSAS: (i) run CAT-FTPL for rounds $t = 1, \ldots, T$, (ii) compute the empirical marginal distributions $\tilde{\mu}_i^t$ for each t, and (iii) run Algorithm 3 to obtain $\tilde{w}_i$. In Appendix C.3, we show an example of this procedure applied to Example 4.6, approximately recovering w_1 and w_2 that implement an ϵ-NE. Taken together, our analysis leads to a high-probability *finite-time* bound on the SPR of the computed approximate compact vectors.

Theorem 5.8. *Suppose CAT-FTPL is run for T rounds in a 2p0s-GSAS with payoffs $u_i : A \to [-1, 1]$, and the empirical marginals $(\tilde{\mu}_1, \tilde{\mu}_2)$ are used in Algorithm 3 to obtain compact vectors $\tilde{w}_i$. Then, for any $p \in (0, 1)$, with probability at least $1 - p$, the SPR of the strategies induced by $\tilde{w}_i$ satisfies $\mathrm{SPR}^2 \leq O\left(\frac{1}{p\sqrt{T}} + \frac{1}{p^2 T}\right)$, where $O(\cdot)$ hides polynomial factors in $|A_i|$.*

CAT-FTPL contributes $O(1/(p^2 T))$ to the bound, which is dominated by the $O(1/(p\sqrt{T}))$ term from the RSA procedure whenever $\sqrt{T} \gg 1/p$. Note that the bound on SPR is relatively loose, since the RSA analysis bounds the expected squared ℓ_2-norm of the error induced by the computed compact vector, whereas SPR is controlled by the ℓ_1-norm of this error. Nevertheless, asymptotic convergence in terms of SPR still holds in probability, and we leave as an open problem the design of SA procedures that improve upon this bound. Moreover, we show in Section 6 that our procedure empirically computes $\tilde{w}$ with low SPR in large 2p0s-GSAS.

6 Experimental Results

Experiment 1: Comparison with LP solver. In light of our compact representation results, we investigate the *scalability* of CAT-FTPL compared to existing solvers for GSAS with exponentially sized support $\mathcal{S}$. Since there are no existing baselines for GSAS solvers, we compare our method against reductions that allow for the application of linear programs (LPs). Note, however, that the LP method requires full access to ρ, while CAT-FTPL only requires sample access to ρ. Due to the inherent differences between LP solvers and our online learning approach, the computed solutions are not directly comparable in terms of output quality, but only in *runtime*. In Figure 1, we compare the wallclock runtime of CAT-FTPL and Gurobi's linear program solver on randomly generated GSAS (full definition in Definition C.1). Fixing a time budget of 360 seconds, we found that Gurobi's LP solver, when applied to the sequence-form representation (von Stengel, 1996) of the GSAS that encodes all action subsets, could only solve games of up to size 11×11. Comparatively, CAT-FTPL was able to obtain low marginal SPR, i.e., $\mathrm{SPR}(\tilde{\mu}_1, \tilde{\mu}_2) \leq 0.05$, for much larger games (350×350) within the same time budget, demonstrating the scalability of our approach. Additional experimental details are given in Appendix C.2.1. Finally, beyond comparisons with LP solvers, we also compare our method with a reduction that allows for the application of Bayesian game solvers (details in Appendix C.4).

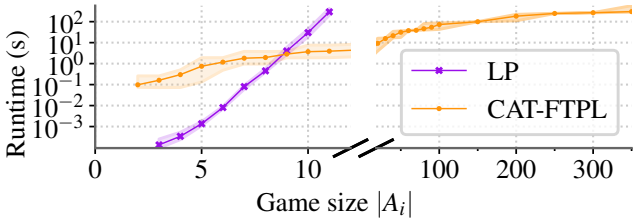
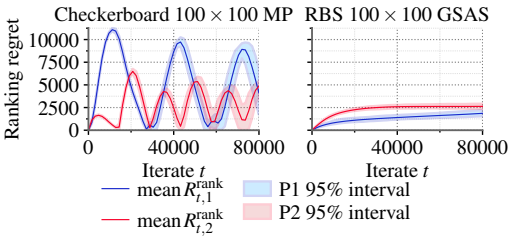
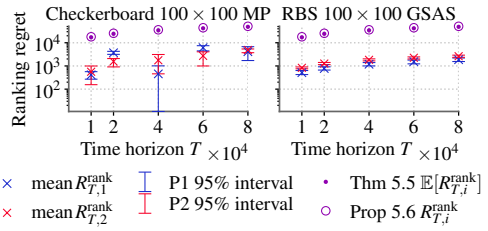


Figure 1: Wallclock time to solve randomly generated GSAS by CAT-FTPL and Gurobi LP solver. Plot shows the average over 20 runs with shaded region showing the range (max and min of wallclock time) of values across runs.

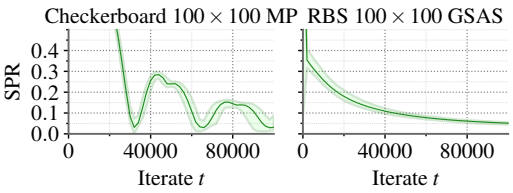


(a) Ranking regret $R^{\text{rank}}_{t,i}$ for intermediate iterates with time horizon $T = 80000$. Mean and 95% central interval from 100 runs shown.

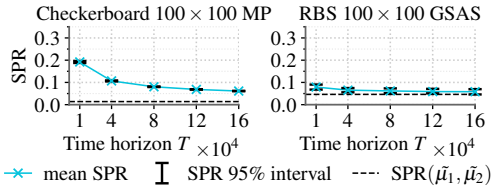


(b) Ranking regret $R^{\text{rank}}_{T,i}$ of final iterate of CAT-FTPL for various time horizons T. Results from 100 runs for each T and theoretical bounds shown.

Figure 2: Ranking regret of CAT-FTPL for two 2p0s-GSAS. Mean, central 95% interval and theoretical expected/high-probability bounds on maximum ranking regret from Theorem 5.5 and Proposition 5.6 are shown. We observed $R^{\text{rank}}_{T,i}$ to be much lower in practice than the bound in Theorem 5.5.



(a) SPR of the time-averaged empirical marginals $\tilde{\mu}^t_i$ played by CAT-FTPL at intermediate iterates for time horizon $T = 100000$.



(b) SPR of strategies induced by the w^T_i vectors computed by Algorithm 3 for various time horizons T.

Figure 3: SPR for two 2p0s-GSAS. For each game, we repeated each experiment for 100 runs and plotted the mean and 95% central interval over the runs.

Experiment 2: Convergence in large GSAS. We seek to corroborate our convergence results in Section 5 in terms of ranking regret and SPR. We focus on two large GSAS: (i) 100×100 Random Biased Support GSAS ("RBS") (cf. Definition C.2) and (ii) Checkerboard 100×100 Matching Pennies (cf. Definition C.3), which are specifically constructed to be challenging due to the payoff structure and stochastic action availabilities. Additional experimental setup details, including estimating $R^{\text{rank}}_{T,i}$ and SPR, as well as further experiments on other GSAS, are given in Appendix C.2. In Figure 2 we focus on the ranking regret, with Figure 2a showing $R^{\text{rank}}_{t,i}$ at all intermediate iterates $t \in \{1, ..., T\}$ for a single time horizon T. Figure 2b shows the final ranking regret at time horizon T for various values of T, and we find the observed $R^{\text{rank}}_{T,i}$ to be bounded by Theorem 5.5 and Proposition 5.6, consistent with sublinear growth with high probability. In Figure 3 we focus on the SPR, with Figure 3a showing the time-average marginal $\tilde{\mu}^t_i$ at the intermediate iterates $t \in \{1, ..., T\}$ played by CAT-FTPL, which we observe converges to low SPR marginals even in challenging games by time horizon T. In Figure 3b, we show the SPR of the final compact strategy w^T_i computed by Algorithm 3 for various time horizons T, using as input marginals $(\tilde{\mu}_1, \tilde{\mu}_2)$ produced by CAT-FTPL. We empirically observe that the SPR corroborates Theorem 5.8, indicating that our proposed method is effective in computing compact strategies representing ϵ-NE in 2p0s-GSAS.

7 Discussion and Future Work

In this paper we have taken the first step toward characterizing and computing *compact* Nash equilibria in games with stochastic action sets. Our analysis leaves open several promising future research directions. These include (i) characterizing and studying convergence to appropriate notions of *correlated* equilibria in general-sum GSAS, (ii) modeling and solving extensive-form and Markovian variations of GSAS, (iii) exploring efficient strategy representations even after relaxing the independence assumption, and (iv) designing more computationally efficient, parameter-free ranking-regret algorithms, particularly methods that avoid repeated geometric resampling.

AI USE STATEMENT

In this work, we used generative AI tools to assist in cleaning up and refining author-written drafts of proofs, as well as to find counterexamples (in particular, the counterexample in Appendix D.2, in order to obtain irrational solutions—the proof idea itself came from the authors). We have not used generative AI tools for developing theoretical models or conceptual frameworks, proposing/refining hypotheses, designing or providing feedback on research methodology or experiments, and interpreting results. Generating synthetic datasets, translation, cleaning/reformatting datasets, and supporting qualitative/thematic data analysis are not applicable to this work. Additionally, we used generative AI tools for writing code to implement our proposed algorithms and to sparingly identify relevant literature. We have reviewed all AI-assisted work: all suggested edits to proofs were verified by hand, a manual literature survey was done to ensure that the identified research was valid, and LLM-generated code was verified and tested for correctness. We take full responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

While our results are primarily theoretical, we acknowledge that there could be potential societal and ethical consequences of our work, none of which we feel must be specifically highlighted.

REPRODUCIBILITY STATEMENT

Our theoretical results are each accompanied with a proof in the appendix. The code used to generate the experimental figures throughout the paper, alongside the implementation of our algorithms, is provided as supplementary material.

REFERENCES

Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2):235–256, 2002.

Yaron Azrieli and John Rehbeck. Marginal stochastic choice. *Theoretical Economics*, 21(2):373–401, 2026.

Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal, Second Series*, 19(3):357–367, 1967.

Michel Benaïm and Olivier Raimond. A class of self-interacting processes with applications to games and reinforced random walks. *SIAM Journal on Control and Optimization*, 48(7):4707–4730, 2010.

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(47):1307–1324, 2007.

Craig Boutilier, Alon Cohen, Avinatan Hassidim, Yishay Mansour, Ofer Meshi, Martin Mladenov, and Dale Schuurmans. Planning and learning with stochastic action sets. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, pp. 4674–4682. AAAI Press, 2018. ISBN 9780999241127.

Mario Bravo and Mathieu Faure. Reinforcement learning with restrictions on the action set. *SIAM Journal on Control and Optimization*, 53(1):287–312, 2015.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. From external to swap regret 2.0: An efficient reduction for large action spaces. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pp. 1216–1222, 2024.

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pp. 235–254. SIAM, 2011.

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems*, 34:27604–27616, 2021.

Jing Dong, Jingyu Wu, Siwei Wang, Baoxiang Wang, and Wei Chen. Taming the exponential action set: Sublinear regret and fast convergence to Nash equilibrium in online congestion games. *arXiv preprint arXiv:2306.13673*, 2023.

Simone Drago, Marco Mussi, and Alberto Maria Metelli. Sleeping reinforcement learning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 14439–14498, 2025.

Erez Druk and Yuval Ishai. Linear-time encodable codes meeting the Gilbert-Varshamov bound and their cryptographic applications. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, ITCS '14, pp. 169–182, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450326988. doi: 10.1145/2554797.2554815. URL https://doi.org/10.1145/2554797.2554815.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Stochastic regret minimization in extensive-form games. In *International Conference on Machine Learning*, volume 119, pp. 3018–3028. PMLR, 2020.

Gabriele Farina, Chung-Wei Lee, Haipeng Luo, and Christian Kroer. Kernelized multiplicative weights for 0/1-polyhedral games: Bridging the gap between learning in extensive-form and normal-form games. In *International Conference on Machine Learning*, pp. 6337–6357. PMLR, 2022.

Yoav Freund and Robert E Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103, 1999.

Kaito Fujii. Bayes correlated equilibria, no-regret dynamics in Bayesian games, and the price of anarchy. In *The Thirty Eighth Annual Conference on Learning Theory*, pp. 2190–2191. PMLR, 2025.

Pierre Gaillard, Aadirupa Saha, and Soham Dan. One arrow, two kills: A unified framework for achieving optimal regret guarantees in sleeping bandits. In *International Conference on Artificial Intelligence and Statistics*, volume 206, pp. 7755–7773. PMLR, 2023.

E. N. Gilbert. A comparison of signalling alphabets. *The Bell System Technical Journal*, 31(3): 504–522, 1952. doi: 10.1002/j.1538-7305.1952.tb01393.x.

Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential Coding Theory*. University at Buffalo and Carnegie Mellon University, 2025. URL http://www.cse.buffalo.edu/faculty/atri/courses/coding-theory/book/.

James Hannan. Approximation to Bayes risk in repeated play. In *Contributions to the Theory of Games*, volume 3 of *Annals of Mathematics Studies*, pp. 97–139. Princeton University Press, 1957.

John C Harsanyi. Games with incomplete information played by "Bayesian" players, Part II. Bayesian equilibrium points. *Management Science*, 14(5):320–334, 1968.

Jason Hartline, Vasilis Syrgkanis, and Eva Tardos. No-regret learning in Bayesian games. *Advances in Neural Information Processing Systems*, 28, 2015.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Satyen Kale, Chansoo Lee, and Dávid Pál. Hardness of online sleeping combinatorial optimization problems. *Advances in Neural Information Processing Systems*, 29, 2016.

Varun Kanade and Thomas Steinke. Learning hurdles for sleeping experts. *ACM Transactions on Computation Theory (TOCT)*, 6(3), 2014.

Varun Kanade, H Brendan McMahan, and Brent Bryan. Sleeping experts and bandits with stochastic action availability and adversarial rewards. In *Artificial Intelligence and Statistics*, volume 5, pp. 272–279. PMLR, 2009.

Robert Kleinberg, Alexandru Niculescu-Mizil, and Yogeshwer Sharma. Regret bounds for sleeping experts and bandits. *Machine Learning*, 80(2–3):245–272, 2010.

Tomáš Kocák, Gergely Neu, Michal Valko, and Rémi Munos. Efficient learning by implicit exploration in bandit problems with side observations. *Advances in Neural Information Processing Systems*, 27, 2014.

Andreas Kontogiannis, Vasilis Pollatos, Gabriele Farina, Panayotis Mertikopoulos, and Ioannis Panageas. Efficient kernelized learning in polyhedral games beyond full-information: From Colonel Blotto to congestion games. *Advances in Neural Information Processing Systems*, 38, 2025.

Harold W. Kuhn. Extensive games and the problem of information. In *Contributions to the Theory of Games*, Annals of Mathematics Studies, pp. 193–216. Princeton University Press, 1953.

John Lazarsfeld, Georgios Piliouras, Ryann Sim, and Andre Wibisono. Fast and furious symmetric learning in zero-sum games: Gradient descent as fictitious play. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291, pp. 3527–3577. PMLR, 2025.

MV Menon and Hans Schneider. The spectrum of a nonlinear operator associated with a matrix. *Linear Algebra and its Applications*, 2(3):321–334, 1969.

A. S. Nemirovski and D. B. Yudin. Cesari convergence of the gradient method of approximating saddle points of convex-concave functions. *Soviet Mathematics Doklady*, 19(2):258–269, 1978.

Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

Gergely Neu and Gábor Bartók. Importance weighting without importance weights: An efficient algorithm for combinatorial semi-bandits. *Journal of Machine Learning Research*, 17(154):1–21, 2016.

Gergely Neu and Michal Valko. Online combinatorial optimization with stochastic decision sets and adversarial losses. *Advances in Neural Information Processing Systems*, 27, 2014.

Quan M Nguyen and Nishant Mehta. Near-optimal per-action regret bounds for sleeping bandits. In *International Conference on Artificial Intelligence and Statistics*, volume 238, pp. 2827–2835. PMLR, 2024.

Ioannis Panageas, Stratis Skoulakis, Luca Viano, Xiao Wang, and Volkan Cevher. Semi bandit dynamics in congestion games: Convergence to Nash equilibrium and no-regret guarantees. In *International Conference on Machine Learning*, pp. 26904–26930. PMLR, 2023.

Binghui Peng and Aviad Rubinstein. The complexity of approximate (coarse) correlated equilibrium for incomplete information games. In *The Thirty Seventh Annual Conference on Learning Theory*, pp. 4158–4184. PMLR, 2024.

Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26, 2013.

Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. In *Proceedings of the 29th International Conference on Machine Learning*, ICML'12, pp. 1571–1578, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851.

Tim Roughgarden. The price of anarchy in games of incomplete information. *ACM Transactions on Economics and Computation (TEAC)*, 3(1):1–20, 2015.

Aadirupa Saha, Pierre Gaillard, and Michal Valko. Improved sleeping bandits with stochastic action sets and adversarial rewards. In *International Conference on Machine Learning*, volume 119, pp. 8357–8366. PMLR, 2020.

Jon Schneider and Julian Zimmert. Optimal cross-learning for contextual bandits with unknown context distributions. *Advances in Neural Information Processing Systems*, 36:51862–51880, 2023.

Lloyd S Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10): 1095–1100, 1953.

Vasilis Syrgkanis. Bayesian games and the smoothness framework. *arXiv preprint arXiv:1203.5155*, 2012.

Vasilis Syrgkanis and Eva Tardos. Composable and efficient mechanisms. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pp. 211–220, 2013.

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E Schapire. Fast convergence of regularized learning in games. *Advances in Neural Information Processing Systems*, 28, 2015.

Rom Rubenovich Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akad. Nauk, SSSR*, 117:739–741, 1957.

J von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928.

Bernhard von Stengel. Efficient computation of behavior strategies. *Games and Economic Behavior*, 14(2):220–246, 1996.

APPENDIX

This supplementary material contains an overview of additional related work in Appendix A, proofs omitted from the main paper for space considerations in Appendix B, further experimental results and details in Appendix C, and formal bit-complexity results in Appendix D.

A ADDITIONAL RELATED WORK

Games with Action Set Restrictions. Benaïm & Raimond (2010) and Bravo & Faure (2015) studied an algorithm called Markovian Fictitious Play (MFP) in repeated two-player normal-form games where the action sets are restricted at each round. Unlike our setting, the action restrictions depend on the players' previous actions and are encoded via an *exploration matrix*. MFP requires players to compute a best response at each round and is shown to converge almost surely to NE in two-player zero-sum and potential games.

Learning in Games with Exponential Action Sets. Panageas et al. (2023) and Dong et al. (2023) studied semi-bandit learning in congestion games which admit exponentially large action sets, leading to slow convergence of standard methods. While their approach also utilizes learning over a compact set of 'facilities', their setting focuses on computing NE of the original congestion game (i.e., NE which are implementable in a sense that is introduced in Definition 4.1). Comparatively, in GSAS, the set of NE can change drastically, and we seek to compute the NE of the game induced by the stochastic action sets, not $\mathcal{G}_{\text{orig}}$ itself. Moreover, our emphasis is also on computing compact representations of NE, which allow players to efficiently play the game optimally.

The study of regret minimization in games with structured strategy sets beyond congestion games has an extensive literature, primarily thanks to the canonical connection between no-regret learning and game-theoretic equilibria (Freund & Schapire, 1999; Cesa-Bianchi & Lugosi, 2006). Farina et al. (2022) and Kontogiannis et al. (2025) studied regret minimization in *polyhedral* games (i.e., games with combinatorial structure which might have exponential action sets), and Farina et al. (2020) studied regret minimization under stochasticity (i.e., Monte-Carlo sampling of regrets). These papers focus on deriving regret bounds, whereas we additionally study efficient methods for equilibrium representation.

A.1 ON CONNECTIONS TO BAYESIAN GAMES

Since a GSAS can be viewed as a variant of Bayesian games, it is natural to discuss existing methods for online learning in Bayesian games. There are two primary access models for ρ where algorithms for Bayesian games can be applied to GSAS.

First, we consider the case where players only have sample or oracle access to ρ. In this setting, Hartline et al. (2015) studied no-regret learning in finite Bayesian games, requiring only sample access to ρ. However, there are two key distinctions: (i) the equilibrium concept of concern in Hartline et al. (2015) is coarse correlated equilibria (CCE) in general-sum Bayesian games, and so they focus on no-*external*-regret algorithms, and (ii) their convergence result is an almost sure convergence statement and does not admit an explicit finite convergence rate. In comparison, we obtain finite-time convergence guarantees via a more fine-grained analysis.

Similarly, recent methods can approximately solve for correlated equilibria in general-sum Bayesian games (Dagan et al., 2024; Peng & Rubinstein, 2024) using only sample access to ρ, but requiring knowledge of K, the number of types. Their algorithm, called multi-scale MWU, can in principle be applied to the 2p0s-GSAS setting, but the runtime grows as $O(|A|K\log(|A|K))$ (note that the runtime per iteration depends linearly on K). Comparatively, we exploit the additional structure of GSAS to design a procedure that has an improved rate and runtime that does *not* depend on the number of types. In Figure 8, we compare the wallclock runtimes of multi-scale MWU and our method in randomly generated GSAS, showing that our method can solve significantly larger games than multi-scale MWU in a given time budget.

The second setting is in the case that ρ is explicitly known, which makes equilibrium computation easier, since GSAS can be explicitly written as a Bayesian game where action subsets are modeled as types. Then, one could expand the GSAS into an expanded normal-form game, similar to how

one might expand a Bayesian game into agent-form/induced normal-form, and solve it directly in this space. To improve scalability, one can also avoid the full-blown normal-form LP by converting the game into sequence-form, which models Nature as a chance node in a game tree that determines the players' types. This representation is well-known to be computationally more efficient than the full normal-form representation, and solving this can be done with an off-the-shelf game solver which obtains a Bayesian Nash equilibrium corresponding to a Nash equilibrium in the GSAS. However, these methods have a dependence on the size of the sequence-form representation, which again can be linear in the number of types. Conversely, our procedure is designed to exploit the additional structure afforded by GSAS to avoid this blow-up. This observation is corroborated in our experiments (cf. Figure 1), where our method is compared explicitly to an LP solver as applied to the sequence-form of the game.

B Omitted Proofs

B.1 Derivation of Equation (4)

$$\begin{aligned}
U_i(\pi) &= \mathbb{E}_{S\sim\rho}\mathbb{E}_{a\sim\pi(S)}[u_i(a)] \\
&= \sum_{S\in\mathcal{S}} \left(\prod_{j\in[n]} \rho_j(S_j)\right) \sum_{a\in S} u_i(a) \prod_{j\in[n]} \pi_j(a_j \mid S_j) \\
&= \sum_{a\in A} u_i(a) \sum_{\substack{S\in\mathcal{S}\\ a\in S}} \prod_{j\in[n]} \rho_j(S_j)\pi_j(a_j \mid S_j) \\
&= \sum_{a\in A} u_i(a) \sum_{S\in\mathcal{S}} \prod_{j\in[n]} \rho_j(S_j)\pi_j(a_j \mid S_j)\mathbf{1}\{a_j \in S_j\} \\
&= \sum_{a\in A} u_i(a) \prod_{j\in[n]} \sum_{S_j\in\mathcal{S}_j} \rho_j(S_j)\pi_j(a_j \mid S_j)\mathbf{1}\{a_j \in S_j\} \\
&= \sum_{a\in A} u_i(a) \prod_{j\in[n]} P[a_j;\rho_j,\pi_j].
\end{aligned}$$

The second equality holds by expanding the expectations, using Assumption 3.3 to factor ρ and the definition of the joint strategy to factor the conditional action distribution. The third exchanges two finite sums, the fourth replaces the constraint $a \in S$ with action-availability indicators, the fifth factors the sum over the Cartesian product, and the final equality follows from the definition.

B.2 Proof of Proposition 4.2

Proposition 4.2. *Consider a GSAS where $\pi = (\pi_1, \ldots, \pi_n)$ is a strategy profile that implements $\mu = (\mu_1, \ldots, \mu_n)$. Then π is an ϵ-Nash equilibrium if and only if for all $i \in [n]$, $U_i(\mu) \geq \max_{\mu_i'\in M_i} U_i(\mu_i', \mu_{-i}) - \epsilon$.*

Proof. ($\Longleftarrow$) Consider a strategy profile π which implements μ.

We have that $U_i(\mu) \geq \max_{\mu_i'\in M_i} U_i(\mu_i', \mu_{-i}) - \epsilon$. Expanding the expression for the expected payoff of π, we get

$$\begin{aligned}
U_i(\pi) &= \sum_{a\in A} u_i(a) \prod_{j\in[n]} \mathbb{P}[a_j;\rho_j,\pi_j] \\
&= U_i(\mu) \\
&\geq \max_{\mu_i'\in M_i} U_i(\mu_i', \mu_{-i}) - \epsilon \\
&= \max_{\pi_i'} \sum_{a_i\in A_i} u_i(a)\mathbb{P}[a_i;\rho_i,\pi_i'] \prod_{j\neq i} \mathbb{P}[a_j;\rho_j,\pi_j] - \epsilon \\
&= \max_{\pi_i'} U_i(\pi_i', \pi_{-i}) - \epsilon,
\end{aligned}$$

where we utilize the implementability of μ and the fact that $\mu_i(a_i) = \mathbb{P}[a_i; \rho_i, \pi_i]$.

($\implies$) Conversely, suppose that π is an ϵ-Nash equilibrium. Fix i and an arbitrary $\mu_i' \in M_i$, and choose a strategy π_i' that implements μ_i'. Then

$$U_i(\mu) = U_i(\pi) \geq U_i(\pi_i', \pi_{-i}) - \epsilon = U_i(\mu_i', \mu_{-i}) - \epsilon.$$

Taking the maximum over $\mu_i' \in M_i$ proves the claim. □

B.3 Proof of Proposition 4.3

Proposition 4.3 (Minimax theorem for 2p0s-GSAS)**.** *For a 2p0s-GSAS, a strategy profile (π_1^*, π_2^*) is a NE if and only if the associated (μ_1^*, μ_2^*) is a saddle point of the function $U = U_1 = -U_2$, i.e., $\mu_1^* \in \operatorname{argmax}_{\mu_1 \in M_1} \min_{\mu_2 \in M_2} U(\mu_1, \mu_2)$ and $\mu_2^* \in \operatorname{argmin}_{\mu_2 \in M_2} \max_{\mu_1 \in M_1} U(\mu_1, \mu_2)$, where $\max_{\mu_1 \in M_1} \min_{\mu_2 \in M_2} U(\mu_1, \mu_2) = \min_{\mu_2 \in M_2} \max_{\mu_1 \in M_1} U(\mu_1, \mu_2)$.*

Proof. The proof follows directly from von Neumann's minimax theorem (von Neumann, 1928) and the fact that $U(\pi) = U(\mu)$ if π implements μ. □

B.4 Proof of Proposition 4.4

Proposition 4.4. *Consider a GSAS $\mathcal{G} = (\mathcal{G}_{\text{orig}}, \mathcal{S}, \rho)$. Let $x^* = (x_1^*, \ldots, x_n^*)$ be an ϵ-NE of $\mathcal{G}_{\text{orig}}$, $x_i^* \in \Delta(A_i)$. If $\mu^* = x^*$ is implementable in $\mathcal{G}$ by $\pi^* = (\pi_1^*, \ldots, \pi_n^*)$, then π^* is an ϵ-NE in $\mathcal{G}$.*

Proof. Let π^* implement $\mu^* = x^*$. Any unilateral deviation π_i' in the GSAS induces an implementable marginal $\mu_i' \in M_i \subseteq \Delta(A_i)$. Since x^* is an ϵ-Nash equilibrium of G_{orig},

$$u_i(x_i^*, x_{-i}^*) \geq u_i(\mu_i', x_{-i}^*) - \epsilon.$$

As π^* implements x^*,

$$U_i(\pi^*) = u_i(x_i^*, x_{-i}^*), \quad U_i(\pi_i', \pi_{-i}^*) = u_i(\mu_i', x_{-i}^*).$$

Therefore, $U_i(\pi^*) \geq U_i(\pi_i', \pi_{-i}^*) - \epsilon$ for every player i and every unilateral deviation π_i', and by definition π^* is an ϵ-Nash equilibrium of G. □

B.5 Proof of Theorem 4.5

Theorem 4.5. *Let π_i implement $\mu_i \in M_i$. Then, $\forall \delta \in (0, 1)$, there exists $w_i \in \Delta(A_i)$ prescribing a strategy π_i' where $\pi_i'(a_i \mid S_i) := \frac{w_i(a_i)\mathbb{1}\{a_i \in S_i\}}{\sum_{a_i' \in A_i} w_i(a_i')\mathbb{1}\{a_i' \in S_i\}}$. Furthermore, π_i' induces a marginal μ_i' with total variation distance $d_{\text{TV}}(\mu_i, \mu_i') < \delta$.*

Proof. We start by assuming that every action is available with strictly positive probability, i.e., $\forall i \in [n] \, \forall a_i \in A_i$, $\Pr_{S_i \sim \rho_i}[a_i \in S_i] > 0$, and that every availability set is available with strictly positive probability, i.e., $\forall i \in [n] \, \forall S_i \in \mathcal{S}_i$, $\rho_i(S_i) > 0$. At the end of the proof, we will relax this assumption.

The proof will require the following linear algebraic result.

Theorem B.1 (Menon & Schneider (1969))**.** *Let $X \in \mathbb{R}_{\geq 0}^{n \times m}$ be a nonnegative matrix and $r \in \mathbb{R}_{>0}^n$, $c \in \mathbb{R}_{>0}^m$. Then there exist $u \in \mathbb{R}_{>0}^n$ and $v \in \mathbb{R}_{>0}^m$ such that $P = \operatorname{diag}(u) X \operatorname{diag}(v)$ has row sums r and column sums c if and only if there exists a nonnegative matrix $Y \in \mathbb{R}_{\geq 0}^{n \times m}$ with row sums r, column sums c, and $\operatorname{supp}(Y) = \operatorname{supp}(X)$.*

Consider any player $i \in [n]$ and, overloading notation for the purposes of this proof, fix some $\epsilon \in (0, \frac{\delta}{|A_i|})$. We will use Theorem B.1 to show the existence of w_i. Going forward, we apply

Theorem B.1 for availability sets that have strictly positive support on ρ. For each availability set $S_i \in \mathcal{S}_i$, let us define a conditional distribution $\tilde{\pi}_i^{(\epsilon)}(S_i)$ as

$$\tilde{\pi}_i^{(\epsilon)}(a_i \mid S_i) := \begin{cases} (1-\epsilon)\pi_i(a_i \mid S_i) + \frac{\epsilon}{|S_i|}, & \text{if } a_i \in \mathcal{S}_i, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

This perturbation term $\frac{\epsilon}{|S_i|}$ ensures that $\tilde{\pi}_i^{(\epsilon)}$ has full support on every S_i. Let $\tilde{\mu}_i^{(\epsilon)}$ be the corresponding marginal distribution over A_i, i.e., $\tilde{\mu}_i^{(\epsilon)}(a_i) := \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i)\tilde{\pi}_i^{(\epsilon)}(a_i \mid S_i)$.

Let us also define a matrix $X_i \in \{0,1\}^{|\mathcal{S}_i| \times |A_i|}$ as the indicator of action availability, i.e., $X_i(S_i, a_i) := \mathbb{1}\{a_i \in S_i\}$, and a nonnegative matrix $Y_i^{(\epsilon)} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}_i| \times |A_i|}$ with $Y_i^{(\epsilon)}(S_i, a_i) := \rho_i(S_i)\tilde{\pi}_i^{(\epsilon)}(a_i \mid S_i)$.

The row sums of $Y_i^{(\epsilon)}$ are given by

$$\sum_{a_i \in A_i} Y_i^{(\epsilon)}(S_i, a_i) = \sum_{a_i \in A_i} \rho_i(S_i)\tilde{\pi}_i^{(\epsilon)}(a_i \mid S_i) \tag{8}$$

$$= \rho_i(S_i) \left[\sum_{a_i \in A_i} \mathbb{1}\{a_i \in \mathcal{S}_i\} \left((1-\epsilon)\pi_i(a_i \mid S_i) + \frac{\epsilon}{|S_i|} \right) \right] \tag{9}$$

$$= \rho_i(S_i) \left[(1-\epsilon) \sum_{a_i \in S_i} [\pi_i(a_i \mid S_i)] + \epsilon \left(\sum_{a_i \in S_i} \frac{1}{|S_i|} \right) \right] \tag{10}$$

$$= \rho_i(S_i). \quad \text{(as both inner sums equal 1)} \tag{11}$$

The column sums of $Y_i^{(\epsilon)}$ are given by $\sum_{S_i \in \mathcal{S}_i} Y_i^{(\epsilon)}(S_i, a_i) = \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i)\tilde{\pi}_i^{(\epsilon)}(a_i \mid S_i) =: \tilde{\mu}_i^{(\epsilon)}(a_i)$, which is the marginal distribution of $\tilde{\pi}_i^{(\epsilon)}$.

Finally, note that $Y_i^{(\epsilon)}(S_i, a_i) \neq 0$ if and only if $a_i \in S_i$. This follows from the definition of $\tilde{\pi}_i^{(\epsilon)}(a_i \mid S_i)$ (Equation (7)). Thus, $\text{supp}(Y_i^{(\epsilon)}) = \{(S_i, a_i) \mid a_i \in S_i\} = \text{supp}(X_i)$. Since the matrix $Y_i^{(\epsilon)}$ has row sums $\rho_i(S_i)$, column sums $\tilde{\mu}_i^{(\epsilon)}$ and $\text{supp}(Y_i^{(\epsilon)}) = \text{supp}(X_i)$, the hypothesis of Theorem B.1 holds for the triple $(X_i, \rho_i, \tilde{\mu}_i^{(\epsilon)})$. Thus, there exist vectors $u_i^{(\epsilon)} \in \mathbb{R}_{>0}^{|\mathcal{S}_i|}$ and $v_i^{(\epsilon)} \in \mathbb{R}_{>0}^{|A_i|}$ such that $P_i^{(\epsilon)} = \text{diag}\left(u_i^{(\epsilon)}\right) X_i \, \text{diag}\left(v_i^{(\epsilon)}\right)$ has row sums ρ_i and column sums $\tilde{\mu}_i^{(\epsilon)}$.

Now, we can define a probability distribution $\pi_i'(S_i)$ over A_i for every availability set S_i as

$$\pi_i'(a_i \mid S_i) := \frac{P_i^{(\epsilon)}(S_i, a_i)}{\sum_{a_i' \in A_i} P_i^{(\epsilon)}(S_i, a_i')} \tag{12}$$

$$= \frac{u_i^{(\epsilon)}(S_i) X_i(S_i, a_i) v_i^{(\epsilon)}(a_i)}{\sum_{a_i' \in A_i} u_i^{(\epsilon)}(S_i) X_i(S_i, a_i') v_i^{(\epsilon)}(a_i')} \tag{13}$$

$$= \frac{X_i(S_i, a_i) v_i^{(\epsilon)}(a_i)}{\sum_{a_i' \in A_i} X_i(S_i, a_i') v_i^{(\epsilon)}(a_i')} \tag{14}$$

$$= \frac{v_i^{(\epsilon)}(a_i) \mathbb{1}\{a_i \in S_i\}}{\sum_{a_i' \in A_i} v_i^{(\epsilon)}(a_i') \mathbb{1}\{a_i' \in S_i\}} \tag{15}$$

$$= \frac{\frac{v_i^{(\epsilon)}(a_i)}{\|v_i^{(\epsilon)}\|_1} \mathbb{1}\{a_i \in S_i\}}{\sum_{a_i' \in A_i} \frac{v_i^{(\epsilon)}(a_i')}{\|v_i^{(\epsilon)}\|_1} \mathbb{1}\{a_i' \in S_i\}} \tag{16}$$

$$= \frac{w_i(a_i) \mathbb{1}\{a_i \in S_i\}}{\sum_{a_i' \in A_i} w_i(a_i') \mathbb{1}\{a_i' \in S_i\}}, \tag{17}$$

where we define $w_i(a_i) := \frac{v_i^{(\epsilon)}(a_i)}{\|v_i^{(\epsilon)}\|_1}$ and note that due to the rescaling, $w_i \in \Delta(A_i)$. Next, the marginal probability distribution $\mu_i' \in M_i$ induced by π_i' is

$$\mu_i'(a_i) = \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i)\pi_i'(a_i \mid S_i) \tag{18}$$

$$= \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i) \frac{P_i^{(\epsilon)}(S_i, a_i)}{\sum_{a_i' \in A_i} P_i^{(\epsilon)}(S_i, a_i')}, \tag{19}$$

where the latter equality follows by Equation (12). As the row sums of $P_i^{(\epsilon)}$ are ρ_i, we have $\sum_{a_i' \in A_i} P_i^{(\epsilon)}(S_i, a_i') = \rho_i(S_i)$ and so

$$\mu_i'(a_i) = \sum_{S_i \in \mathcal{S}_i} P_i^{(\epsilon)}(S_i, a_i).$$

By Theorem B.1, $P_i^{(\epsilon)}$ and $Y_i^{(\epsilon)}$ have equal column sums and thus

$$\mu_i'(a_i) = \sum_{S_i \in \mathcal{S}_i} Y_i^{(\epsilon)}(S_i, a_i) \tag{20}$$

$$= \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i)\tilde{\pi}_i^{(\epsilon)}(a_i \mid S_i) \tag{21}$$

$$= \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i)\mathbb{1}\{a_i \in S_i\} \left[(1-\epsilon)\pi_i(a_i \mid S_i) + \frac{\epsilon}{|S_i|}\right], \tag{22}$$

where Equation (21) follows from the definition of $Y_i^{(\epsilon)}$. Rearranging and noting $\pi_i(a_i \mid S_i) = 0$ when $a_i \notin S_i$,

$$\mu_i'(a_i) = (1-\epsilon) \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i)\pi_i(a_i \mid S_i) + \sum_{S_i \in \mathcal{S}_i} \frac{\epsilon \rho_i(S_i)\mathbb{1}\{a_i \in S_i\}}{|S_i|} \tag{23}$$

$$= (1-\epsilon)\mu_i(a_i) + \sum_{S_i \in \mathcal{S}_i} \frac{\epsilon \rho_i(S_i)\mathbb{1}\{a_i \in S_i\}}{|S_i|} \tag{24}$$

$$= \mu_i(a_i) + \epsilon \left(-\mu_i(a_i) + \sum_{S_i \in \mathcal{S}_i} \frac{\rho_i(S_i)\mathbb{1}\{a_i \in S_i\}}{|S_i|}\right). \tag{25}$$

We can now bound $\mu_i'(a_i)$ on both sides. Using Equation (25) and since $\mu_i(a_i) \in [0, 1]$,

$$\mu_i'(a_i) \geq \mu_i(a_i) - \epsilon\mu_i(a_i) \tag{26}$$

$$\geq \mu_i(a_i) - \epsilon. \tag{27}$$

Similarly, from Equation (25) and utilizing the fact that $|S_i| \geq 1$,

$$\mu_i'(a_i) \leq \mu_i(a_i) + \epsilon \sum_{S_i \in \mathcal{S}_i} \frac{\rho_i(S_i)\mathbb{1}\{a_i \in S_i\}}{|S_i|} \tag{28}$$

$$\leq \mu_i(a_i) + \epsilon \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i) \tag{29}$$

$$\leq \mu_i(a_i) + \epsilon. \tag{30}$$

Thus, for all $a_i \in A_i$, $|\mu_i'(a_i) - \mu_i(a_i)| \leq \epsilon$.

The total variation distance between μ_i and μ_i' can then be bounded by

$$d_{\mathrm{TV}}(\mu_i, \mu_i') = \frac{1}{2} \sum_{a_i \in A_i} |\mu_i(a_i) - \mu_i'(a_i)| \tag{31}$$

$$\leq \frac{1}{2} \sum_{a_i \in A_i} \epsilon \tag{32}$$

$$\leq \frac{1}{2} |A_i| \frac{\delta}{|A_i|} \tag{33}$$

$$< \delta, \tag{34}$$

where Equation (33) uses $\epsilon < \delta/|A_i|$ and Equation (34) uses $\delta > 0$.

It now remains to relax our assumption requiring all actions to be available with nonzero probability, and all availability sets to be observed with nonzero probability. Indeed, any GSAS $\mathcal{G}$ not meeting this assumption can be reduced to an equivalent GSAS $\mathcal{G}'$ with actions and availability sets that are never available removed. Obviously, this process does not affect the game structure and the player strategies. The compact vector w_i' in the reduced GSAS $\mathcal{G}'$ can then be converted to an equivalent compact vector w_i in the original GSAS $\mathcal{G}$ by setting the entries for never-available actions to zero, and leaving entries for observed actions unchanged. This completes the proof. □

B.6 Illustration of proof via Example 4.6

We illustrate the proof of Theorem 4.5 using Example 4.6 from the main text. Utilizing ideas from the proof of Theorem 4.5 and the statement of Theorem B.1, we define a matrix $X_i \in \{0,1\}^{|\mathcal{S}_i| \times |A_i|}$ as the indicator of action availability, i.e., $X_i(S_i, a_i) := \mathbb{1}\{a_i \in S_i\}$, and a nonnegative matrix $Y_i^{(\epsilon)} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}_i| \times |A_i|}$ with $Y_i^{(\epsilon)}(S_i, a_i) := \rho_i(S_i)\tilde{\pi}_i^{(\epsilon)}(a_i \mid S_i)$, where $\tilde{\pi}_i^{(\epsilon)}$ is a conditional distribution over actions defined in Equation (7). Then for every $\epsilon \in (0,1)$, the matrix $Y_i^{(\epsilon)}$ has row sums ρ_i, column sums $\tilde{\mu}_i^{(\epsilon)}$, and $\mathrm{supp}(Y_i^{(\epsilon)}) = \mathrm{supp}(X_i)$. For the row player, the above construction gives us matrices X_1 and $Y_1^{(\epsilon)}$ given by

$$X_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } Y_1^{(\epsilon)} = \begin{pmatrix} \frac{3-\epsilon}{12} & \frac{\epsilon}{6} & \frac{3-\epsilon}{12} \\ 0 & \frac{2-\epsilon}{4} & \frac{\epsilon}{4} \end{pmatrix}.$$

Theorem B.1 guarantees that there exist vectors $u_i^{(\epsilon)} \in \mathbb{R}_{>0}^{|\mathcal{S}_i|}$ and $v_i^{(\epsilon)} \in \mathbb{R}_{>0}^{|A_i|}$ such that $P_i^{(\epsilon)} = \mathrm{diag}\left(u_i^{(\epsilon)}\right) X_i \,\mathrm{diag}\left(v_i^{(\epsilon)}\right)$ has row sums ρ_i and column sums $\tilde{\mu}_i^{(\epsilon)}$. In our case, by solving the system

$$\begin{cases} \mathrm{diag}(u_1^{(\epsilon)}) X_1 \,\mathrm{diag}(v_1^{(\epsilon)})\mathbf{1} = Y_1^{(\epsilon)}\mathbf{1} \\ \mathrm{diag}(v_1^{(\epsilon)}) X_1^\top \,\mathrm{diag}(u_1^{(\epsilon)})\mathbf{1} = {Y_1^{(\epsilon)}}^\top \mathbf{1}, \end{cases}$$

we get

$$u_1^{(\epsilon)} = \lambda \left(1, \frac{6}{3+\epsilon}\right) \text{ and } v_1^{(\epsilon)} = \frac{1}{12\lambda} \left(3-\epsilon, \frac{(6-\epsilon)(3+\epsilon)}{\epsilon+9}, \frac{(3+2\epsilon)(3+\epsilon)}{\epsilon+9}\right),$$

for any constant $\lambda > 0$. Setting $\lambda = \frac{1}{2}$ and letting $\epsilon \to 0$, we get $v_1^{(\epsilon)} \to w_1 = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)$. With similar calculations, we obtain $w_2 = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right)$. With w_1 and w_2, we have $\pi_1(S_{11}) = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)$, $\pi_1(S_{12}) = \left(0, \frac{2}{3}, \frac{1}{3}\right)$, $\pi_2(S_{21}) = \left(\frac{2}{3}, 0, \frac{1}{3}\right)$, and $\pi_2(S_{22}) = \left(0, \frac{2}{3}, \frac{1}{3}\right)$, which are consistent with the marginal distribution μ^* and also yield expected payoffs of 0 for each player. Therefore, π is also a Nash equilibrium.

B.7 Proof of Proposition 5.3

Proposition 5.3. *Consider a 2p0s-GSAS $\mathcal{G}$ satisfying Assumption 3.3. For every sequence of strategies,* $\mathrm{SPR}(\bar{\mu}_1^T, \bar{\mu}_2^T) = \frac{R_{T,1}^{\mathrm{rank}} + R_{T,2}^{\mathrm{rank}}}{T} =: \epsilon_T$. *Consequently, every strategy profile implementing $(\bar{\mu}_1^T, \bar{\mu}_2^T)$ is an ϵ_T-Nash equilibrium. In particular, if $R_{T,i}^{\mathrm{rank}} = o(T)$ for both players, then the time-averaged marginals converge to the set of NE of $\mathcal{G}$ in terms of SPR.*

Proof. We first show that in GSAS, rankings contain a best fixed response in hindsight. Fix a player i and define

$$G_i(a_i) := \sum_{t=1}^{T} U_i(a_i; \pi_{-i}^t), \quad a_i \in A_i.$$

For every π_i,

$$\begin{aligned}\sum_{t=1}^{T} U_i(\pi_i, \pi_{-i}^t) &= \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i) \sum_{a_i \in S_i} \pi_i(a_i \mid S_i) G_i(a_i) \\ &\leq \sum_{S_i \in \mathcal{S}_i} \rho_i(S_i) \max_{a_i \in S_i} G_i(a_i).\end{aligned}$$

A ranking that orders the actions in decreasing order of G_i attains the last expression for every S_i, with arbitrary tie-breaking. Therefore,

$$\max_{\tau_i \in \mathcal{T}_i} \sum_{t=1}^{T} U_i(\pi_i^{\tau_i}, \pi_{-i}^t) = \max_{\pi_i} \sum_{t=1}^{T} U_i(\pi_i, \pi_{-i}^t).$$

Let $U := U_1 = -U_2$. The marginal representation of payoffs gives

$$\begin{aligned}R_{T,1}^{\mathrm{rank}} + R_{T,2}^{\mathrm{rank}} &= \max_{\mu_1 \in M_1} \sum_{t=1}^{T} U(\mu_1, \mu_2^t) - \min_{\mu_2 \in M_2} \sum_{t=1}^{T} U(\mu_1^t, \mu_2) \\ &= T \left[\max_{\mu_1 \in M_1} U(\mu_1, \bar{\mu}_2^T) - \min_{\mu_2 \in M_2} U(\bar{\mu}_1^T, \mu_2) \right],\end{aligned}$$

where the second equality follows from bilinearity. The gap is the sum of the two unilateral-deviation gains at $(\bar{\mu}_1^T, \bar{\mu}_2^T)$, so each gain is at most ϵ_T. Combining the above with Proposition 4.2 then gives the result for every strategy profile implementing these marginals. □

Corollary B.2 (Empirical-marginal residual)**.** *For $x_i \in \Delta(A_i)$, define*

$$\widehat{\mathrm{SPR}}(x_1, x_2) := \max_{\mu_1 \in M_1} U(\mu_1, x_2) - \min_{\mu_2 \in M_2} U(x_1, \mu_2).$$

Then, for every $p \in (0, 1)$, with probability at least $1 - p$,

$$\widehat{\mathrm{SPR}}(\tilde{\mu}_1^T, \tilde{\mu}_2^T) \leq \frac{R_{T,1}^{\mathrm{rank}} + R_{T,2}^{\mathrm{rank}}}{T} + \sqrt{\frac{2|A_1|}{pT}} + \sqrt{\frac{2|A_2|}{pT}}.$$

Proof. By concentration of the empirical marginals and a union bound, with probability at least $1 - p$,

$$\|\tilde{\mu}_i^T - \bar{\mu}_i^T\|_1 \leq \sqrt{\frac{2|A_i|}{pT}}, \quad i \in \{1, 2\}.$$

Because $u_i \in [-1, 1]$, expected payoffs are 1-Lipschitz in each marginal under the ℓ_1-norm. Hence,

$$\widehat{\mathrm{SPR}}(\tilde{\mu}_1^T, \tilde{\mu}_2^T) \leq \mathrm{SPR}(\bar{\mu}_1^T, \bar{\mu}_2^T) + \sum_{i=1}^{2} \|\tilde{\mu}_i^T - \bar{\mu}_i^T\|_1.$$

The result follows from Proposition 5.3. □

B.8 Counterexamples for Sleeping External and Internal Regret (Remark 5.4)

B.8.1 Sleeping External Regret

We first define a notion of Sleeping External Regret (SE-regret) which was introduced in (Blum & Mansour, 2007; Kleinberg et al., 2010).

Definition B.3 (Sleeping External Regret)**.** *For any action $a_i' \in A_i$, the* sleeping external regret *(*SE-regret*) for player i is defined as:*

$$R_{T,i}^{\mathsf{EXT}}(a_i') \coloneqq \mathbb{E}_{S\sim\rho}\left[\mathbb{E}_{a\sim\pi(S)}\left[\sum_{t=1}^{T}\mathbb{1}\{a_i'\in S_i^t\}(u_i(a_i',a_{-i}^t)-u_i(a_i^t,a_{-i}^t))\right]\right].$$

In other words, sleeping external regret captures player i's expected payoff gain from switching to a_i' whenever it is available, regardless of the action originally chosen. In the case where a player's SE-regret grows sublinearly, i.e., $\max_{a_i'} R_{T,i}^{\mathsf{EXT}}(a_i') = o(T)$, they are said to have *no-SE-regret*.

Counterexample for SE-regret We construct a single-player GSAS $\mathcal{G} = (\mathcal{G}_{\text{orig}}, \mathcal{S}, \rho)$ where $\mathcal{G}_{\text{orig}}$ has three actions a_1, a_2, a_3. Let $\mathcal{S} = \{S_1 = \{a_1, a_2\}, S_2 = \{a_2, a_3\}\}$ with $\rho(S_1) = \rho(S_2) = 0.5$. The payoff function is given as $u(a_1) = 0.01$, $u(a_2) = 0.02$ and $u(a_3) = 1$. Suppose the player plays uniformly in S_1 and plays a_3 with probability 1 in S_2. The SE-regret per round for each action is as follows:

$$\begin{aligned}
\frac{R_T^{\mathsf{EXT}}(a_1)}{T} &= \frac{1}{2}(0.01-0.015) = -0.0025,\\
\frac{R_T^{\mathsf{EXT}}(a_2)}{T} &= \frac{1}{2}\left((0.02-1)+(0.02-0.015)\right) = -0.4875,\\
\frac{R_T^{\mathsf{EXT}}(a_3)}{T} &= 0.
\end{aligned}$$

Thus for each action a, $[R_T^{\mathsf{EXT}}(a)]_+ = 0$; hence it has no-SE-regret. But the strategy is not a Nash equilibrium, since a profitable deviation would be to play a_2 with probability 1 in S_1 instead; its gain is $\frac{1}{2}(0.02-0.015) = 0.0025$ per round. Therefore, no-SE-regret does not suffice to guarantee convergence to Nash equilibrium in GSAS.

B.8.2 Sleeping Internal Regret

Sleeping internal regret (SI-regret) is a stronger notion of sleeping regret introduced in Gaillard et al. (2023) to study the sleeping bandit problem (Kleinberg et al., 2010).

Definition B.4 (Sleeping Internal Regret)**.** *For any pair of actions $\hat{a}_i \in A_i$ and $\hat{a}_i' \in A_i$, the* sleeping internal regret *(*SI-regret*) for player i in T rounds is*

$$R_{T,i}^{\mathsf{INT}}(\hat{a}_i \to \hat{a}_i') \coloneqq \mathbb{E}_{S\sim\rho}\left[\mathbb{E}_{a\sim\pi(S)}\left[\sum_{t=1}^{T}\mathbb{1}\{a_i^t=\hat{a}_i, \hat{a}_i'\in S_i^t\}\cdot\left(u_i(\hat{a}_i',a_{-i}^t)-u_i(a_i^t,a_{-i}^t)\right)\right]\right].$$

Similar to SE-regret, if a player's SI-regret grows sublinearly for each action pair $(\hat{a}_i, \hat{a}_i')$, i.e., $\max_{\hat{a}_i,\hat{a}_i'} R_{T,i}^{\mathsf{INT}}(\hat{a}_i \to \hat{a}_i') = o(T)$, they are said to have *no-SI-regret*. The intuition is that player i does not regret not playing action $\hat{a}_i'$ (if $\hat{a}_i'$ was available) every time they played $\hat{a}_i$, for any $\hat{a}_i$, $\hat{a}_i'$. With respect to the counterexample for SE-regret above, one can easily see that the SI-regret is *not* sublinear in T, since $R_T^{\mathsf{INT}}(a_1 \to a_2)/T = 1/2 \cdot 1/2 \cdot (0.02 - 0.01) = 0.0025 > 0$ at each round.

Analogous to the relationship between the non-sleeping variants of internal and external regret, no-SI-regret implies no-SE-regret (though the converse does not hold). This follows from the definitions of SI-regret and SE-regret:

$$R_{T,i}^{\mathsf{EXT}}(a_i') = \sum_{a_i\in A_i} R_{T,i}^{\mathsf{INT}}(a_i \to a_i').$$

Nevertheless, even though SI-regret is stronger than SE-regret, we show that minimizing SI-regret does *not* suffice for Nash convergence.

Counterexample for SI-regret We construct a 2p0s-GSAS $\mathcal{G} = (\mathcal{G}_{\text{orig}}, \mathcal{S}, \rho)$. In $\mathcal{G}_{\text{orig}}$, player 1 has actions x, y, z while player 2 has actions L, R. The payoff matrix is given as

	L	R
x	0	1
y	1	0
z	1	1

Let $\mathcal{S}_1 = \{S_{11}, S_{12}\}$, where $S_{11} = \{x, y\}$ and $S_{12} = \{x, y, z\}$, with $\rho_1(S_{11}) = \rho_1(S_{12}) = 1/2$. Let $\mathcal{S}_2 = \{S_{21}\}$, where $S_{21} = \{L, R\}$. Thus action z is available to player 1 with probability $1/2$, while both actions are available for player 2.

Consider the strategy π_i^t defined as follows:

- when t is odd, set $\pi_1^t(S_{11}) = x, \pi_1^t(S_{12}) = y, \pi_2^t(S_{21}) = L$;
- when t is even, set $\pi_1^t(S_{11}) = y, \pi_1^t(S_{12}) = x, \pi_2^t(S_{21}) = R$.

Thus, on odd rounds player 1 plays x from S_{11} and y from S_{12}, while player 2 plays L; on even rounds player 1 reverses x and y, while player 2 plays R. We now show that both players have no-SI-regret for even T.

For player 1, consider the deviation $x \to y$. On an odd round, player 2 plays L and this deviation is triggered when $S_1^t = S_{11}$, which has probability $1/2$. Its gain is $u_1(y, L) - u_1(x, L) = 1 - 0 = 1$. Similarly, on an even round, player 2 plays R and this deviation is triggered when $S_1^t = S_{12}$, which has probability $1/2$. Its gain is $u_1(y, R) - u_1(x, R) = 0 - 1 = -1$. Hence they cancel out and $R_{T,1}^{\mathsf{INT}}(x \to y) = 0$. Similarly $R_{T,1}^{\mathsf{INT}}(y \to x) = 0$.

Now consider deviations into z. As z is only available with S_{12}, we have $u_1(z, L) - u_1(y, L) = u_1(z, R) - u_1(x, R) = 0$. Therefore we also have $R_{T,1}^{\mathsf{INT}}(x \to z) = R_{T,1}^{\mathsf{INT}}(y \to z) = 0$.

For player 2, it is easy to show by symmetry that $R_{T,2}^{\mathsf{INT}}(L \to R) = R_{T,2}^{\mathsf{INT}}(R \to L) = 0$. Therefore, we have $R_{T,1}^{\mathsf{INT}} = R_{T,2}^{\mathsf{INT}} = 0$ as required. By definition, our strategy profile π_i induces

$$\bar{\mu}_1 = \left(\frac{1}{2}, \frac{1}{2}, 0\right), \quad \bar{\mu}_2 = \left(\frac{1}{2}, \frac{1}{2}\right).$$

We can calculate the corresponding payoff as $U_1(\bar{\mu}_1, \bar{\mu}_2) = 1/2$. Now, consider the alternative strategy π_1' for player 1 given as $\pi_1'(S_{11}) = y, \pi_1'(S_{12}) = z$, which induces

$$\mu_1' = \left(0, \frac{1}{2}, \frac{1}{2}\right),$$

with payoff

$$\begin{aligned} U_1(\mu_1', \bar{\mu}_2) &= \frac{1}{2}\left(\frac{u_1(y, L) + u_1(y, R)}{2}\right) + \frac{1}{2}\left(\frac{u_1(z, L) + u_1(z, R)}{2}\right) \\ &= \frac{1}{2}\left(\frac{1}{2}\right) + \frac{1}{2}(1) = \frac{3}{4}. \end{aligned}$$

Hence we have that

$$U_1(\mu_1', \bar{\mu}_2) - U_1(\bar{\mu}_1, \bar{\mu}_2) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4} > 0,$$

and therefore $(\bar{\mu}_1, \bar{\mu}_2)$ is not a Nash equilibrium. Hence, no-SI-regret does not guarantee convergence to NE in GSAS.

B.9 CAT-FTPL Details

Using player i's observed payoff, we define the observed loss $\ell_{i,t}(a_i) := \frac{1}{2}(1 - u_i(a_i, a_{-i}^t)) \in [0, 1]$. Let $\widehat{L}_i^{t-1}$ be the cumulative estimated-loss vector. CAT-FTPL draws independent standard Gumbel

perturbations $Z_{i,t}(a_i)$ and plays

$$a_i^t \in \operatorname*{argmin}_{a_i \in S_i^t} \left\{ \eta_i \widehat{L}_i^{t-1}(a_i) - Z_{i,t}(a_i) \right\}. \tag{35}$$

Equivalently, the induced strategy for player i is

$$\pi_i^t(a_i \mid S_i) = \frac{\mathbb{1}\{a_i \in S_i\} e^{-\eta_i \widehat{L}_i^{t-1}(a_i)}}{\sum_{b_i \in S_i} e^{-\eta_i \widehat{L}_i^{t-1}(b_i)}}. \tag{36}$$

Thus, each perturbation vector induces a random ranking, but the algorithm stores only one score per action of the GSAS. Going forward, π_i^t denotes the induced strategy as in (36), and a_i^t denotes its sampled action. It remains to estimate the losses of actions that are not observed at a given round. For each $a_i \in S_i^t$, draw auxiliary sets $\widetilde{S}_{i,t,1}, \widetilde{S}_{i,t,2}, \ldots \overset{\text{i.i.d.}}{\sim} \rho_i$ and independent stopping indicators $B_{i,t,k}(a_i) \sim \text{Bernoulli}(\gamma_i/(1+\gamma_i))$. Define the number of trials until the first success as

$$H_{i,t}(a_i) := \min\{k \geq 1 : a_i \in \widetilde{S}_{i,t,k} \text{ or } B_{i,t,k}(a_i) = 1\}. \tag{37}$$

Putting these elements together, we present CAT-FTPL in Algorithm 1.

Algorithm 1 CAT-FTPL for player i

Input: $\eta_i, \gamma_i > 0$
$\widehat{L}_i^0 \leftarrow 0 \in \mathbb{R}_{\geq 0}^{|A_i|}$
for $t = 1, \ldots, T$ **do**
  Observe S_i^t, draw Gumbel perturbations and select a_i^t according to (35)
  Observe $\ell_{i,t}(a_i)$ for every $a_i \in S_i^t$
  $\widehat{\ell}_{i,t} \leftarrow 0 \in \mathbb{R}_{\geq 0}^{|A_i|}$
  for $a_i \in S_i^t$ **do**
    Generate $H_{i,t}(a_i)$ by (37) and set $\widehat{\ell}_{i,t}(a_i) \leftarrow \ell_{i,t}(a_i) H_{i,t}(a_i)/(1+\gamma_i)$
  end for
  $\widehat{L}_i^t \leftarrow \widehat{L}_i^{t-1} + \widehat{\ell}_{i,t}$
end for

B.10 Proof of Theorem 5.5

Theorem 5.5. *A player using CAT-FTPL has expected pathwise ranking regret bounded as*

$$\mathbb{E}\left[R_{T,i}^{\text{rank}}\right] \leq 2\left(\log|A_i|\left(\frac{1}{\eta_i} + \frac{1}{\gamma_i}\right) + |A_i| T(\eta_i + \gamma_i)\right). \tag{6}$$

In particular, with parameters $\eta_i = \gamma_i = \sqrt{\frac{\log|A_i|}{|A_i|T}}$, *we obtain* $\mathbb{E}\left[R_{T,i}^{\text{rank}}\right] \leq 8\sqrt{|A_i| T \log|A_i|}$.

Proof. Let $\mathcal{F}_{t-1}$ be the history before round t, i.e., the information generated by all play and algorithmic randomness revealed through the end of round $t-1$. Then under simultaneous play and Assumption 3.3, we have $S_i^t \perp a_{-i}^t \mid \mathcal{F}_{t-1}$. Now fix a player i and omit its index. Write $d := |A|$, $q(a) := \Pr(a \in S_t)$, and $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}_{t-1}]$. Let

$$\bar{\ell}_t(a) := \mathbb{E}_t[\ell_t(a)] = \frac{1 - U_i(a; \pi_{-i}^t)}{2},$$

and let μ_t and μ^τ be the marginals induced by π^t and ranking τ, respectively. As action a can be selected only when it is available, every implementable marginal obeys

$$\mu(a) \leq q(a). \tag{38}$$

We first record the properties of the loss estimator. Conditional on $\mathcal{F}_{t-1}$, $H_t(a)$ is geometric on $\{1, 2, \ldots\}$ with parameter $p_a = (q(a)+\gamma)/(1+\gamma)$ and is independent of $\mathbb{1}\{a \in S_t\}\ell_t(a)$. Therefore

$$\mathbb{E}_t[\hat{\ell}_t(a)] = \frac{q(a)}{q(a)+\gamma}\bar{\ell}_t(a), \quad \mathbb{E}_t[\hat{\ell}_t(a)^2] \le \frac{2q(a)}{(q(a)+\gamma)^2}. \tag{39}$$

We also have the one-sided exponential-moment bound

$$\mathbb{E}_t\left[e^{\gamma(\hat{\ell}_t(a)-\bar{\ell}_t(a))}\right] \le 1. \tag{40}$$

To see why this holds, let $s := \gamma/(1+\gamma)$ and $M_a(s) := \mathbb{E}[e^{sH_t(a)}]$. Since $e^{sHx} \le 1 + x(e^{sH} - 1)$ for $x \in [0, 1]$,

$$\mathbb{E}_t[e^{\gamma\hat{\ell}_t(a)}] \le 1 + q(a)\bar{\ell}_t(a)(M_a(s) - 1).$$

For $q(a) > 0$,

$$M_a(s) - 1 = \frac{e^s - 1}{1 - (1 - p_a)e^s} \le \frac{\gamma}{q(a)},$$

because $s \le \log(1+\gamma)$. The case $q(a) = 0$ is immediate. Hence the above display is at most $1 + \gamma\bar{\ell}_t(a) \le e^{\gamma\bar{\ell}_t(a)}$, proving (40).

Let $\hat{L}_t := \sum_{s=1}^{t} \hat{\ell}_s$ and define the analysis potential

$$\Phi_t := \mathbb{E}_{S\sim\rho_i}\left[\log \sum_{a\in S} e^{-\eta\hat{L}_{t-1}(a)}\right].$$

By the Gumbel-max identity, the strategy induced by (35) is (36). Thus $e^{-x} \le 1 - x + x^2/2$ for $x \ge 0$ gives

$$\Phi_{t+1} - \Phi_t \le -\eta\langle\mu_t, \hat{\ell}_t\rangle + \frac{\eta^2}{2}\sum_a \mu_t(a)\hat{\ell}_t(a)^2. \tag{41}$$

Moreover, $\Phi_1 \le \log d$, and a ranking that orders actions by increasing $\hat{L}_T(a)$ minimizes the cumulative estimated loss in every availability set. Hence

$$\Phi_{T+1} \ge -\eta \min_\tau \sum_{t=1}^{T}\langle\mu^\tau, \hat{\ell}_t\rangle.$$

Summing (41) yields, pathwise,

$$\sum_t\langle\mu_t, \hat{\ell}_t\rangle - \min_\tau\sum_t\langle\mu^\tau, \hat{\ell}_t\rangle \le \frac{\log d}{\eta} + \frac{\eta}{2}\sum_{t,a}\mu_t(a)\hat{\ell}_t(a)^2. \tag{42}$$

By (38) and (39),

$$\mathbb{E}\left[\sum_{t,a}\mu_t(a)\hat{\ell}_t(a)^2\right] \le 2dT. \tag{43}$$

Now define the pathwise ranking regret in terms of conditional mean losses,

$$\mathcal{R}_T^\ell := \sum_t\langle\mu_t, \bar{\ell}_t\rangle - \min_\tau\sum_t\langle\mu^\tau, \bar{\ell}_t\rangle.$$

Adding and subtracting the estimated losses gives

$$\begin{aligned}\mathcal{R}_T^\ell \le &\left[\sum_t\langle\mu_t, \hat{\ell}_t\rangle - \min_\tau\sum_t\langle\mu^\tau, \hat{\ell}_t\rangle\right] \\ &+ \sum_t\langle\mu_t, \bar{\ell}_t - \hat{\ell}_t\rangle + \max_a\sum_t(\hat{\ell}_t(a) - \bar{\ell}_t(a)).\end{aligned}$$

The expected second term is at most $d\gamma T$ by (38)–(39). Iterating (40), followed by log-sum-exp and Jensen's inequality, gives

$$\mathbb{E}\left[\max_a \sum_t (\hat{\ell}_t(a) - \bar{\ell}_t(a))\right] \le \frac{\log d}{\gamma}. \tag{44}$$

The maximum over actions is sufficient because every ranking marginal μ^τ belongs to $\Delta(A_i)$. Combining (42), (43), and (44) gives

$$\mathbb{E}[\mathcal{R}_T^\ell] \le \frac{\log d}{\eta} + \eta d T + d\gamma T + \frac{\log d}{\gamma}.$$

Finally, the normalization for loss implies $R_{T,i}^{\mathrm{rank}} = 2\mathcal{R}_T^\ell$. This proves (6); substituting the prescribed parameters completes the proof. □

B.11 Proof of Proposition 5.6

Proposition 5.6. *Suppose players $i \in [n]$ run CAT-FTPL in a GSAS. Then for any $p_1, \dots, p_n \in (0,1)$ with $\sum_i p_i < 1$, $\Pr\left(\bigcap_{i=1}^n \left\{R_{T,i}^{\mathrm{rank}} \le \frac{16}{p_i}\sqrt{|A_i| T \log |A_i|}\right\}\right) \ge 1 - \sum_{i=1}^n p_i$.*

Proof. First, note that ranking regret could be negative: an adaptive learner may outperform every fixed ranking in hindsight. We nevertheless only need to seek an upper-tail bound. Indeed, write $[x]_+ := \max\{x, 0\}$, and for every $r \ge 0$,

$$\{R_T^{\mathrm{rank}} > r\} = \{[R_T^{\mathrm{rank}}]_+ > r\}.$$

It therefore suffices to control the nonnegative random variable $[R_T^{\mathrm{rank}}]_+$, to which Markov's inequality can be applied. Fix a player and omit its index, and write $d := |A|$. The proof of Theorem 5.5 gives, pathwise,

$$\frac{1}{2} R_T^{\mathrm{rank}} \le \frac{\log d}{\eta} + \frac{\eta}{2} V_T + L_T + C_T, \tag{45}$$

where

$$V_T := \sum_{t,a} \mu_t(a) \hat{\ell}_t(a)^2,$$
$$L_T := \sum_t \langle \mu_t, \bar{\ell}_t - \hat{\ell}_t \rangle,$$
$$C_T := \max_a \sum_t \big(\hat{\ell}_t(a) - \bar{\ell}_t(a)\big), \quad \bar{\ell}_t(a) := \mathbb{E}_t[\ell_t(a)].$$

Since the first two terms on the right are nonnegative, taking positive parts gives

$$\frac{1}{2}[R_T^{\mathrm{rank}}]_+ \le \frac{\log d}{\eta} + \frac{\eta}{2} V_T + [L_T]_+ + [C_T]_+. \tag{46}$$

We first bound $\mathbb{E}[V_T]$. From (39), we have

$$\mathbb{E}_t[\hat{\ell}_t(a)^2] \le \frac{2q(a)}{(q(a)+\gamma)^2},$$

Consequently,

$$\mathbb{E}_t\left[\sum_{a \in A} \mu_t(a) \hat{\ell}_t(a)^2\right] \le 2 \sum_{a \in A} \frac{\mu_t(a) q(a)}{(q(a)+\gamma)^2}$$
$$\le 2 \sum_{a \in A} \frac{q(a)^2}{(q(a)+\gamma)^2} \le 2d.$$

Summing over t and applying the tower rule therefore gives

$$\mathbb{E}[V_T] \le 2dT. \tag{47}$$

To bound $\mathbb{E}[[L_T]_+]$, let $Y_t := \langle \mu_t, \bar{\ell}_t - \hat{\ell}_t \rangle$. Using (38) and (39),

$$0 \le \mathbb{E}_t[Y_t] = \sum_a \mu_t(a) \frac{\gamma}{q(a)+\gamma} \bar{\ell}_t(a) \le d\gamma.$$

Consider the martingale-difference sequence $D_t := Y_t - \mathbb{E}_t[Y_t]$. Jensen's inequality gives

$$\begin{aligned}\mathbb{E}_t[D_t^2] &\le \sum_a \mu_t(a) \mathbb{E}_t[\hat{\ell}_t(a)^2] \\ &\le 2 \sum_a \frac{\mu_t(a) q(a)}{(q(a)+\gamma)^2} \le 2d.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}[[L_T]_+] &\le d\gamma T + \mathbb{E}\left[\left(\sum_t D_t\right)_+\right] \\ &\le d\gamma T + \sqrt{\sum_t \mathbb{E}[D_t^2]} \le d\gamma T + \sqrt{2dT}.\end{aligned} \tag{48}$$

For $\mathbb{E}[[C_T]_+]$, define

$$Z_{T,a} := \sum_t \big(\hat{\ell}_t(a) - \bar{\ell}_t(a)\big).$$

Recall from (40) that we have

$$\mathbb{E}_t\left[e^{\gamma(\hat{\ell}_t(a) - \bar{\ell}_t(a))}\right] \le 1,$$

and the tower rule implies

$$\mathbb{E}[e^{\gamma Z_{T,a}}] \le 1.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[e^{\gamma [C_T]_+}\right] &\le 1 + \mathbb{E}[e^{\gamma C_T}] \\ &\le 1 + \sum_{a \in A} \mathbb{E}[e^{\gamma Z_{T,a}}] \le d + 1.\end{aligned}$$

Applying Jensen's inequality gives

$$\mathbb{E}[[C_T]_+] \le \frac{\log(d+1)}{\gamma} \le \frac{\log d + 1}{\gamma}. \tag{49}$$

Combining these estimates (47), (48), (49) with (46) and setting $\eta = \gamma = \sqrt{\log d/(dT)}$ gives

$$\mathbb{E}[[R_T^{\mathrm{rank}}]_+] \le \left(8 + \frac{2}{\log d} + 2\sqrt{\frac{2}{\log d}}\right) \sqrt{dT \log d}.$$

For $d \ge 2$, we have

$$8 + \frac{2}{\log d} + 2\sqrt{\frac{2}{\log d}} \le 8 + \frac{2}{\log 2} + 2\sqrt{\frac{2}{\log 2}} < 16,$$

so

$$\mathbb{E}[[R_T^{\mathrm{rank}}]_+] \le 16\sqrt{dT \log d}.$$

Finally, Markov's inequality and a union bound over the players prove the proposition. □

B.12 Proof of Proposition 5.7

Algorithm 2 Computing a compact vector for player i

Input: η_t
$\theta_i^0 \leftarrow \mathbf{0}_{|A_i|}$, $\tilde{\mu}_i^0 \leftarrow \mathbf{0}_{|A_i|}$
for $t = 1, \dots, T$ **do**
  Observe S_i^t and play $\pi_i^t(S_i^t)$
  $\tilde{\mu}_i^t \leftarrow \frac{t-1}{t}\tilde{\mu}_i^{t-1} + \frac{1}{t}\pi_i^t(S_i^t)$
  $G_i^t(a_i) \leftarrow \tilde{\mu}_i^t(a_i) - \frac{\mathbb{1}\{a_i \in S_i^t\}\exp(\theta_i^{t-1}(a_i))}{\sum_{a_i' \in S_i^t}\exp(\theta_i^{t-1}(a_i'))}, \forall a_i \in A_i$
  $\theta_i^t \leftarrow \theta_i^{t-1} + \eta_t G_i^t$
end for
return $w_i^T = \exp(\theta_i^T)/\sum_{a_i \in A_i}\exp(\theta_i^T(a_i))$

Proposition 5.7. *Let w_i^T be the weight vector produced by Algorithm 2. If $\frac{1}{T}\sum_{t=1}^T \pi_i^t(S_i^t) \to \mu_i^*$ almost surely as $T \to \infty$, and that $\sum_{t=1}^\infty \eta_t = \infty$ and $\sum_{t=1}^\infty \eta_t^2 < \infty$, then the marginal induced by w_i^T converges to μ_i^* almost surely. In particular, almost surely, for every $\delta > 0$, w_i^T is a δ-approximate compact representation of μ_i^* for all sufficiently large T.*

Proof. Suppress the player indices and consider

$$p_\theta(a \mid S) := \frac{\mathbb{1}\{a \in S\}e^{\theta(a)}}{\sum_{b \in S} e^{\theta(b)}}, \qquad \Phi(\theta) := \mathbb{E}_{S\sim\rho} \log \sum_{a \in S} e^{\theta(a)}.$$

Then $\nabla\Phi(\theta) = \hat{\mu}(\exp\theta)$. For each S, the function $\log\sum_{a\in S} e^{\theta(a)}$ has Hessian $\mathrm{diag}(p) - pp^\top$ for $p = p_\theta(\cdot \mid S)$, with eigenvalues in $[0, 1]$, so Φ is convex and its gradient is 1-Lipschitz. The bounded martingale differences $\pi^t(\cdot \mid S^t) - \mu^t$ satisfy

$$\Pr\left(\|\tilde{\mu}^t - \bar{\mu}^t\|_\infty > \varepsilon\right) \le 2|A|e^{-t\varepsilon^2/2}$$

by the Azuma–Hoeffding inequality (Azuma, 1967; Hoeffding, 1963) and a union bound over actions. Since these probabilities are summable for every $\varepsilon > 0$, $\tilde{\mu}^t - \bar{\mu}^t \to 0$ almost surely. Each $\bar{\mu}^t$ belongs to the closed set M, and $\mu^* \in M$ almost surely. Consider the case $\mu^* \in M$ and define

$$f(\theta) := \Phi(\theta) - \langle \mu^*, \theta\rangle, \qquad f_* := \inf_\theta f(\theta).$$

Since μ^* is implementable, $\langle\mu^*, \theta\rangle \le \mathbb{E}_{S\sim\rho}\max_{a\in S}\theta(a) \le \Phi(\theta)$, so $f \ge 0$, and $\|\nabla f\|_2 \le \sqrt{2}$ since $\nabla\Phi$ and μ^* are probability distributions. We claim that, for every $\delta > 0$, some $c_\delta > 0$ satisfies

$$f(\theta) \ge f_* + \delta \quad \Longrightarrow \quad \|\nabla f(\theta)\|_2 \ge c_\delta. \tag{50}$$

To see why this is the case, towards a contradiction, take θ_k with $f(\theta_k) \ge f_* + \delta$ and $\nabla f(\theta_k) \to 0$. The finitely many coordinates $p_{\theta_k}(a \mid S) \in [0, 1]$ admit a subsequence with $p_{\theta_k}(\cdot \mid S) \to q_S$ for every S. Since $\sum_S \rho(S)q_S = \mu^*$, for every θ,

$$\lim_k [f(\theta_k) - f(\theta)] = \sum_{S,a} \rho(S)q_S(a)\log\frac{p_\theta(a \mid S)}{q_S(a)} \le 0. \tag{51}$$

Here the summation is over $\rho(S) > 0$ and $a \in S$ with $q_S(a) > 0$. Then, the inequality in Equation (51) follows from $\log x \le x - 1$ and $\sum_{a:q_S(a)>0} p_\theta(a \mid S) \le 1$. Thus $f(\theta) \ge f_* + \delta$ for every θ, contradicting the definition of f_*.

Let $\mathcal{H}_t$ be the history and chosen strategy before S^t is drawn, and set

$$\xi_t := \nabla\Phi(\theta^{t-1}) - p_{\theta^{t-1}}(\cdot \mid S^t).$$

Since $\mathbb{E}[\xi_t \mid \mathcal{H}_t] = 0$, $\|\xi_t\|_2 \le \sqrt{2}$, and $\sum_t \eta_t^2 < \infty$, the partial sums of $\sum_t \eta_t \xi_t$ are martingales with bounded second moments and thus converge almost surely. Fix an outcome for which these partial sums converge and $\tilde{\mu}^t \to \mu^* \in M$. For the remainder of the proof, define

$$\vartheta^t := \theta^t + \sum_{k>t} \eta_k \xi_k.$$

The series tail tends to zero, so $\vartheta^t - \theta^t \to 0$. Subtracting $\eta_t \xi_t$ from the update in Algorithm 2 gives

$$\vartheta^t - \vartheta^{t-1} = \theta^t - \theta^{t-1} - \eta_t \xi_t = \eta_t[\tilde{\mu}^t - \nabla\Phi(\theta^{t-1})],$$

so $\|\vartheta^t - \vartheta^{t-1}\|_2 \le \sqrt{2}\eta_t$.

Set $\varepsilon_t := \|\tilde{\mu}^t - \mu^*\|_2 + \|\vartheta^{t-1} - \theta^{t-1}\|_2 \to 0$. The gradient bound and its Lipschitz property give

$$f(\vartheta^t) - f(\vartheta^{t-1}) \le -\eta_t \|\nabla f(\vartheta^{t-1})\|_2^2 + \sqrt{2}\eta_t \varepsilon_t + \eta_t^2, \tag{52}$$

$$|f(\vartheta^t) - f(\vartheta^{t-1})| \le 2\eta_t \to 0. \tag{53}$$

Fix some $\delta > 0$. For all sufficiently large t, $\sqrt{2}\varepsilon_t + \eta_t \le c_\delta^2/2$ and $2\eta_t \le \delta$. By (52) and (50), whenever $f(\vartheta^{t-1}) \ge f_* + \delta$, the subsequent value $f(\vartheta^t)$ decreases by at least $c_\delta^2 \eta_t/2$. Hence for sufficiently large t, the sequence $\{f(\vartheta^t)\}$ must fall below $f_* + \delta$: otherwise, summing these decreases contradicts $f \ge f_*$ because $\sum_t \eta_t = \infty$. It then stays below $f_* + 2\delta$: indeed, a step starting below $f_* + \delta$ increases the value by at most δ by (53) and a step starting at or above $f_* + \delta$ decreases it. Since $\delta > 0$ is arbitrary and $0 \le f(\vartheta^t) - f_* < 2\delta$, we conclude $f(\vartheta^t) \to f_*$.

Finally, the Lipschitz gradient gives

$$f_* \le f(\theta - \nabla f(\theta)) \le f(\theta) - \tfrac{1}{2}\|\nabla f(\theta)\|_2^2.$$

Hence $\nabla f(\vartheta^t) \to 0$. Since $\vartheta^t - \theta^t \to 0$, $\hat{\mu}(w^T) = \nabla\Phi(\theta^T) \to \mu^*$ almost surely, as required. □

B.13 Proof of Theorem 5.8

Algorithm 3 Finite-time RSA for player i

Input: $\tilde{\mu}_i \in \Delta(A_i)$
$\theta_i^1 \leftarrow \mathbf{0}_{|A_i|}$, $\lambda \leftarrow T^{-1/2}$, $s \leftarrow \lfloor T/2 \rfloor + 1$
for $t = 1, \dots, T$ **do**
  Draw $S_i^t \sim \rho_i$
  $\eta_t \leftarrow 2/(\lambda(t+1))$
  $G_i^t(a_i) \leftarrow \tilde{\mu}_i(a_i) - \frac{\mathbb{1}\{a_i \in S_i^t\}\exp(\theta_i^t(a_i))}{\sum_{a_i' \in S_i^t}\exp(\theta_i^t(a_i'))} - \lambda\theta_i^t(a_i), \forall a_i \in A_i$
  $\theta_i^{t+1} \leftarrow \theta_i^t + \eta_t G_i^t$
end for
$\tilde{\theta}_i \leftarrow \sum_{t=s}^T (\eta_t \theta_i^t)/\sum_{t=s}^T \eta_t$
return $\tilde{w}_i = \exp(\tilde{\theta}_i)/\sum_{a_i \in A_i}\exp(\tilde{\theta}_i(a_i))$

Nemirovski & Yudin (1978) initially proposed the use of Cesàro means to avoid non-convergence or slow convergence for Lipschitz convex functions in stochastic approximation, a method they referred to as *robust stochastic approximation* (RSA). Algorithm 3 adapts this framework to a quadratically regularized objective with decreasing stepsize; to accommodate this, we employ a trick known as *suffix averaging*, which was initially proposed in Rakhlin et al. (2012). We mostly follow the analysis of Nemirovski et al. (2009), using strong convexity to bound the error in minimizing the regularized objective after T timesteps. We then account for the quadratic regularizer and the difference between the empirical and true marginals to obtain the following bound.

Lemma B.5. *Fix a player i and an input $\tilde{\mu}_i \in \Delta(A_i)$ to Algorithm 3. For every implementable marginal $\mu_i \in M_i$, its output satisfies*

$$\mathbb{E}\big[\|\hat{\mu}_i(\tilde{w}_i) - \mu_i\|_2^2\big] \le \frac{192 + 4\log|A_i|}{\sqrt{T}} + 2\|\tilde{\mu}_i - \mu_i\|_2^2. \tag{54}$$

Proof. Suppress the player indices and use Φ from the preceding proof. The function

$$F_\lambda(\theta) := \Phi(\theta) - \langle \tilde{\mu}, \theta \rangle + \frac{\lambda}{2}\|\theta\|_2^2$$

is coercive and λ-strongly convex, so its unique minimizer θ^* satisfies

$$\nabla\Phi(\theta^*) - \tilde{\mu} + \lambda\theta^* = 0. \tag{55}$$

Since $0 \le \lambda\eta_t \le 1$, the update

$$\theta^{t+1} = (1 - \lambda\eta_t)\theta^t + \lambda\eta_t \frac{\tilde{\mu} - p_{\theta^t}(\cdot \mid S^t)}{\lambda}$$

and $\theta^1 = 0$ imply that $\lambda\|\theta^t\|_2 \le \sqrt{2}$ by induction. Thus $-G^t$ is an unbiased stochastic gradient of F_λ with norm at most $2\sqrt{2}$.

Let $D_t = \mathbb{E}\|\theta^t - \theta^*\|_2^2$. Expanding the squared distance norm and using strong convexity gives

$$2\eta_t \mathbb{E}[F_\lambda(\theta^t) - F_\lambda(\theta^*)] \le (1 - \lambda\eta_t)D_t - D_{t+1} + 8\eta_t^2. \tag{56}$$

Since the LHS of the above is nonnegative, using $\eta_t = 2/[\lambda(t+1)]$ yields $D_t \le 32/[\lambda^2(t+1)]$ by induction. The initial bound $D_1 \le 2/\lambda^2$ follows from (55).

For the averaging bound, note that $-\lambda\eta_t D_t \le 0$, so omitting this term only increases the right-hand side. Using the suffix averaging trick from Rakhlin et al. (2012), we sum (56) over $t = s, \dots, T$ where $s = \lfloor T/2 \rfloor + 1$, which gives, by convexity

$$\mathbb{E}[F_\lambda(\tilde{\theta}) - F_\lambda(\theta^*)] \le \frac{D_s + 8\sum_{t=s}^T \eta_t^2}{2\sum_{t=s}^T \eta_t} \le \frac{48}{\lambda T}. \tag{57}$$

The last step uses the inequalities $s + 1 \ge (T+2)/2$, $\sum_{t=s}^T \eta_t \ge T/[\lambda(T+1)]$, and $\sum_{t=s}^T \eta_t^2 \le 2\sum_{t=s}^T \eta_t/[\lambda(s+1)]$.

For any feasible μ, we have $\Phi(\theta) - \langle \mu, \theta \rangle \ge 0$ and $\Phi(0) \le \log|A_i|$. Using (55) and $F_\lambda(\theta^*) \le F_\lambda(0)$ therefore gives

$$\begin{aligned}
\|\nabla\Phi(\theta^*) - \mu\|_2^2 &= \|\tilde{\mu} - \mu\|_2^2 + \lambda^2\|\theta^*\|_2^2 - 2\lambda\langle \tilde{\mu} - \mu, \theta^* \rangle \\
&\le \|\tilde{\mu} - \mu\|_2^2 + 2\lambda[\Phi(0) - \Phi(\theta^*) + \langle \mu, \theta^* \rangle] \\
&\le \|\tilde{\mu} - \mu\|_2^2 + 2\lambda \log|A_i|.
\end{aligned}$$

Convexity and 1-smoothness of Φ, together with (55), imply

$$\begin{aligned}
\|\nabla\Phi(\theta) - \nabla\Phi(\theta^*)\|_2^2 &\le 2[\Phi(\theta) - \Phi(\theta^*) - \langle \nabla\Phi(\theta^*), \theta - \theta^* \rangle] \\
&\le 2[F_\lambda(\theta) - F_\lambda(\theta^*)].
\end{aligned}$$

Combining these bounds with (57) and the triangle inequality yields

$$\mathbb{E}\|\nabla\Phi(\tilde{\theta}) - \mu\|_2^2 \le \frac{192}{\lambda T} + 4\lambda \log|A_i| + 2\|\tilde{\mu} - \mu\|_2^2.$$

Substituting $\lambda = T^{-1/2}$ and $\hat{\mu}(\tilde{w}) = \nabla\Phi(\tilde{\theta})$ gives the statement as required. □

Now we are ready to prove Theorem 5.8.

Theorem 5.8. *Suppose CAT-FTPL is run for T rounds in a 2p0s-GSAS with payoffs $u_i : A \to [-1, 1]$, and the empirical marginals $(\tilde{\mu}_1, \tilde{\mu}_2)$ are used in Algorithm 3 to obtain compact vectors $\tilde{w}_i$. Then, for any $p \in (0, 1)$, with probability at least $1 - p$, the SPR of the strategies induced by $\tilde{w}_i$ satisfies* $\mathrm{SPR}^2 \le O\left(\frac{1}{p\sqrt{T}} + \frac{1}{p^2 T}\right)$*, where $O(\cdot)$ hides polynomial factors in $|A_i|$.*

Proof. Let $d = \max_i |A_i|$. The differences $\pi_i^t(\cdot \mid S_i^t) - \mu_i^t$ have conditional mean zero, so cross terms from different timesteps vanish in expectation. Since each $\pi_i^t(\cdot \mid S_i^t)$ is a probability vector, it holds that

$$\mathbb{E}\|\pi_i^t(\cdot \mid S_i^t) - \mu_i^t\|_2^2 = \mathbb{E}\|\pi_i^t(\cdot \mid S_i^t)\|_2^2 - \mathbb{E}\|\mu_i^t\|_2^2 \le 1.$$

Thus, expanding the squared norms gives

$$\mathbb{E}\sum_{i=1}^{2} \|\tilde{\mu}_i^T - \bar{\mu}_i^T\|_2^2 = \frac{1}{T^2}\sum_{i=1}^{2}\sum_{t=1}^{T} \mathbb{E}\|\pi_i^t(\cdot \mid S_i^t) - \mu_i^t\|_2^2 \le \frac{2}{T}.$$

We can apply Lemma B.5 conditionally on the CAT-FTPL history with inputs $\tilde{\mu}_i^T$ and $\bar{\mu}_i^T$. Taking expectations yields

$$\mathbb{E}\sum_{i=1}^{2} \|\hat{\mu}_i(\tilde{w}_i) - \bar{\mu}_i^T\|_2^2 \le O\left(\frac{1+\log d}{\sqrt{T}}\right).$$

Then, Markov's inequality, with failure probability $p/2$, gives

$$\sum_{i=1}^{2} \|\hat{\mu}_i(\tilde{w}_i) - \bar{\mu}_i^T\|_2^2 \le O\left(\frac{1+\log d}{p\sqrt{T}}\right). \tag{58}$$

Meanwhile, setting $p_1 = p_2 = p/4$ in Proposition 5.6, and combining with Proposition 5.3 give

$$\mathrm{SPR}(\bar{\mu}_1^T, \bar{\mu}_2^T) \le O\left(\frac{\sqrt{d\log d}}{p\sqrt{T}}\right)$$

with failure probability at most $p/2$. Note that both bounds hold with probability at least $1-p$ by a union bound.

Since payoffs lie in $[-1,1]$, for every $\nu_1 \in M_1$ and $\nu_2 \in M_2$,

$$\begin{aligned} U\big(\nu_1, \hat{\mu}_2(\tilde{w}_2)\big) &\le U(\nu_1, \bar{\mu}_2^T) + \|\hat{\mu}_2(\tilde{w}_2) - \bar{\mu}_2^T\|_1, \\ U\big(\hat{\mu}_1(\tilde{w}_1), \nu_2\big) &\ge U(\bar{\mu}_1^T, \nu_2) - \|\hat{\mu}_1(\tilde{w}_1) - \bar{\mu}_1^T\|_1. \end{aligned}$$

Taking the maximum over ν_1 in the first inequality and the minimum over ν_2 in the second then subtracting gives, by definition of SPR,

$$\mathrm{SPR}\big(\hat{\mu}_1(\tilde{w}_1), \hat{\mu}_2(\tilde{w}_2)\big) \le \mathrm{SPR}(\bar{\mu}_1^T, \bar{\mu}_2^T) + \|\hat{\mu}_1(\tilde{w}_1) - \bar{\mu}_1^T\|_1 + \|\hat{\mu}_2(\tilde{w}_2) - \bar{\mu}_2^T\|_1. \tag{59}$$

Recall that for every a, b, c we have $(a+b+c)^2 \le 2a^2 + 2(b+c)^2 \le 2a^2 + 4b^2 + 4c^2$. Moreover, applying Cauchy–Schwarz we get that $\|\hat{\mu}_i(\tilde{w}_i) - \bar{\mu}_i^T\|_1 \le \sqrt{d}\|\hat{\mu}_i(\tilde{w}_i) - \bar{\mu}_i^T\|_2$. Hence, it follows that

$$\begin{aligned} \mathrm{SPR}\big(\hat{\mu}_1(\tilde{w}_1), \hat{\mu}_2(\tilde{w}_2)\big)^2 &\le 2\,\mathrm{SPR}(\bar{\mu}_1^T, \bar{\mu}_2^T)^2 + 4d\sum_{i=1}^{2} \|\hat{\mu}_i(\tilde{w}_i) - \bar{\mu}_i^T\|_2^2 \\ &\le O\left(\frac{d(1+\log d)}{p\sqrt{T}} + \frac{d\log d}{p^2 T}\right). \end{aligned}$$

This proves that the stated bound holds with probability at least $1-p$. □

C Additional Experiments

C.1 Game definitions

In this section we formally define the games used in our experiments.

Definition C.1 (Random 2p0s-GSAS)**.** *A random GSAS has $|A_i| = n$ actions for each player, with payoffs given by*

$$A_{i,j} \sim \text{Uniform}[-1, 1],$$

and for each action we sample $p_{a_i} \sim \text{Uniform}[3/10, 5/10]$ when generating the game. Sampling $S \sim \rho$ then, independently for each player, selects actions a_i to be available independently with probability p_{a_i} and redraws if no action is present.

Definition C.2 ($n \times n$ Random Biased Support ("RBS") GSAS)**.** *An $n \times n$ RBS GSAS has $|A_i| = n$ actions for each player, with payoffs given by*

$$A_{i,j} \sim \text{Uniform}[-1, 1],$$

except that the upper-left block has a matching pennies structure given by $A_{1,1} = A_{2,2} = 1$ and $A_{1,2} = A_{2,1} = 0$ and that the first two actions which are always available. For the other actions we sample $p_{a_i} \sim \text{Uniform}[1/100, 2/100]$. Sampling $S \sim \rho$ then, independently for each player, selects actions a_i to be available independently with probability p_{a_i} and redraws if no action is present.

Definition C.3 (Checkerboard $n \times n$ Matching Pennies ("Checkerboard MP"))**.** *A checkerboard $n \times n$ MP has $|A_i| = n$ actions for each player, with payoffs given by*

$$A_{i,j} \coloneqq \begin{cases} 1 & \text{if } i \equiv j \pmod 2, \\ -1 & \text{otherwise}. \end{cases}$$

Action availabilities for player 1 are drawn uniformly at random from

$$\{\{i\} \cup \{2, 4, \ldots, n\} \mid i \in \{1, 3, \ldots, n-1\}\},$$

and player 2 always has access to all actions. Here n is required to be even.

Intuitively, player 1 wins if they pick an action with the same parity as player 2, and player 1 has access to all even actions and a single odd action chosen uniformly at random, while player 2 has access to all actions.

Definition C.4 (Biased $n \times n$ Rock-Paper-Scissors ("Biased RPS"))**.** *A biased $n \times n$ RPS has $|A_i| = n \geq 3$ actions for each player, with payoffs given by*

$$A_{i,j} \coloneqq \begin{cases} -1 & \text{if } j \equiv i + 1 \pmod n, \\ 1 & \text{if } j \equiv i - 1 \pmod n, \\ 0 & \text{otherwise}. \end{cases}$$

With probability $2/n$, player 1 only has their first action available; otherwise player 1 has all their actions available. Player 2 always has all actions available.

This is a special case of the generalized $n \times n$ Rock-Paper-Scissors defined by Lazarsfeld et al. (2025).

Definition C.5 (Biased $n \times n$ matching pennies ("Biased MP"))**.** *A biased $n \times n$ MP has $|A_i| = n \geq 3$ actions for each player, with payoffs given by*

$$A_{i,j} \coloneqq \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise}. \end{cases}$$

Player 1 has with probability $4/5$ only actions $\{a_1, \ldots, a_{\lfloor 3n/5 \rfloor}\}$ available, and with probability $1/5$ only actions $\{a_{\lfloor 2n/5 \rfloor}, \ldots, a_n\}$ available. Player 2 always has all actions available.

C.2 Additional Experimental Details

All experiments were run on a 2021 MacBook Pro with 32 GB of RAM and an 8-core Apple M1 Pro chip. The Gurobi commercial solver was allowed to use any number of threads. Gurobi optimizer version 13.0.1 build v13.0.1rc0 (mac64[arm] - Darwin 24.6.0 24G90) was used. Python version 3.14.2 with NumPy version 2.5.3 and SciPy version 1.18.1 were used. Central 95% intervals, where reported, were computed using `numpy.quantile`. To ensure reproducibility, all randomness used in the algorithms was generated by a NumPy random generator seeded with the current run number, i.e., if an experiment performed 100 runs, the rth run seeded the random number generator with r.

C.2.1 Experiment 1: Comparison with LP solver

To provide a suitable comparison between CAT-FTPL and Gurobi, we first construct the sequence-form representation (von Stengel, 1996) of the GSAS, which encodes all possible action subsets in the game, and apply Gurobi's LP solver to this expanded form, recording the wallclock convergence time. For CAT-FTPL, we use the doubling trick to find a time horizon $T \geq 1000$ with $\mathrm{SPR}(\tilde{\mu_1}, \tilde{\mu_2}) \leq 0.05$ and report the wallclock time of Algorithm 1 for the final run. We repeat the experiment 20 times and show the average and range across all runs in Figure 1. As described in Section 6, with a 360-second time budget, Gurobi could solve a GSAS of size 11×11, via an LP with 13313 variables and a linear constraint matrix of size 13313×13313 with 126904320 nonzero entries, taking ≈ 293 seconds to solve on average. CAT-FTPL was able to obtain low SPR for much larger games (350×350) within the same time budget.

Nevertheless, we reiterate that comparing the *solution quality* of these methods is moot: our method is an online regret-minimizing algorithm, and so the stopping criterion can be thought of as a proxy for terminating when the time-averaged ranking regret is low enough. This is not comparable at all to the interior point methods used by Gurobi's LP solver. The experiment is specifically designed to investigate the runtime difficulties that arise due to exponentially many action availability sets in GSAS, which is corroborated in Figure 1. For a more direct comparison with some implications on solution quality, it is instructive for the reader to consider the experiment on multi-scale MWU in Appendix C.4. Since multi-scale MWU is an *online* algorithm that is designed for Bayesian games, we can to some extent compare the quality of the solutions obtained.

Some additional remarks are in order. In Figure 1, we plot the *wallclock* runtime reported by Gurobi, which does *not* include the time to construct the sequence-form representation or the time to build the Gurobi model, which in practice can also be slow for large games. In other words, CAT-FTPL significantly outperforms Gurobi for large games, even without considering the additional preprocessing required for Gurobi. In line with our theoretical results, we run CAT-FTPL with $\eta_i = \gamma_i = \sqrt{\frac{\log|A_i|}{|A_i|T}}$.

The sequence-form linear program has $2^n + n2^{n-1} + 1$ variables, and a constraint matrix of the same dimension, which quickly becomes infeasible to solve as observed.

C.2.2 Experiment 2: Convergence in large GSAS

We find that using smaller λ values in Algorithm 3 induces faster convergence in practice, and so run Algorithm 3 with $\lambda = \frac{0.1}{\sqrt{T}}$. We run CAT-FTPL with the parameters from Theorem 5.5, i.e., using $\eta_i = \gamma_i = \sqrt{\frac{\log|A_i|}{|A_i|T}}$. All runs of Algorithm 3 for a given game use the same time-average marginal $(\tilde{\mu}_1, \tilde{\mu}_2)$, computed as the output of a CAT-FTPL run with time horizon $T = 120000$, to isolate the error contributed to the SPR to only that from the compact representation procedure Algorithm 3.

In Figure 4, we show further regret plots on several additional 2p0s-GSAS: Example 4.6, biased 50×50 Rock-Paper-Scissors (Definition C.4) and biased 50×50 matching pennies (Definition C.5). We show the saddle point residual on these games for the CAT-FTPL marginals in Figure 5 and of the robust averaging w_i^t in Figure 6. All results match our theoretical bounds as in Section 6.

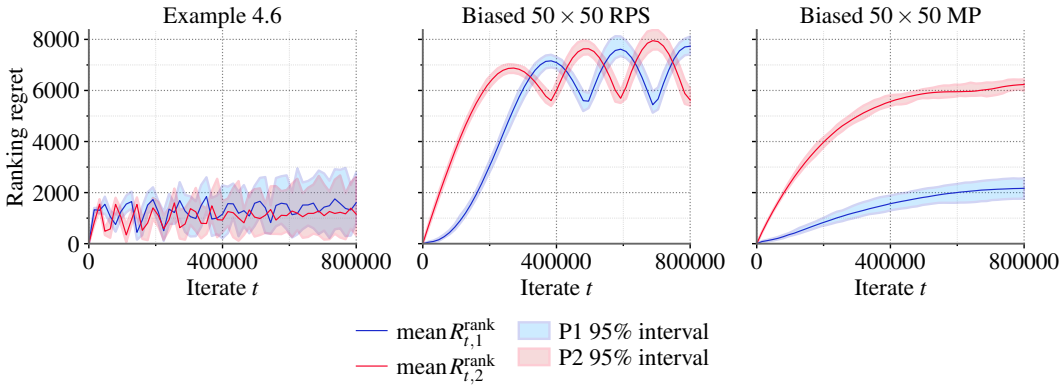
C.2.3 Estimating the saddle point residual in experiments

Since we focus on the 2p0s case, we can consider the *saddle point residual* of a strategy pair (π_1, π_2). In particular, let $U = U_1 = -U_2$. The SPR expressed in terms of strategies (rather than implementable marginals) is given as

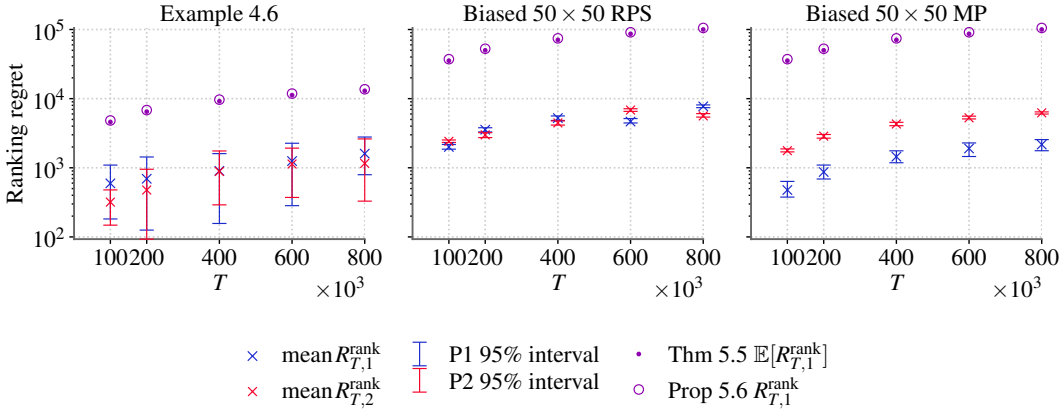
Definition C.6 (Saddle point residual (SPR) of a strategy pair)**.**

$$\begin{aligned}\mathrm{SPR}(\pi_1, \pi_2) &= [U(\pi_1, \pi_2) - \min_{\pi_2'} U(\pi_1, \pi_2')] + [\max_{\pi_1'} U(\pi_1', \pi_2) - U(\pi_1, \pi_2)] \\ &= \max_{\pi_1'} U(\pi_1', \pi_2) - \min_{\pi_2'} U(\pi_1, \pi_2').\end{aligned}$$

It is easy to see that the saddle point residual of (π_1, π_2) is equal to the saddle point residual of their marginals (cf. Definition 5.2) and is zero if and only if (π_1, π_2) is a Nash equilibrium.



(a) $R^{\text{rank}}_{t,i}$ for $T = 800000$. Mean and 95% central interval from 20 runs shown. Note the theoretical bounds from Theorem 5.5 and Proposition 5.6 apply only to the final iterate at $T = 800000$.



(b) Ranking regret of final iterate of CAT-FTPL for various T. Theoretical bounds from from Theorem 5.5 and Proposition 5.6 are also shown. Results from 20 runs for each T shown.

Figure 4: Ranking regret of CAT-FTPL for three 2p0s-GSAS. Average, central 95% interval and theoretical expected/high-probability bounds on maximum ranking regret are shown. We observed $R^{\text{rank}}_{t,i}$ to be much lower in practice than the bound in Theorem 5.5.

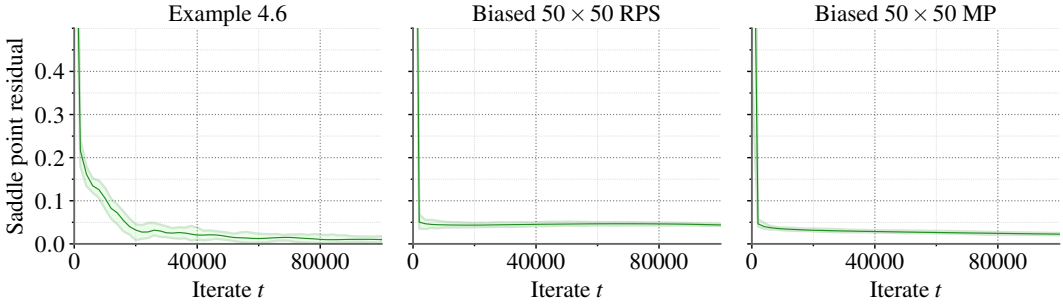


Figure 5: Saddle point residuals of the time-averaged empirical marginal pairs produced by CAT-FTPL for three 2p0s-GSAS. We repeated each experiment 100 times and plot the average and the 95% central interval across the repetitions.

We seek to compute the saddle point residual of the strategies π_i both induced by the compact vectors from Algorithm 2 and by the marginal distributions produced by CAT-FTPL. However, $U(\pi_1, \pi_2) = \mathbb{E}_{S\sim\rho}\left[\mathbb{E}_{a\sim\pi(S)}[u_1(a)]\right]$ is infeasible to calculate when we only have sampling access to ρ, or when its support is too large to enumerate. We thus wish to calculate where possible, and estimate when not, the saddle point residual of the empirical marginals in three different regimes:

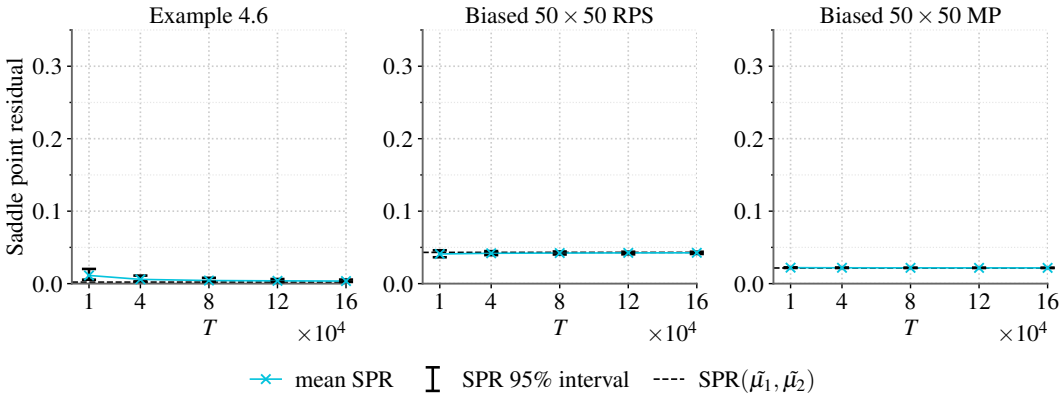


Figure 6: Saddle point residuals of the strategy pairs induced by the weight vectors w_i^T learned by Algorithm 3 on the marginal produced by CAT-FTPL for three 2p0s-GSAS. We repeated each experiment 100 times and plot the mean and the 95% central interval of the saddle point residual across the repetitions.

(i) ρ is known and has small support, (ii) only sampling access to ρ is available, and (iii) when ρ is such that each action is available independently across the action set. For each regime, we describe the calculation only for $\max_{\pi_1'} U(\pi_1', \pi_2)$, noting the procedure for $\min_{\pi_2'} U(\pi_1, \pi_2')$ is similar. Although the empirical marginals may not be implementable, the chance that an empirical marginal is significantly different from an implementable marginal is negligible for the large number of timesteps used in our experiments.

(i) Known ρ with small support regime. If we know $\rho = (\rho_1, \rho_2)$ we can directly calculate $\max_{\pi_1'} U(\pi_1', \pi_2)$. For a marginal distribution μ_2 induced by π_2, we compute

$$\max_{\pi_1'} U(\pi_1', \pi_2) = \sum_{S_1 \in \text{supp}(\rho_1)} \mathbb{P}[S_1] \max_{a_1 \in S_1} \left[\sum_{a_2 \in A_2} \mu_2(a_2) U(a_1, a_2) \right].$$

If we instead have a strategy π_2, we can first calculate the expected payoff for each action $a_1 \in A_1$ as

$$\mathbb{E}_{S_2 \sim \rho_2}[U(a_1; \pi_2(S_2))] = \sum_{S_2 \in \text{supp}(\rho_2)} \rho_2(S_2) U(a_1; \pi_2(S_2)),$$

and can then proceed the same as for the marginal. We use this procedure for calculating the SPR for Example 4.6 and for small ($n \leq 10$) instances of "Checkerboard MP", "Biased RPS", and "Biased MP".

(ii) Sample ρ access regime. If we only have sample access to ρ, for example when enumerating all possibly observed $S_i \in \text{supp}(\rho_i)$ is infeasible, we repeat the same process as in the known ρ regime but sampling from ρ as needed, instead of enumerating all possible action availabilities, to estimate the SPR.

We use this regime to estimate the SPR for large ($n > 10$) instances of "Checkerboard MP", "Biased RPS", and "Biased MP".

(iii) Actions available independently regime. Consider a ρ_i such that action a_i's availability is determined by independently sampling from a Bernoulli distribution with parameter p_{a_i}, and resampling all actions if no action is available. If we know the Bernoulli parameters p_{a_i}, after estimating the expected payoff against the opponent $E(a_i)$ for each action $a_i \in A_i$ the same as in the "sample ρ access regime", we can then directly calculate $\mathbb{E}_{S_i \sim \rho_i}[\max_{a_i \in S_i} E(a_i)]$.

We proceed by calculating the best response for player 1, with the procedure identical for player 2. Assume without loss of generality that actions are indexed such that $E(a_j) \geq E(a_{j+1})$ and let $V = \max_{\pi_1'} U(\pi_1', \pi_2) = \mathbb{E}_{S_1 \sim \rho_1}[\max_{a \in S_1} E(a)]$.

Ignoring the case that no action is present and S_1 must be resampled, action a_i contributes $E(a_i)p_{a_i}\prod_{j=1}^{i-1}(1-p_{a_j})$ to V.

We account for the case where we need to resample by solving

$$V = \sum_{i=1}^{|A_1|}\left[E(a_i)p_{a_i}\prod_{j=1}^{i-1}(1-p_{a_j})\right] + V\prod_{i=1}^{|A_1|}(1-p_{a_i})$$
$$= \frac{1}{1-\prod_{i=1}^{|A_1|}(1-p_{a_i})}\sum_{i=1}^{|A_1|}\left[E(a_i)p_{a_i}\prod_{j=1}^{i-1}(1-p_{a_j})\right],$$

which can be calculated in a single pass over the actions after sorting.

We use this to more accurately estimate the saddle point residual of a random 2p0s-GSAS (Definition C.1) and RBS GSAS (Definition C.2).

C.2.4 Calculating the ranking regret in experiments

Naively calculating the ranking regret directly by the definition (cf. Definition 5.1), i.e., by tracking the realized payoff for every ranking strategy $\pi_i^{\tau_i}$, would require tracking $|A_i|!$ many values (the payoff from each ranking) which is infeasible for large GSAS. Thus, to compute $R_{T,i}^{\text{rank}}$ in practice, we:

1. Calculate the expected payoff from each action given the opponents' marginals using Appendix C.2.3.
2. Calculate the optimal ranking $\pi_i^{\tau_i}$ in hindsight.
3. Estimate $\sum_{t=1}^{T} U_i(\pi_i^{\tau_i}, \pi_{-i}^t)$ using the same three calculation regimes as estimating the SPR (the "known ρ with small support" regime, the "sample ρ access" regime or the "actions available independently" regime).
4. Subtract $\sum_{t=1}^{T} U_i(\pi^t)$.

Note that similar to estimating the saddle point residual, the realized payoff may differ from the expected payoff of the strategy, but this difference is negligible for the large values of T used in our experiments.

C.3 Solving Example 4.6 Computationally

Definition C.7. *The original payoff matrix $\mathcal{G}_{\text{orig}}$ of the modified RPS game from Example 4.6 is given by*

	***R**ock*	***P**aper*	***S**cissors*
***R**ock*	0	-1	1
***P**aper*	0.5	0	-0.5
***S**cissors*	-1	1	0

We apply the procedure outlined in Section 5 to solve the GSAS as described in Example 4.6. In particular, we run CAT-FTPL and run Algorithm 3. We show the learned weights per iterate of Algorithm 3 in Figure 7. We find that the weights learned by Algorithm 3 converge to the analytically computed compact Nash equilibrium vector, which is $(\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$ for player 1 and $(\frac{2}{5}, \frac{2}{5}, \frac{1}{5})$ for player 2.

C.4 Comparison with Multi-scale MWU for Bayesian Games

In Figure 8, we show the wallclock running time for multi-scale MWU (Peng & Rubinstein, 2024) as applied to random 2p0s-GSAS (as defined in Definition C.1 and used in Experiment 1) with $\epsilon = 0.3$. Note that the choice of $\epsilon = 0.3$ for multi-scale MWU was a practical one: the theory guarantees

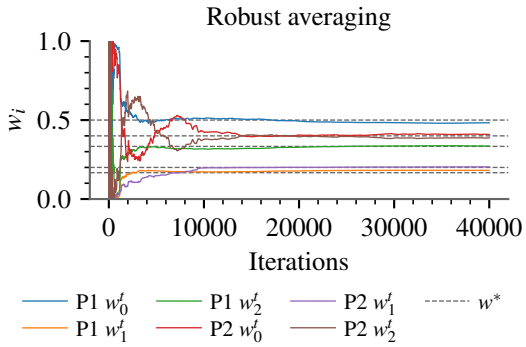


Figure 7: Learned weights by Algorithm 3 using robust averaging on CAT-FTPL output in Example 4.6.

convergence to an ϵ-normal form correlated equilibria of the expanded game, but we empirically observe that the runtime of multi-scale MWU was prohibitively slow for smaller ϵ. As the expanded game is two-player zero-sum, normal form correlated equilibria are equivalent to Nash equilibria.

Recall that multi-scale MWU was developed for Bayesian games (Peng & Rubinstein, 2024; Dagan et al., 2024), and one can reduce a GSAS of size n to a Bayesian game with $2^n - 1$ types per player (one type for each nonempty availability set). In Figure 8, we observe the runtime grows exponentially with game size for multi-scale MWU, similar to the LP solver method in Figure 1. This observation matches the bounds given in (Peng & Rubinstein, 2024): although the bound on the iterations needed is independent of the number of types in the Bayesian game, the bound on the total runtime grows in the number of types.

CAT-FTPL in comparison exhibits runtime polynomial in the game size, matching the theoretical bounds. This difference is expected, as CAT-FTPL can exploit the additional structure in GSAS between availability sets, which is inaccessible to a Bayesian game solver where there may not be structure across types.

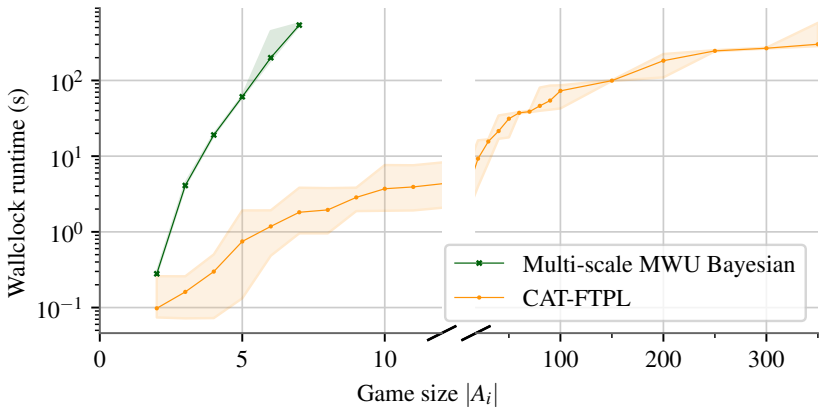


Figure 8: Wallclock runtime for multi-scale MWU on random GSAS (Definition C.1). In a time budget of 360 s the algorithm can solve a 6×6 game, with 609 s on average needed to solve a 7×7 random GSAS. Comparative results from Figure 1 for our method are shown, scaling to 350×350 random GSAS within the 360 s time budget.

D On the Bit Complexity of GSAS Solutions

Before proceeding, we make two important points. First, we are *not* interested in bit representations that are as large as the number of entries of ρ (which could be exponential in $|A_i|$) — if we allowed that, then one may as well do away with the compact representation altogether and work with the naïve distribution mapping action sets to distributions over actions, an approach that is neither practical nor enlightening. Second, this section is only concerned with *representing* (approximate) equilibria, and has little to do with explicit algorithms to compute the representations.

D.1 Bit-Compact Representations

Fix some family Γ of 2p0s-GSAS, for example, those meeting Assumption 3.3. We want to be able to represent an ϵ-NE strategy for any GSAS from Γ. Informally, this representation should allow us to play an ϵ-NE when given an availability set, without needing to perform further game-specific calculations.[2] As a concrete example, for $\Gamma = \{\mathcal{G} \mid \mathcal{G}$ a 2p0s-GSAS meeting Assumption 3.3$\}$, we know from Theorem 4.5 that for every $\mathcal{G} \in \Gamma$ there exists a vector $w_i \in \Delta(A_i)$ and the following algorithm: given w_i and an availability set S_i, renormalize w_i over the available actions, will play an ϵ-NE for $\mathcal{G}$. Note that to strictly meet the definition of bit-compact representation would require showing we can approximate the numbers in w_i with polynomially many bits.

While we do not want to unnecessarily restrict how such bit-compact ϵ-NE representations work, we do require that they admit a *playing algorithm* which is independent of any particular GSAS in Γ. For any GSAS in Γ, given the right bit string and an availability set, the *playing algorithm* should play a valid ϵ-NE. To this end, Definition D.1 formally captures 'bit-compactness' of ϵ-NE in 2p0s-GSAS:

Definition D.1 (Bit-compact representable ϵ-NE in 2p0s-GSAS)**.** *Let $\Gamma \subseteq \{\mathcal{G} \mid \mathcal{G}$ is a 2p0s-GSAS$\}$ be a family of GSAS, with $\mathcal{G} \in \Gamma$ a particular GSAS from the family. We denote $\mathcal{G}$'s actions for player i by $A_{\mathcal{G},i}$ and availability sets by $\mathcal{S}_{\mathcal{G},i}$. We say that Γ has* bit-compactly representable ϵ-NE *if there exists a polynomial p and an algorithm R, taking as input a bit string $x_{\mathcal{G},i}$ and an availability set $\mathcal{S}_{\mathcal{G},i}$, such that $\forall \epsilon \in (0,1), \forall \mathcal{G} \in \Gamma, \forall i \in \{1,2\}, \exists x_{\mathcal{G},i} \in \{0,1\}^{p\left(|A_{\mathcal{G},i}|, \log\left(\frac{1}{\epsilon}\right)\right)}$, $\forall S_i \in \mathcal{S}_{\mathcal{G},i}$, $R(S_i, x_{\mathcal{G},i}) \in \Delta(S_i)$ and the strategy profile played by R is an ϵ-NE. We call such an $x_{\mathcal{G},i}$ for a GSAS $\mathcal{G}$ a* bit-compact representation *and let $k := |x_{\mathcal{G},i}|$ be the length of the bit string.*

Crucially, R is fixed for Γ and only takes in $x_{\mathcal{G},i}$ and S_i, but not the entire GSAS $\mathcal{G}$. This means that it can only operate on the compact representation afforded by $x_{\mathcal{G},i}$, and cannot rely on external quantities like ρ unless they are stored either implicitly or explicitly in $x_{\mathcal{G},i}$. Furthermore, note that k is required to be polynomial in $|A_i|$ and $\log\left(\frac{1}{\epsilon}\right)$, while in general ρ could require $\Omega(2^{|A_i|})$ bits to encode. The omission of $\mathcal{G}$ was explicitly chosen to ensure any bit-compact representation contains all the information needed to play an ϵ-NE, but without specifying how that information is represented.

Note that having a bit-compact representation of a GSAS family is necessary, but not sufficient, to have a practical method to solve that family: one additionally needs a practical method to compute the bit string, and for the playing algorithm to run in polynomial time in its input length. However, our procedure in Section 5 meets these additional requirements, with CAT-FTPL and Algorithm 3 providing a practical method to compute w and the playing algorithm used by Theorem 4.5 having a polynomial runtime.

Finally, while Definition D.1 is specified for 2p0s-GSAS, it can be broadened to general GSAS by ensuring that the joint strategy played by R is an ϵ-NE. The definition can also be easily generalized to cover solution concepts such as (coarse) correlated equilibria, or indeed any set of strategies meeting some desired criteria.

D.2 Polynomial Bit-Compactness under Assumption 3.3

In light of the above, a nuanced but important question is: while the NE of a 2p0s-GSAS meeting Assumption 3.3 can be approximately represented for player i with $O(|A_i|)$ real numbers given

[2] For example, we cannot allow the naïve strategy of solving the GSAS from scratch when presented with an availability set.

by w_i^*, can the entries of w_i^* be written in a polynomial number of *bits* that are used to represent the original payoff matrix $\mathcal{G}_{\text{orig}}$ (whose elements are represented by rational numbers)? If this were indeed the case, then such a w_i^* can be used as a bitstring for $x_{\mathcal{G},i}$, and that would immediately be a bit-compact representation in the sense of Definition D.1.

Unfortunately, the answer to the above is negative. Indeed, the proof of Theorem 4.5 relies on w_i^* being the solution to a system of quadratics (cf. Appendix B.6). It is well-known that solutions to quadratic systems can be algebraic, even if the entries of the payoff matrices are rational. Concretely, we give a counterexample showing that there exists a 2p0s-GSAS $\mathcal{G}$ for which a compactly represented equilibrium w^* (in the sense of Theorem 4.5) contains irrational numbers as its entries, even under Assumption 3.3, assuming payoffs in $\mathcal{G}_{\text{orig}}$ and probabilities in ρ are all rational.

Consider the following 2p0s-GSAS (with independent action availabilities, fulfilling Assumption 3.3) where only player 1 has stochastic action sets. We have $|A_1| = |A_2| = 3$ and

$$\mathcal{G}_{\text{orig}} = \begin{array}{c|ccc} A_1 \backslash A_2 & 1 & 2 & 3 \\ \hline 1 & 1/2 & 0 & 0 \\ 2 & 0 & 1/2 & 0 \\ 3 & 0 & 0 & 1/3 \end{array}$$

where $\mathcal{S}$ and ρ are given by

$$\begin{array}{c|c} \mathcal{S}_1 & \rho_1 \\ \hline \{1\} & \frac{1}{7} \\ \{2\} & \frac{1}{7} \\ \{3\} & \frac{1}{7} \\ \{1,2\} & \frac{1}{7} \\ \{1,3\} & \frac{1}{7} \\ \{2,3\} & \frac{1}{7} \\ \{1,2,3\} & \frac{1}{7} \end{array}, \quad \begin{array}{c|c} \mathcal{S}_2 & \rho_2 \\ \hline \{1,2,3\} & 1 \end{array},$$

i.e., player 1's availability set is sampled uniformly from the nonempty subsets of A_1, while player 2 has every action always available. It is easy to verify that the unique equilibrium of $\mathcal{G}_{\text{orig}}$ is given by both players playing actions 1, 2, and 3 with probabilities $[2/7, 2/7, 3/7]$ respectively. We know from Proposition 4.4 that if this distribution is implementable by both players, then it is also an equilibrium in $\mathcal{G}$ in a sense of marginal distributions as described by (μ_1^*, μ_2^*). Indeed, this can be achieved in several ways. For example, the following gives the joint probabilities for an equilibrium strategy of player 1:

$$\begin{array}{c|ccc|c} \mathcal{S}_1 & 1 & 2 & 3 & \rho_1 \\ \hline \{1\} & \frac{1}{7} & 0 & 0 & \frac{1}{7} \\ \{2\} & 0 & \frac{1}{7} & 0 & \frac{1}{7} \\ \{3\} & 0 & 0 & \frac{1}{7} & \frac{1}{7} \\ \{1,2\} & \frac{1}{14} & \frac{1}{14} & 0 & \frac{1}{7} \\ \{1,3\} & \frac{1}{28} & 0 & \frac{3}{28} & \frac{1}{7} \\ \{2,3\} & 0 & \frac{1}{28} & \frac{3}{28} & \frac{1}{7} \\ \{1,2,3\} & \frac{1}{28} & \frac{1}{28} & \frac{1}{14} & \frac{1}{7} \\ \hline \mu_1^* & \frac{2}{7} & \frac{2}{7} & \frac{3}{7} & 1 \end{array},$$

which unfortunately does not obey our compact representation. However, the following implements the exact same marginal probabilities μ_1^*:

$\mathcal{S}_1$	1	2	3	ρ_1
$\{1\}$	$\frac{1}{7}$	0	0	$\frac{1}{7}$
$\{2\}$	0	$\frac{1}{7}$	0	$\frac{1}{7}$
$\{3\}$	0	0	$\frac{1}{7}$	$\frac{1}{7}$
$\{1,2\}$	$\frac{1}{14}$	$\frac{1}{14}$	0	$\frac{1}{7}$
$\{1,3\}$	$\frac{s-3}{28}$	0	$\frac{7-s}{28}$	$\frac{1}{7}$
$\{2,3\}$	0	$\frac{s-3}{28}$	$\frac{7-s}{28}$	$\frac{1}{7}$
$\{1,2,3\}$	$\frac{5-s}{28}$	$\frac{5-s}{28}$	$\frac{s-3}{14}$	$\frac{1}{7}$
μ_1^*	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{3}{7}$	1

,

where $s = \sqrt{17}$ does indeed obey our compact representation given by

$$w_1^* = \left[\frac{5-s}{4}, \frac{5-s}{4}, \frac{s-3}{2}\right],$$

which is clearly irrational.

The above counterexample may look discouraging, as it shows that the normalized proportional weights need not be rational, so the decimal/binary representation cannot be written exactly with finite bits. However, we show that there exist ϵ-NE compactly represented by $\tilde{w}^*$ that do enjoy polynomial bit-complexity. This can be achieved easily by a simple dyadic truncation procedure.

For a positive weight vector w, we require the following notation:

$$w(S) = \sum_{a \in S} w_a, \qquad \hat{\mu}(w)_a = \sum_{S \ni a} \rho(S) \frac{w_a}{w(S)}. \tag{60}$$

In other words, $\hat{\mu}(w)$ is the marginal induced by playing the GSAS using w. Then, we have the following result:

Theorem D.2. *For every implementable marginal $\mu_i \in M_i$ and each $\delta \in (0,1)$, there exists a δ-approximate compact representation w_i of μ_i in the sense of Theorem 4.5 and a rational vector $\tilde{w}_i \in \Delta(A_i)$ that can be represented in $O(|A_i| \log(|A_i|/\delta))$ bits such that $\|\tilde{w}_i - w_i\|_1 \leq \delta$ and $\|\hat{\mu}_i(\tilde{w}_i) - \mu_i\|_1 \leq \delta$.*

Proof. In the remainder of the proof, we suppress the player index for brevity. We will require a technical lemma that establishes an appropriate choice of w so that the number of bits required to represent $\tilde{w}$ is $O(|A| \log(|A|/\delta))$.

Lemma D.3. *Let $\mu \in M$. For every $\eta \in (0,1)$, there exists $w \in \Delta(A) \cap \mathbb{R}_{>0}^{|A|}$ such that*

$$\|\hat{\mu}(w) - \mu\|_1 \leq 2\eta, \quad \textit{and} \quad \min_a w_a \geq \frac{1}{|A|}\left(\frac{\eta}{|A|}\right)^{|A|-1}. \tag{61}$$

Proof. Consider a π that implements μ. For $a \in S$, and similarly to the proof of Theorem 4.5 we define the conditional distribution:

$$\pi^\eta(a|S) = (1-\eta)\pi(a|S) + \frac{\eta}{|S|}, \tag{62}$$

which ensures full support for π^η for all S. Let μ^η be the induced marginal of π^η. Taking the expectation over action availabilities with respect to ρ gives:

$$\|\mu^\eta - \mu\|_1 \leq \sum_S \rho(S) \|\pi^\eta(\cdot|S) - \pi(\cdot|S)\|_1 \leq 2\eta, \tag{63}$$

where the first inequality holds due to the triangle inequality and the second holds because the ℓ_1-norm between two probability distributions is at most 2. Then, the same matrix-scaling argument in the proof of Theorem 4.5 ensures that there exist positive weights v such that $\hat{\mu}(v) = \mu^\eta$.

Next, we seek to modify the weights v without changing the induced marginal, so as to ensure a lower bound on the smallest element of w. First, let us relabel the actions so that $v_1 \geq \cdots \geq v_{|A|} > 0$. Moreover, for each $j < |A|$, consider the sets $H_j = \{1, \ldots, j\}$ and the ratios $r_j = \frac{v_{j+1}}{v_j}$. Next, we show that for any admissible availability set, one can rescale the weights without changing the proportional probabilities within the availability set.

Suppose $r_j < \eta/|A|$. Then, for any availability set S that satisfies $S \cap H_j \neq \varnothing$ and $S \setminus H_j \neq \varnothing$ (i.e., an availability set that contains at least one action in H_j and one outside of H_j), it holds that:

$$\sum_{a \in S \setminus H_j} \frac{v_a}{v(S)} \leq \frac{|S \setminus H_j|}{|S \cap H_j|} \cdot r_j < \frac{\eta |S \setminus H_j|}{|S|} \leq \sum_{a \in S \setminus H_j} \pi^\eta(a|S).$$

The strict inequality above holds because $|S| \leq |A| \cdot |S \cap H_j|$. Hence, whenever an availability set contains actions both in and out of H_j, π^η assigns strictly more total probability to actions outside H_j than the normalized weights do. In particular, it holds that $\sum_{a \in S \setminus H_j} \left(\pi^\eta(a|S) - \frac{v_a}{v(S)} \right) \geq 0$ for every availability set in the game. This implies that if such an availability set occurs with strictly positive probability, the induced marginals satisfy $\sum_{a \notin H_j} (\mu_a^\eta - \hat{\mu}_a(v)) = \sum_S \rho(S) \sum_{a \in S \setminus H_j} (\pi^\eta(a \mid S) - \frac{v_a}{v(S)}) > 0$. This contradicts equality of the induced marginals as established above. Hence, there does not exist any availability set with positive probability that contains actions both in and out of H_j.

The preceding argument ensures that multiplying $v_{j+1}, \ldots, v_{|A|}$ (i.e., the weights outside of H_j) by some constant c does *not* change the strategy induced by the rescaled compact vector. In our setting, we can select $c = \frac{\eta}{|A| r_j}$, which ensures that the ratio after rescaling is $r'_j = \frac{v'_{j+1}}{v'_j} = \frac{\eta}{|A|}$. This also preserves the ordering and leaves all other consecutive ratios unchanged. Hence, we can perform this rescaling for each j with $r_j < \frac{\eta}{|A|}$, then normalize the resulting vector to obtain a suitable w. By construction, $\hat{\mu}(w) = \mu^\eta$, and $\frac{w_{j+1}}{w_j} \geq \frac{\eta}{|A|}$ for each $j < |A|$. Since we have that $w_1 \geq 1/|A|$, it holds that

$$\min_a w_a = w_{|A|} = w_1 \prod_{j=1}^{|A|-1} \frac{w_{j+1}}{w_j} \geq \frac{1}{|A|} \left(\frac{\eta}{|A|} \right)^{|A|-1}. \tag{64}$$

This concludes the proof of the lemma. □

In the remainder of the proof, we apply Lemma D.3 with $\eta = \frac{\delta}{4}$ to obtain a $w \in \Delta(A)$ that satisfies the conditions in (61). We consider an ordering of the elements of w so that $w_1 = \max_a w_a$, where w_a is the a-th element of w and $1 \leq a \leq |A|$. We first show that $\|\tilde{w} - w\|_1 \leq \delta$. Select $b = \lceil \log_2(\frac{4|A|}{\delta}) \rceil$ and $e_a = \lceil \log_2 \frac{1}{w_a} \rceil$. Then, we can define q_a to be w_a truncated downward, i.e.,

$$q_a = \frac{\lfloor 2^{b+e_a} w_a \rfloor}{2^{b+e_a}}. \tag{65}$$

Since $2^{-e_a} \leq w_a < 2^{1-e_a}$, it holds that $2^b \leq \lfloor 2^{b+e_a} w_a \rfloor < 2^{b+1}$. Thus, $0 \leq w_a - q_a < 2^{-b-e_a} \leq 2^{-b} w_a \leq 2^{-b}$. Hence, each q_a is positive and rational.

However, q may sum to below 1. Let the deficit be $\kappa = 1 - ||q||_1 \geq 0$. To fix this deficit in our approximate $\tilde{w}$, we can add κ to q_1 corresponding to $w_1 = \max_a w_a$ such that

$$\tilde{w}_a = \begin{cases} q_1 + \kappa & a = 1 \\ q_a & a \neq 1 \end{cases}.$$

Clearly, $\tilde{w}$ is rational since κ is also rational. Then, we have

$$\begin{aligned}||\tilde{w} - w||_1 &= |w_1 - \tilde{w}_1| + \sum_{a \neq 1} |w_a - \tilde{w}_a| \\ &\leq \kappa + ||q - w||_1 \\ &\leq |A| \cdot 2^{-b} + |A| \cdot 2^{-b} \\ &= 2|A| \cdot 2^{-b} \\ &\leq \delta.\end{aligned}$$

Note that despite the above, it remains to control the number of bits needed to represent $\tilde{w}$. In particular, notice that e_a could require arbitrarily many bits without a lower bound on w_a.

Using Lemma D.3, we proceed to show the number of bits representing $\tilde{w}$. In particular, by selecting $\delta = 4\eta$, the lemma ensures that $\max_a e_a \leq \left\lceil \log_2 |A| + (|A| - 1) \log_2 \frac{4|A|}{\delta} \right\rceil \leq |A| \cdot b$. Each truncation in Equation (65) requires $b + 1$ bits, each e_a requires $O(\log(|A| \cdot b + 2))$ bits, and $\tilde{w}_1 = q_1 + \kappa$ also requires $O(|A|b)$ bits as its denominator divides $2^{b + \max_a e_a}$. Thus, overall, we require $O(|A|(b + \log(|A|b + 2))) = O(|A|b) = O(|A| \log(|A|/\delta))$ bits. This is because we have $b = O(\log(|A|/\delta))$ and $\log |A| \leq b$.

Next, we need to establish that the induced marginal $\hat{\mu}(\tilde{w})$ approximates μ, i.e., $\|\hat{\mu}(\tilde{w}) - \mu\|_1 \leq \delta$. First, by summing over actions, we get that

$$\kappa = \sum_a (w_a - q_a) \leq 2^{-b}. \tag{66}$$

For $a \neq 1$, we have that $|\tilde{w}_a - w_a| \leq 2^{-b} w_a$. For action 1, since $w_1 \geq 1/|A_i|$, we have $0 \leq \tilde{w}_1 - w_1 = \sum_{a \neq 1} (w_a - q_a) \leq 2^{-b} \leq |A| 2^{-b} w_1$.

Thus, for each a, it holds that $|\tilde{w}_a - w_a| \leq |A| 2^{-b} w_a$. Towards obtaining the marginals corresponding to $\tilde{w}$ and w, using the notation established in Equation (60), we have:

$$\begin{aligned}\sum_{a \in S} \left| \frac{\tilde{w}_a}{\tilde{w}(S)} - \frac{w_a}{w(S)} \right| &\leq \sum_{a \in S} \tilde{w}_a \left| \frac{1}{\tilde{w}(S)} - \frac{1}{w(S)} \right| + \frac{\sum_{a \in S} |\tilde{w}_a - w_a|}{w(S)} \\ &= \frac{|w(S) - \tilde{w}(S)| + \sum_{a \in S} |\tilde{w}_a - w_a|}{w(S)} \\ &\leq \frac{2 \sum_{a \in S} |\tilde{w}_a - w_a|}{w(S)} \\ &\leq 2|A| 2^{-b} \leq \frac{\delta}{2}\end{aligned}$$

The penultimate inequality uses the fact that $|\tilde{w}_a - w_a| \leq |A| 2^{-b} w_a$. Finally, taking the empirical average over availability sets according to ρ gives

$$\|\hat{\mu}(\tilde{w}) - \hat{\mu}(w)\|_1 \leq \delta/2. \tag{67}$$

From Lemma D.3, we also have that $\|\hat{\mu}(w) - \mu\|_1 \leq 2\eta = \delta/2$. Thus,

$$\|\hat{\mu}(\tilde{w}) - \mu\|_1 \leq \|\hat{\mu}(\tilde{w}) - \hat{\mu}(w)\|_1 + \|\hat{\mu}(w) - \mu\|_1 \leq \delta, \tag{68}$$

as required. □

As a corollary, it follows that for any 2p0s-GSAS that meets Assumption 3.3, the dyadic truncation procedure above ensures the existence of an *approximate* bit-compact representation of NE.

Corollary D.4. *Every 2p0s-GSAS that satisfies Assumption 3.3 admits, for every $\epsilon \in (0, 1)$, an ϵ-NE which, for player i, can be represented by $O(|A_i| \log(|A_i|/\epsilon))$ bits.*

Proof. Let (μ_1^*, μ_2^*) be marginals corresponding to NE in a 2p0s-GSAS meeting the assumption. We can apply Theorem D.2 to each μ_i^* with $\delta = \epsilon/2$, which ensures $\|\hat{\mu}_i(\tilde{w}_i) - \mu_i^*\|_1 \leq \epsilon/2$. Then, applying the SPR inequality from the proof of Theorem 5.8 (Equation (59)), we get

$$\mathrm{SPR}(\hat{\mu}_1(\tilde{w}_1), \hat{\mu}_2(\tilde{w}_2)) \leq \mathrm{SPR}(\mu_1^*, \mu_2^*) + \|\hat{\mu}_1(\tilde{w}_1) - \mu_1^*\|_1 + \|\hat{\mu}_2(\tilde{w}_2) - \mu_2^*\|_1 \leq \epsilon,$$

as required. □

D.3 Non-Existence of Bit-Compact Nash Equilibria in 2p0s-GSAS

In this section, we provide a result that establishes non-existence of bit-compact approximate NE strategies in 2p0s-GSAS that do not meet Assumption 3.3.

Theorem D.5 (No ϵ-NE bit-compact representation)**.** *The set*

$$\Gamma = \{\mathcal{G} \mid \mathcal{G} \textit{ is a 2p0s-GSAS}\}$$

does not have a bit-compact representation of ϵ-NE.

Theorem D.5 further justifies the use of Assumption 3.3 in our analysis: indeed, once the assumption is relaxed, there does not exist an algorithm that can play a bit-compact representation of ϵ-NE in the family of 2p0s-GSAS.

Proof sketch. Our proof of Theorem D.5 is conceptually similar to counting arguments, such as those in (Arora & Barak, 2009, Chapter 6), to show circuit lower bounds: we construct a doubly exponentially large family of games, all of which must have different bit-compact representations in order to play an ϵ-NE. The family of interest arises from a mapping from a binary function f to a 2p0s-GSAS, with a matching pennies setup and m additional 'signal' actions for player 1 which are strictly dominated by Heads. The availability sets are such that player 1 always has Heads and Tails available, while the availability of the signal actions is used as input to f, with player 2 having exactly one of Heads or Tails available, as determined by f. Since player 2 can never deviate from their single valid strategy, a Nash Equilibrium is for player 1 to use the availability of its signal actions to also play Heads or Tails according to f.

We use the Gilbert-Varshamov bound (Gilbert, 1952; Varshamov, 1957) from coding theory, which is written explicitly in Theorem D.6 following (Guruswami et al., 2025, Section 4.2) and Druk & Ishai (2014). We use this bound to construct an exponentially large family of games where for all games in the family, and for a constant fraction of the sampled signal availabilities, player 2 receives a different available action. We use a counting argument to establish that if there exists an ϵ-NE bit-compact representation, when $\epsilon < \frac{1}{4}$ for at least two games in the family player 1 must receive the same bit string and must play the same way on both games. This collision allows us to establish a contradiction, since by construction the same strategy cannot be an ϵ-NE for both games.

Theorem D.6 (Gilbert-Varshamov Bound for $q = 2$)**.** *For every $0 < \delta < \frac{1}{2}$ there exists a family of binary linear codes $\mathcal{C}$ with rate $R(\mathcal{C}) \geq 1 - H_2(\delta)$ and relative distance $\delta(\mathcal{C}) \geq \delta$.*

From the Gilbert-Varshamov bound there exists a set C of binary strings of length l with

$$|C| \geq \frac{2^l}{\mathrm{Vol}_2(\delta l, l)} \geq 2^{l(1-H_2(\delta))} \quad \text{and} \min_{\substack{c_1, c_2 \in C \\ c_1 \neq c_2}} d_{\text{Hamming}}(c_1, c_2) \geq \delta l, \tag{69}$$

where $\mathrm{Vol}_2(\delta l, l)$ is the volume of a Hamming ball of radius δl in $\{0,1\}^l$ and $H_2(\delta) = -\delta \log_2(\delta) - (1-\delta)\log_2(1-\delta)$ is the binary entropy function.

Proof of Theorem D.5. **Game generator definition.** We first define $\mathcal{G}(f)$, which constructs a 2p0s-GSAS from any binary function $f : \{0,1\}^m \to \{0,1\}$.

Let $A_1 = \{H, T, 1, 2, \ldots, m\}$ and $A_2 = \{H, T\}$ be the action sets of players 1 and 2 respectively. Moreover, let the (zero-sum) payoff function be

$$u_1(a_1, a_2) = -u_2(a_1, a_2) = \begin{cases} 1 & \text{if } a_1 = a_2, \\ 0 & \text{if } a_1 \neq a_2, a_1 \in \{H, T\}, \\ -1 & \text{if } a_1 \in \{1, 2, \ldots, m\}. \end{cases} \tag{70}$$

Let the game have availability sets $\mathcal{S}_1 = \{S_{1,j} \mid j \subseteq \{1, 2, \ldots, m\}\}$ with $S_{1,j} = \{H, T\} \cup j$, and $\mathcal{S}_2 = \{S_{2,H}, S_{2,T}\}$ with $S_{2,H} = \{H\}, S_{2,T} = \{T\}$. The distribution over availability sets is

$$\rho(S_{1,j}, S_2) = \begin{cases} \frac{1}{2^m} & \text{if } f(\mathbb{1}_j) = 1, S_2 = \{H\}, \\ \frac{1}{2^m} & \text{if } f(\mathbb{1}_j) = 0, S_2 = \{T\}, \\ 0 & \text{otherwise,} \end{cases} \tag{71}$$

where $\mathbb{1}_j \in \{0,1\}^m$ with $\mathbb{1}_j(i) \coloneqq \begin{cases} 1 & i \in j \\ 0 & i \notin j \end{cases}$ is the Boolean indicator vector.

Game family construction. With this method to generate a 2p0s-GSAS from a binary function f, we now construct a suitable family of games. Assume toward a contradiction that there exists an ϵ-NE bit-compact representation R. Fix an $\epsilon < \frac{1}{4}$ and let δ be a constant such that $\frac{1}{2} > \delta > 2\epsilon$. From the Gilbert-Varshamov bound (69), we know there exists a set of binary strings $C \subseteq \{0,1\}^{2^m}$ with relative distance δ and $|C| \geq 2^{(1-H_2(\delta))2^m}$. Therefore, for any R, with $k \in \text{poly}\left(m + 2, \log\left(\frac{1}{\epsilon}\right)\right)$ the length of the bit-compact representation bit string for a game, there exists an m_0 such that $\forall m \geq m_0, 2^{2^{m(1-H_2(\delta))}} > 2^k$.

Fix such an $m \geq m_0$ and let C be such a set of binary strings. Given a codeword $c \in C$, we construct a binary function by outputting the value of the codeword at the index of the input $f_c : \{0, 1, \ldots, 2^m - 1\} \to \{0, 1\}, f_c(x) = c_x$.

We thus consider the set of GSAS given by $\gamma = \{\mathcal{G}(f_c) \mid c \in C\}$. Note γ is a strict subset of the set of all 2p0s-GSAS, and so if γ does not have an ϵ-NE bit-compact representation, neither does the set of all 2p0s-GSAS. As $|\gamma| = |C| \geq 2^{(1-H_2(\delta))2^m}$, but there are only $2^{\text{poly}\left(m+2,\log\left(\frac{1}{\epsilon}\right)\right)}$ possible bit-compact representation bit strings, there exist games $\mathcal{F}, \mathcal{K} \in \gamma$ such that $\mathcal{F} \neq \mathcal{K}$ but $x^{\mathcal{F}} = x^{\mathcal{K}} =: x$.

Cannot be an ϵ-NE in both games. We proceed by showing that the strategy played by R using the bit string x cannot be an ϵ-NE for both $\mathcal{F}$ and $\mathcal{K}$.

Let $\pi_2^{\mathcal{G}}$ be the single valid strategy for player 2 in game $\mathcal{G} \in \gamma$ (noting that any game in γ has a single action choice available for player 2 at all times).

Let π_R be the strategy output by $R(S_1, x)$, and $\pi_{R'}$ the strategy that plays H whenever π_R plays $\{1, 2, \ldots, m\}$ but is otherwise identical to π_R. As actions $\{1, 2, \ldots, m\}$ are strictly dominated by $\{H\}$, $U(\pi_{R'}, \pi_2) \geq U(\pi_R, \pi_2)$. We can thus assume without loss of generality that actions $\{1, 2, \ldots, m\}$ are never played by $R(S_1, x)$.

Let $\pi_1^{\mathcal{G}}(S_1) = \begin{cases} H & \text{if } f_{\mathcal{G}}(\mathbb{1}_{S_1}) = 1 \\ T & \text{if } f_{\mathcal{G}}(\mathbb{1}_{S_1}) = 0 \end{cases}$ be the strategy for player 1 in a GSAS $\mathcal{G} \in \gamma$ which plays the element that, by definition of ρ, player 2 is forced to play. Moreover, note that $U_1^{\mathcal{G}}(\pi_1^{\mathcal{G}}, \pi_2^{\mathcal{G}}) = 1$. Since player 2 cannot deviate from the single valid strategy, and 1 is the maximal game value, $\pi^{\mathcal{G}}$ is a NE.

Then, since by definition R outputs an ϵ-NE,

$$U_1^{\mathcal{F}}(\pi_1^{\mathcal{F}}, \pi_2^{\mathcal{F}}) - \epsilon \leq U_1^{\mathcal{F}}(\pi_R, \pi_2^{\mathcal{F}}) \tag{72}$$

$$1 - \epsilon \leq U_1^{\mathcal{F}}(\pi_R, \pi_2^{\mathcal{F}}) \tag{73}$$

$$U_1^{\mathcal{F}}(\pi_R, \pi_2^{\mathcal{F}}) = \mathbb{E}_{S \sim \rho}[U_1^{\mathcal{F}}(R(S_1, x), \pi_2^{\mathcal{F}}(S_2))] \tag{74}$$

$$= \sum_{S \in \mathcal{S}} \Pr[S] \cdot U_1^{\mathcal{F}}(R(S_1, x), \pi_2^{\mathcal{F}}(S_2)). \tag{75}$$

By definition of γ and the corresponding ρ for any game (recalling that $\pi_2^{\mathcal{F}}$ is restricted to a single output determined by $f_{\mathcal{F}}$ and π_R never plays actions $\{1, 2, \ldots, m\}$) we can expand $U_1^{\mathcal{F}}$ from Equation (75) to get

$$U_1^{\mathcal{F}}(\pi_R, \pi_2^{\mathcal{F}}) = \frac{1}{2^m} \sum_{S_1 \in \mathcal{S}_1} \left\{R(S_1, x)_H \cdot f_{\mathcal{F}}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_{\mathcal{F}}(\mathbb{1}_{S_1}))\right\} \tag{76}$$

$$1 - \epsilon \leq \frac{1}{2^m} \sum_{S_1 \in \mathcal{S}_1} \left\{R(S_1, x)_H \cdot f_{\mathcal{F}}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_{\mathcal{F}}(\mathbb{1}_{S_1}))\right\}. \tag{77}$$

We now partition player 1's availability sets $\mathcal{S}_1$ into two disjoint sets: $S_=$, where the binary functions that generated $\mathcal{F}$ and $\mathcal{K}$ are equal, using the availability set as a Boolean indicator vector as input (cf. Equation (71)), and $S_\neq$, where the binary functions differ. Formally, we have

$$S_= = \{S_1 \in \mathcal{S}_1 \mid f_\mathcal{F}(\mathbb{1}_{S_1}) = f_\mathcal{K}(\mathbb{1}_{S_1})\} \tag{78}$$

$$S_\neq = \{S_1 \in \mathcal{S}_1 \mid f_\mathcal{F}(\mathbb{1}_{S_1}) \neq f_\mathcal{K}(\mathbb{1}_{S_1})\}. \tag{79}$$

Note that $S_= \cup S_\neq = \mathcal{S}_1, S_= \cap S_\neq = \varnothing$. Splitting Equation (77) to separately sum over these sets gives

$$\begin{aligned} 1 - \epsilon \leq \frac{1}{|\mathcal{S}_1|} \sum_{S_1 \in S_=} \big\{ R(S_1, x)_H \cdot f_\mathcal{F}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_\mathcal{F}(\mathbb{1}_{S_1})) \big\} \\ + \frac{1}{|\mathcal{S}_1|} \sum_{S_1 \in S_\neq} \big\{ R(S_1, x)_H \cdot f_\mathcal{F}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_\mathcal{F}(\mathbb{1}_{S_1})) \big\}. \end{aligned} \tag{80}$$

As payoffs are bounded by 1,

$$\frac{1}{|\mathcal{S}_1|} \sum_{S_1 \in S_=} \big\{ R(S_1, x)_H \cdot f_\mathcal{F}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_\mathcal{F}(\mathbb{1}_{S_1})) \big\} \leq \frac{|S_=|}{|\mathcal{S}_1|}, \tag{81}$$

which, combining with Equation (80), gives

$$1 - \frac{|S_=|}{|\mathcal{S}_1|} - \epsilon \leq \frac{1}{|\mathcal{S}_1|} \sum_{S_1 \in S_\neq} \big\{ R(S_1, x)_H \cdot f_\mathcal{F}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_\mathcal{F}(\mathbb{1}_{S_1})) \big\}. \tag{82}$$

Then, as $R(S_1, x) \in \Delta(\{H, T\})$, we have $R(S_1, x)_H + R(S_1, x)_T = 1$, and since $f_\mathcal{F}$ is always either 0 or 1, we get

$$\begin{aligned} |S_\neq| = \sum_{S_1 \in S_\neq} \big\{ R(S_1, x)_H \cdot f_\mathcal{F}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_\mathcal{F}(\mathbb{1}_{S_1})) \\ + R(S_1, x)_H \cdot (1 - f_\mathcal{F}(\mathbb{1}_{S_1})) + R(S_1, x)_T \cdot f_\mathcal{F}(\mathbb{1}_{S_1}) \big\}. \end{aligned} \tag{83}$$

Splitting the sum and substituting into Equation (82) gives

$$1 - \frac{|S_=|}{|\mathcal{S}_1|} - \epsilon \leq \frac{|S_\neq|}{|\mathcal{S}_1|} - \frac{1}{|\mathcal{S}_1|} \sum_{S_1 \in S_\neq} \big\{ R(S_1, x)_H \cdot (1 - f_\mathcal{F}(\mathbb{1}_{S_1})) + R(S_1, x)_T \cdot f_\mathcal{F}(\mathbb{1}_{S_1}) \big\} \tag{84}$$

$$\epsilon \geq \frac{1}{|\mathcal{S}_1|} \sum_{S_1 \in S_\neq} \big\{ R(S_1, x)_H \cdot (1 - f_\mathcal{F}(\mathbb{1}_{S_1})) + R(S_1, x)_T \cdot f_\mathcal{F}(\mathbb{1}_{S_1}) \big\}. \tag{85}$$

Since in $S_\neq$ we have $f_\mathcal{K} \neq f_\mathcal{F}$, it follows that

$$\epsilon \geq \frac{1}{|\mathcal{S}_1|} \sum_{S_1 \in S_\neq} \big\{ R(S_1, x)_H \cdot f_\mathcal{K}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_\mathcal{K}(\mathbb{1}_{S_1})) \big\}. \tag{86}$$

Since the availability sets are defined by truth tables read from two binary strings from C, and $\min_{c_1, c_2 \in C; c_1 \neq c_2} d_{\text{Hamming}}(c_1, c_2) \geq \delta 2^m$, $f_\mathcal{F}$ and $f_\mathcal{K}$ must differ on at least a δ fraction of the inputs. Thus $|S_\neq|/|\mathcal{S}_1| \geq \delta \implies |S_=|/|\mathcal{S}_1| \leq 1 - \delta$. Adding the payoff component contributed by both $S_=$ from Equation (81) and $S_\neq$ from Equation (86) gives

$$1 - \delta + \epsilon \geq \frac{1}{|\mathcal{S}_1|} \sum_{S_1 \in \mathcal{S}_1} \big\{ R(S_1, x)_H \cdot f_\mathcal{K}(\mathbb{1}_{S_1}) + R(S_1, x)_T \cdot (1 - f_\mathcal{K}(\mathbb{1}_{S_1})) \big\} \tag{87}$$

$$1 + \epsilon - \delta \geq \mathbb{E}_{S \sim \rho_\mathcal{K}}[U_1^\mathcal{K}(R(S_1, x), \pi_2^\mathcal{K}(S_2))] \tag{88}$$

$$U_1^\mathcal{K}(\pi_1^\mathcal{K}, \pi_2^\mathcal{K}) + \epsilon - \delta \geq U_1^\mathcal{K}(\pi_R, \pi_2^\mathcal{K}). \tag{89}$$

As by definition $\delta > 2\epsilon$ and by the Gilbert-Varshamov bound $\delta < 1/2$, it follows that

$$U_1^\mathcal{K}(\pi_1^\mathcal{K}, \pi_2^\mathcal{K}) - \epsilon > U_1^\mathcal{K}(\pi_R, \pi_2^\mathcal{K}), \tag{90}$$

where $\epsilon < \frac{1}{4}$. Since R must be able to play an ϵ-NE on $\mathcal{K}$, we have a contradiction. □