Safe Exploitative Play with Untrusted Type Beliefs

Tongxin Li[1]* **Tinashe Handina**[2] **Shaolei Ren**[3] **Adam Wierman**[2]

[1]School of Data Science
The Chinese University of Hong Kong, Shenzhen, China
litongxin@cuhk.edu.cn

[2]Computing + Mathematical Sciences
California Institute of Technology, Pasadena, USA
{thandina, adamw}@caltech.edu

[3]Electrical & Computer Engineering
University of California, Riverside, USA
shaolei@ucr.edu

Abstract

The combination of the Bayesian game and learning has a rich history, with the idea of controlling a single agent in a system composed of multiple agents with unknown behaviors given a set of types, each specifying a possible behavior for the other agents. The idea is to plan an agent's own actions with respect to those types which it believes are most likely to maximize the payoff. However, the type beliefs are often learned from past actions and likely to be incorrect. With this perspective in mind, we consider an agent in a game with type predictions of other components, and investigate the impact of incorrect beliefs to the agent's payoff. In particular, we formally define a tradeoff between risk and opportunity by comparing the payoff obtained against the optimal payoff, which is represented by a gap caused by trusting or distrusting the learned beliefs. Our main results characterize the tradeoff by establishing upper and lower bounds on the Pareto front for both normal-form and stochastic Bayesian games, with numerical results provided.

1 Introduction

> "*The Chinese symbol for crisis is composed of two elements: one signifies danger and the other opportunity.*" — Lewis Mumford, 1944

The famous interpretation of the Chinese word for 'crisis' (known as 危機), although based on mistaken etymology, captures the dual nature of risk and opportunity. It provides an interesting metaphor for the inherent complexities within real-world multi-agent systems. These systems, where risk and opportunity are often inextricably linked, play a pivotal role across diverse domains, ranging from human-AI collaboration [1] and cyber-physical systems [2], to highly competitive environments like real-time strategy games [3] and poker [4].

In conventional applied and theoretical frameworks, it is typically assumed that all agents either cooperate or adhere to pre-defined policies [5, 6, 7, 8]. However, real-world scenarios often defy these simplifications, presenting agents that display a spectrum of behaviors ranging from cooperative, to heterogeneous, irrational, or even adversarial [9]. This deviation from expected behavior patterns

*Correspondence to: Tongxin Li <litongxin@cuhk.edu.cn>.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

complicates the dynamics of multi-agent systems, as agents cannot reliably predict the actions of their counterparts. The uncertainty regarding whether to trust or distrust predictions naturally leads to a critical tradeoff between the coexisted risk and opportunity, reflecting the dual aspects highlighted in Mumford's remark.

As critical examples, Bayesian games [10, 11] provide an approach for modeling differing types of agents in strategic environments. In these games, players form beliefs about others' types and update beliefs in response to observed actions and choose their actions accordingly. For example, in competitive settings like poker or even in simple games like matching pennies, deviating from the game theoretic optimal (GTO) strategy [12, 13] to exploit weaker opponents can be beneficial, but this approach also relies on potentially flawed type beliefs [6], making it risky to take advantage of such side-information. Similarly, in real-world problems like security games, having a prior distribution of attackers' behavioral types in a Bayesian game setting leads to advantages. However, exploiting incorrect types can be risky compared to just using minimax strategies. Despite the ubiquity of incorrect type beliefs in practical scenarios, limited attention has been paid to explore such a tradeoff, with exceptions in designing heuristically safe and exploitative strategies in specific contexts, such as with Byzantine adversaries [9] and sequential games [14].

Inaccuracies in beliefs about others' types may arise from factors such as using out-of-distribution data to generate priors, changes in opponents' behaviors, or mismatches between hypothesized types and the optimal type space, etc. As noted in [15], it is evident that prior beliefs significantly affect the long-term performance of type-based learning algorithms like the Harsanyi-Bellman Ad Hoc Coordination (HBA). Although algorithms like HBA demonstrate asymptotic convergence to correct predictions or type distributions through various methods of estimating posterior beliefs, and methods exist for detecting inaccuracies in type beliefs using empirical behavioral hypothesis testing [15], the theoretical capabilities of general algorithms remain uncertain.

In general, relying on learned beliefs presents a fundamental tradeoff between the potential payoffs from exploitative play and the risk of incorrect type beliefs. Given the potential inaccuracies in these type beliefs, using them to exploit opponents could lead to high-risk strategies. Conversely, not exploiting these beliefs might result in overly cautious play. Therefore, it is natural to investigate the impact of incorrect beliefs on the agent's payoff, in terms of a tradeoff between trusting or distrusting the beliefs of types provided by type-based learning algorithms (e.g., Bayesian learning [16], best response dynamics [17], and policy iteration with neural networks [18], etc.) in multi-agent systems. To summarize the focus of this paper, we aim to address the following critical question:

What is the fundamental tradeoff between trusting/distrusting type beliefs in games?

Contributions. Motivated by the above question, we analyze the following *payoff gap* that arises from erroneous type beliefs:

$$\Delta(\varepsilon;\pi) := \max_{d(\theta,\theta^\star)\leq\varepsilon}\left(\max_{\phi\in\Pi}\mathsf{Payoff}(\phi,\theta^\star) - \mathsf{Payoff}(\pi(\theta),\theta^\star)\right), \quad \textbf{(payoff gap)} \tag{1}$$

where θ and $\theta^\star$ are predicted and true type beliefs respectively; $d(\cdot,\cdot)$ measures the distance between θ and $\theta^\star$ bounded from above by ε and will be formally specified with concrete model contexts in Section 3 and Section 4, together with the payoff, denoted by $\mathsf{Payoff}(\cdot,\cdot)$ as a function of the true type $\theta^\star$ and the deployed strategy $\pi \in \Pi$. We denote the given strategy space by Π. The agent's strategy $\pi(\theta)$ may depend on the type belief θ. Overall, the payoff difference (1) above quantifies the worst-case gap between the optimal payoff (obtained by an optimal strategy in Π that maximizes $\mathsf{Payoff}(\cdot,\theta^\star)$), and the payoff corresponding to π.

In particular, when the agent's strategy π trusts the belief θ, it takes the opportunity to close the gap in (1) when the belief error ε is small. However, when ε increases, such a strategy incurs a high risk since it trusts the incorrect θ. Evaluating the payoff gap in (1) naturally yields a tradeoff between opportunity and risk, as illustrated on the right of Figure 1. To be more precise, we measure the tradeoff between two important quantities:

(Missed) Opportunity: $\Delta(0;\pi)$, corresponding to the case when the type beliefs are correct;
Risk: $\max_{\varepsilon>0}\Delta(\varepsilon;\pi)$ measuring the payoff difference incurred by worst-case incorrect beliefs.

In summary, the (missed) opportunity measures the discrepancy of a strategy π from the optimal strategy, which is aligned with the ground truth belief $\theta^\star$ of other players in terms of the obtained payoff. Additionally, the risk quantifies how inaccurate beliefs impact the difference in terms of

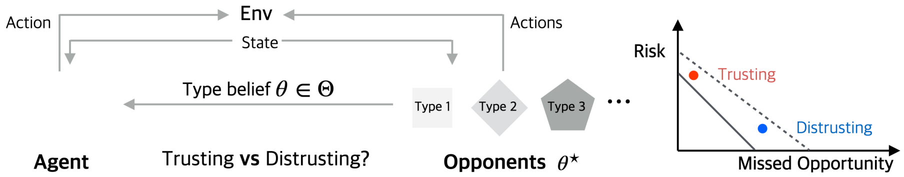


Figure 1: **Left**: A stochastic Bayesian game where an agent interacts with an environment and opponents, with a belief of their types $\theta \in \Theta$. **Right**: The tradeoff between trusting and distrusting type beliefs, with trust leading to higher risk and opportunity and distrust resulting in lower risk and opportunity, implying an opportunity-risk tradeoff with varying strategy π.

payoffs in the worst case. The goal of this paper is to investigate safe and exploitative strategies in Bayesian games that achieve near-optimal opportunity and risk.

Our main results are two-fold. Firstly, in normal-form Bayesian games, we characterize a tradeoff between the opportunity and risk. We consider a strategy as a convex combination of a safe strategy and the best response given type beliefs. Upper bounds on opportunity and risk are provided in Theorem 3.1. Conversely, lower bounds that hold for any mixed strategy are shown in Theorem 3.2. Notably, when the game is fair, and the hypothesis set Θ is sufficiently large, these bounds tightly converge. Secondly, we explore a dynamic setting in stochastic Bayesian games, where an agent, provided with type beliefs about other players, engages in interactions over time, as illustrated in Figure 1. Unlike the normal-form approach, we utilize a value-based strategy that establishes upper bounds on opportunity and risk, as outlined in Theorem 4.1. Additionally, Theorem 4.2 provides lower bounds on opportunity and risk that differ from the upper bounds by multiplicative constants, yielding a characterization of the opportunity-risk tradeoff. Finally, a case study of a security game, simulating a defender protecting an elephant population from illegal poachers, is provided in Section 5.

2 Related Work

Learning Types in Games. The concept of type-based methods dates back to the development of Bayesian games, as first established by Harsanyi in the late 1960s [10, 11]. The concepts of learning and updating beliefs appear in pioneering works like the adaptive learning [19]. While much game theory research, including work by Kalai et al. [20], focuses on equilibrium analysis through Bayesian belief-based learning, other studies, such as those by Nachbar et al. [21, 22] and [23, 24], reveal the challenges players face in making correct predictions while playing optimally under certain game conditions and assumptions. From an application perspective, Southey et al. [25] applied type-based methods to poker, where players' hands are partially hidden. They demonstrated how to maintain and use beliefs to determine the best strategies in this setting. Besides classic results exemplified above, closely related to our results, the recent work by Milec et al. [14] addresses the limitations of Nash equilibrium strategies in two-player extensive-form games, particularly their inability to exploit the weaknesses of sub-optimal opponents. They defined the exploitability of a strategy as the expected payoff that a fully rational opponent can gain beyond the game's base value and introduced a method that ensures safety, defined as an upper limit on exploitability compared to the payoff obtained using a Nash equilibrium strategy. However, the proposed tradeoff between the exploitation of the opponent given the correct model and safety against an opponent who can deviate arbitrarily from the predicted model is applicable specifically to an algorithm that employs a continual depth-limited restricted Nash response.

Online Decision-Making with Predictions. The tradeoff between opportunity and risk analyzed in this work is motivated by the recent progress in algorithms with predictions, also known as learning-augmented algorithms [26, 27] for online decision-making problems such as caching [28, 29, 30], bipartite matching [31], online optimization [32, 33], control [34, 35, 36, 37], valued-based reinforcement learning [38, 39, 40], and real-world applications [41, 42, 43, 44]. This line of work investigates the impact of untrusted predictions on two key metrics known as consistency and robustness, which are defined based on the competitive ratio of the considered contexts. It is also worth highlighting that previous works in decision-making often provide best-of-both-worlds guarantees for both stochastic and adversarial environments [45, 46], while the results in this work shed light on studying the intermediate regimes that do not fully align with either stochastic or adversarial settings and delves into interactive environments formed by players whose behaviors may deviate from the type beliefs.

Stochastic Bayesian Games. Our results presented in Section 4 draw on foundational concepts from stochastic Bayesian games as outlined by Albrecht et al. [47, 15], which merge concepts of the Bayesian games [10, 11] and the stochastic games [48], with applications in cooperative multi-agent reinforcement learning [9]. In [15], three methods—product, sum, and correlated—for integrating observed evidence into posterior beliefs have been explored. Specifically, it has been demonstrated that the Harsanyi-Bellman Ad Hoc Coordination (HBA) eventually make right future predictions under specific conditions using the product posterior beliefs. However, while HBA may eventually align with the correct type distribution, it is not guaranteed to learn it with the product posterior beliefs. Under certain conditions, HBA with the sum and correlated posterior converges to the correct type distribution. Nonetheless, even though methods like empirical behavioral hypothesis testing are available to detect inaccuracies in type beliefs, a comprehensive theoretical analysis is still lacking.

3 Normal-Form Bayesian Games with Untrusted Type Beliefs

Let $\|\cdot\|_1$ denote the ℓ_1-norm and $\|\cdot\|_{\max}$ denote the (element-wise) max norm.

3.1 Problem Setting, Opportunity, and Risk

Consider a normal-form game (NFG) with two players: Player 1 possesses a payoff matrix $A \in \mathbb{R}^{a\times b}$ with $\|A\|_{\max} \leq \alpha$, where the first player has a choices of the rows and the second player has b choices of the columns of A. Player 1 forms a belief about Player 2's mixed strategy, denoted by a distribution ρ over a set of hypothesized strategies Θ that contains a ground truth strategy $y^\star$.[2]

As a concrete example of the *payoff gap* proposed in (1), the following benchmark characterizes the gap between payoffs obtained by a strategy with the machine-learned belief ρ and an optimal strategy knowing $y^\star$ beforehand:

$$\Delta_{\mathsf{NFG}}(\varepsilon;\pi) := \max_{d(\rho,y^\star)\leq\varepsilon}\left(\max_{x\in\mathsf{P}_a} x^\top A y^\star - \pi(\rho)^\top A y^\star\right), \quad \textbf{(NFG payoff gap)} \tag{2}$$

given a fixed policy $\pi : \mathsf{P}_\Theta \to \mathsf{P}_a$ that outputs a mixed strategy of the first player knowing ρ, where $d(\rho, y^\star) := \|\mathbb{E}_\rho[y] - y^\star\|_1$, P_Θ and P_a are the sets of probability distributions on Θ and a different choices of rows respectively.

In particular, we focus on measuring a tradeoff between two important quantities, the *(missed) opportunity* $\Delta_{\mathsf{NFG}}(0;\pi)$ and the *risk* $\max_{\varepsilon>0}\Delta_{\mathsf{NFG}}(\varepsilon;\pi)$. The former measures how far the considered strategy π is away from the optimal strategy knowing the ground truth type $y^\star$ of Player 2 in terms of the payoff obtained; the latter quantifies the worst-case impact of inaccurate belief on the payoff difference.

3.2 Motivating Example: Matching Pennies

Before presenting general results, we illustrate the underlying concepts through a simple normal-form game. Consider the classic matching pennies game as an example, where the mixed strategies for the players are defined as follows.

Suppose $a = b = 2$ and each of the two players has the option to choose either Heads (`H`) or Tails (`T`). Let $y^\star \in [0,1]$ be the true probability that Player 2 plays `H` and $1-y^\star$ the probability of playing `T`. Suppose the hypothesis set Θ contains all possible mixed strategies, each corresponding to a type of Player 2. Then, Player 1 receives a belief of $y^\star$, denoted by y, and chooses a strategy π that depends on y. Similarly, let $x = \pi(y)$ be the probability that Player 1 plays `H` and $1-x$ the probability of playing `T`. If two players' actions match, Player 1 will receive 1 and -1 otherwise. The problem setting is summarized in Figure 2. Given $y^\star$ and x, the expected payoff is therefore $(2y^\star-1)(2x-1)$. The best response of Player 1, depending on y, is characterized by:

$$\mathsf{BR}(y) = \begin{cases} x = 0 & \text{if } y < \frac{1}{2}, \\ x \in [0,1] & \text{if } y = \frac{1}{2}, \\ x = 1 & \text{if } y > \frac{1}{2}. \end{cases} \tag{3}$$

[2] Note that we assume $y^\star \in \Theta$ for the ease of presentation aligning with Assumption 1 in [15], and our results can be easily generalized to the case when Θ does not contain $y^\star$ by modifying the definition of ε to incorporate incomplete or incorrect hypothesized types.

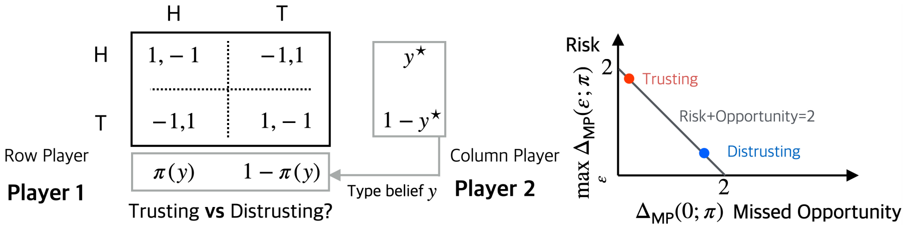


Figure 2: **Left**: Matching Pennies payoff matrix for Player 1 (row player) with type belief y and Player 2 (column player) whose strategy is defined by $y^\star$. **Right**: Opportunity-risk tradeoff that satisfies $\Delta_{\mathsf{MP}}(0;\pi) + \max_\varepsilon \Delta_{\mathsf{MP}}(\varepsilon;\pi) = 2$.

Given a strategy $\pi : [0,1] \to [0,1]$, the payoff gap in (2) for this example is instantiated as

$$\Delta_{\mathsf{MP}}(\varepsilon;\pi) := \max_{|y-y^\star|\leq\varepsilon} 2\Big((2y^\star - 1)\mathsf{BR}(y^\star) - (2y^\star - 1)\pi(y)\Big) \;\; \textbf{(MP payoff gap)}. \qquad (4)$$

Impossibility. We first consider an arbitrary strategy $\pi : [0,1] \to [0,1]$ that misses at most $(1-\lambda)$ opportunity. Suppose π chooses H with a probability $\pi(y)$ and T with a probability $1-\pi(y)$. Therefore, setting $\varepsilon = 0$, the payoff gap must satisfy

$$\Delta_{\mathsf{MP}}(0;\pi) = \max_{y\in[0,1]} 2(2y-1)\left(\mathsf{BR}(y) - \pi(y)\right) \leq 1-\lambda,$$

which implies $\pi(0) \leq (1-\lambda)/2$ and $\pi(1) \geq (1+\lambda)/2$ by setting $y=0$ and $y=1$, respectively. Thus, plugging in $y^\star = 1$ and $y = 0$, we conclude that

$$\max_\varepsilon \Delta_{\mathsf{MP}}(\varepsilon;\pi) \geq \max_{|y-y^\star|\leq 1} 2\left(\mathsf{BR}(y^\star) - \pi(y)\right) \geq 2\left(\mathsf{BR}(1) - \pi(0)\right) \geq 1+\lambda,$$

since $\mathsf{BR}(1) = 1$ by (3). This implies that π has at least $(1+\lambda)$ risk. Note that this argument holds for any arbitrarily chosen π. In a word, any strategy misses at most $1-\lambda$ opportunity must incur at least $1+\lambda$ risk, in terms of the payoff.

Mixed Strategy Existence. Now, let us construct a concrete strategy to derive upper bounds on the (missed) opportunity and risk. The idea is to combine the strategy $\mathsf{BR}(y)$, which exploits the belief y of types, and the Nash equilibrium strategy of this game, known as $\overline{x} = 1/2$, which is a solution of the minimax problem $\min_{y\in[0,1]}\max_{x\in[0,1]}(2y-1)(2x-1)$ that enhances safety.

Fix $\pi(y) := \lambda\mathsf{BR}(y) + (1-\lambda)\overline{x}$ a mixed strategy as a convex combination of the two strategies.

The payoff gap in (4) satisfies

$$\Delta_{\mathsf{MP}}(\varepsilon;\pi) \leq \max_{|y-y^\star|\leq\varepsilon} 2(2y^\star - 1)\Big(\lambda\left(\mathsf{BR}(y^\star) - \mathsf{BR}(y)\right) + (1-\lambda)\left(\mathsf{BR}(y^\star) - 1/2\right)\Big).$$

Opportunity: Therefore, there exists a mixed strategy π such that when $\varepsilon = 0$ (i.e., $\mathsf{BR}(y) = \mathsf{BR}(y^\star)$), its (missed) opportunity is bounded by

$$\Delta_{\mathsf{MP}}(0;\pi) = \max_{y\in[0,1]} 2(2y-1)\left(1-\lambda\right)\left(\mathsf{BR}(y) - 1/2\right) \leq 1-\lambda.$$

Risk: Moreover, maximizing over ε such that $y = 0$, $y^\star = 1$, $\mathsf{BR}(y^\star) - \mathsf{BR}(y) = 1$, the risk for π always satisfies $\max_\varepsilon \Delta_{\mathsf{MP}}(\varepsilon;\pi) \leq 1+\lambda$.

In a word, we find a strategy that misses $1-\lambda$ opportunity and meanwhile has $1+\lambda$ risk.

Pareto Optimality In conclusion, above construction shows that there is a mixed strategy π for the matching pennies that misses $(1-\lambda)$ opportunity and incurs $(1+\lambda)$ risk. Conversely, any strategy that misses at most $(1-\lambda)$ opportunity must have at least $(1+\lambda)$ risk. The segment on the right of Figure 2 represents a Pareto front, and the strategy constructed as a convex combination is confirmed to be Pareto optimal following the arguments above. This motivating example indicates a tight tradeoff between opportunity and risk for matching pennies. In the sequel, we further generalize this result to normal-form games.

3.3 Opportunity-Risk Tradeoff for Normal-Form Games

In general, the opportunity-risk tradeoff depends on the hypothesis set Θ that contains a subset of candidate strategies and the payoff matrix of the game. We state useful definitions to characterize properties of the hypothesis set Θ and the considered normal-form game.

Definition 1. *Given a hypothesis set Θ, we let the diameter of Θ with respect to the ℓ_1-norm be $\eta(\Theta) := \max_{y,z\in\Theta} \|y-z\|_1$. Define the following **type intensity***

$$\kappa(\Theta) := \max_{y,z\in\Theta}\left(\sum_{i:y_i\leq z_i} z_i - \sum_{i:y_i>z_i} z_i\right) \text{ subject to } \sum_{i:y_i\leq z_i} y_i < \sum_{i:y_i>z_i} y_i. \tag{5}$$

Furthermore, with fixed Θ and A, we define the maximum and value of the game by

$$\mu_\Theta(A) := \max_{y\in\Theta}\max_{x\in\mathsf{P}_a}\left|x^\top Ay\right| \textbf{ (maximum)}, \quad \nu_\Theta(A) := \min_{y\in\Theta}\max_{x\in\mathsf{P}_a} x^\top Ay \textbf{ (value)}. \tag{6}$$

The type intensity, as defined in (5), quantifies the divergence between two distributions within the set Θ, specifically those that maximize the given objective. As the density of Θ increases, the value of $\kappa(\Theta)$ approaches 1.

Theorem 3.1 (NFG EXISTENCE)**.** *Fix any Θ and consider a general-sum normal-form game where Player 1 has a payoff matrix $A\in\mathbb{R}^{a\times b}$ with $\mu_\Theta(A)\leq\mu$ and $\nu_\Theta(A)\geq\nu$. For any $0\leq\lambda\leq 1$, there exists a mixed strategy $\pi:\mathsf{P}_\Theta\to\mathsf{P}_a$ for Player 1 that misses $(1-\lambda)(\mu-\nu)$ opportunity and has $(1-\lambda)(\mu-\nu)+\lambda\mu\eta(\Theta)$ risk.*

To show Theorem 3.1, we construct a mixed strategy as follows. Denote by $\overline{x}$ as the following safe/minimax strategy of Player 1, which is the Nash equilibrium when the game is zero-sum:

$$\min_{y\in\Theta}\overline{x}^\top Ay = \max_{x\in\mathsf{P}_a}\min_{y\in\Theta} x^\top Ay. \tag{7}$$

Motivated by the matching pennies example in Section 3.2, Player 1 implements a mixed strategy $\pi(\rho) := \lambda\widetilde{x} + (1-\lambda)\overline{x}$ given the predicted belief ρ as a distribution over types in Θ, where $\widetilde{x}\in\mathsf{P}_a$ is a best response strategy given ρ such that $x^\top A\mathbb{E}_\rho[y]$ is maximized. The detailed proof of Theorem 3.1 is provided in Appendix A.

Moreover, we further show the following impossibility result, indicating that the tradeoff in Theorem 3.1 is tight. We relegate the proof of Theorem 3.2 to Appendix B.

Theorem 3.2 (NFG IMPOSSIBILITY)**.** *For any Θ satisfying $\kappa(\Theta)\geq 0$, there is a payoff matrix $A\in\mathbb{R}^{a\times b}$ with $\mu_\Theta(A)\leq\mu$ and $\nu_\Theta(A)\geq\nu$ such that for any $0\leq\lambda\leq 1$, if any mixed strategy $\pi:\mathsf{P}_\Theta\to\mathsf{P}_a$ for Player 1 misses at most $(1-\lambda)(\mu-\nu)$ opportunity, then it incurs at least $(\kappa(\Theta)\mu-\nu)(1+\lambda)$ risk.*

In particular, Theorem 3.1 and 3.2 together imply the following special case when the hypothesis set Θ contains all possible mixed strategies in P_b. This corollary highlights that for a fair game, the tradeoff is tight and the mixed strategy $\pi(\rho) := \lambda\widetilde{x} + (1-\lambda)\overline{x}$ with $\lambda\in[0,1]$ is Pareto optimal when the hypothesis set contains all possible mixed strategies.

Corollary 3.1 (NFG PARETO OPTIMALITY)**.** *Suppose $\Theta=\mathsf{P}_b$ and the game is fair, where Player 1 has a payoff matrix $A\in\mathbb{R}^{a\times b}$ with $\|A\|_{\max}\leq\alpha$. For any $0\leq\lambda\leq 1$, there exists a mixed strategy π for Player 1 that misses $(1-\lambda)\alpha$ opportunity and has $(1+\lambda)\alpha$ risk. Furthermore, there is a payoff matrix $A\in\mathbb{R}^{a\times b}$ with $\|A\|_{\max}\leq\alpha$ such that for any $0\leq\lambda\leq 1$ if any mixed strategy π for Player 1 misses at most $(1-\lambda)\alpha$ opportunity, then it incurs at least $(1+\lambda)\alpha$ risk.*

It is worth noting that the requirement can be relaxed straightforwardly so that as long as the hypothesis set contains two mixed strategies such that $\kappa(\Theta)=1$ in (5), the statement holds. The proof of Corollary 3.1 can be found in Appendix C.

4 Stochastic Bayesian Games with Untrusted Type Beliefs

In many real-world applications, agents are coupled through a shared state in stochastic Bayesian games (SBG) [47, 15], which combines standard Bayesian games [10, 11] with the stochastic games [48]. In this section, we consider an infinite-horizon time-varying discounted n-player stochastic Bayesian game, represented by a tuple $\mathcal{M} := \langle\mathcal{S},\mathcal{A},\Theta,\sigma,p,r,\gamma\rangle$, and analyze the impact of beliefs on the opportunity-risk tradeoff.

4.1 Preliminaries: MDP for Stochastic Bayesian Games

Let $\mathcal{S}$ be a finite set of states. Write $\mathcal{N} := \{1, \ldots, n\}$. Let $\mathcal{A}(i)$ be the set of finitely many actions that player $i \in \mathcal{N}$ can take at any state $s \in \mathcal{S}$. Furthermore, $\mathcal{A} := \mathcal{A}(1) \times \cdots \times \mathcal{A}(n)$ denotes the set of action profiles $a = (a(i) : i \in \mathcal{N})$ with $a(i) \in \mathcal{A}(i)$. The types of other players are unknown to the i-th player. The set $\Theta = \prod_{j \neq i} \Theta(j)$ is a product type space where each opponent j chooses types from Θ_j. We focus on the decision-making of the i-th player, considering stationary (Markov) strategies.[3] At each time $t \geq 0$ each player $j \neq i$ uses a mixed strategy $\pi_j : \mathcal{S} \times \Theta(j) \to \mathsf{P}_{\mathcal{A}(j)}$ parameterized by a type that depends on the current state and the true type $\theta_j^\star \in \Theta(j)$. We use $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ to denote the reward function for the i-th player, which is assumed to be bounded, as formally defined below.

Assumption 1. *The reward $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ satisfies that $|r(s, a)| \leq r_{\max}$ for all $s \in \mathcal{S}$, $a \in \mathcal{A}$.*

For any pair of states (s, s') and action profile $a \in \mathcal{A}$, we define $p(s'|s, a)$ as a transition probability from s to s' given an action profile a. Finally, let $\gamma \in (0, 1)$ be a discount factor. We define the expected utility of the i-th player using a strategy π_i as the expected discounted total payoff

$$J\left(\pi_i; \theta_{-i}^\star\right) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r\left(s_t, a_t\right)\right], \tag{8}$$

where $\theta_{-i}^\star := (\theta_j^\star \in \Theta(j) : j \neq i)$, with respect to stochastic processes $(s_t \sim p\left(\cdot \mid s_{t-1}, a_{t-1}\right))_{t>0}$ and $(a_t(i) \sim \pi_i\left(s_t\right))_{t\geq 0}$, which generate the state and the action trajectories. The expectation in (8) is taken with respect to all randomness induced by the initial state distribution $s_0 \sim p_0 \in \mathsf{P}_{\mathcal{S}}$, the state transition kernel p, and strategy profile π (as well as opponents' types in Θ).

4.2 Opportunity, Risk, and Type Beliefs

Suppose the i-th player has beliefs of the other players' types, provided by a machine-learned forecaster, denoted by θ_{-i}. For notational simplicity, we write π_i, the strategy for the i-th player as π, and the strategies π_{-i} for the opponents as σ (parameterized by θ_{-i}). Furthermore, we omit the subscripts and denote the type beliefs and true types as θ and $\theta^\star$, and write the payoff in (8) as $J\left(\pi\right)$ if there is no ambiguity. The i-th player uses a strategy $\pi : \mathcal{S} \times \Theta \to \mathsf{P}_{\mathcal{A}(i)}$, which is a function of the type beliefs of the other players. Given a strategy π used by the agent and strategies by other players, denoted by σ, we define the following useful value functions of the game:

$$V^{\pi,\sigma}(s) := \mathbb{E}_{p,\pi,\sigma}\left[\sum_{\tau=t}^{\infty} \gamma^{\tau-t} r\left(s_t, a_t\right) | s_t = s\right], \tag{9}$$

which satisfies the Bellman equation $V^{\pi,\sigma}(s) = \mathbb{E}_{a\sim\pi(s),a'\sim\sigma(s)}\left[\left(r + \gamma\mathbb{P}V^{\pi,\sigma}\right)\left(s, (a, a')\right)\right]$, where we define an operator $\mathbb{P}V^{\pi}(s, (a, a')) := \mathbb{E}_{s'\sim p(\cdot|s,(a,a'))}\left[V^{\pi,\sigma}(s')\right]$.

The following worst-case payoff gap is defined similarly as the one in (2) for normal-form games, which in our context can be considered as the dynamic regret for a strategy π_i against an optimal strategy. Note that an optimal strategy is also a best response strategy $\pi^\star$ knowing the true types of all players in hindsight maximizing (9). The optimal value function given σ and $\pi^\star$ is denoted by $V^{\star,\sigma}(s)$, whose Bellman optimality equation is $V^{\star,\sigma}(s) = \max_{a\in\mathcal{A}}\left(r + \gamma\mathbb{P}^\sigma V^{\star,\sigma}\right)(s, a)$.

Definition 2 (SBG Payoff Gap). *Given a stochastic Bayesian game $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \Theta, \sigma, p, r, \gamma \rangle$, for a fixed strategy π, the payoff gap is defined as*

$$\Delta_{\mathsf{SBG}}(\varepsilon; \pi) := \sup_{d(\theta,\theta^\star)\leq\varepsilon} \sup_{s_0\in\mathcal{S}} \left(V^{\pi^\star,\sigma(\theta^\star)}(s_0) - V^{\pi(\theta),\sigma(\theta^\star)}(s_0)\right), \quad \textbf{(SBG \textit{payoff gap})} \tag{10}$$

where s_0 denotes an initial state and $d\left(\theta, \theta^\star\right) := \max_{s\in\mathcal{S}} \left\|\sigma(s; \theta) - \sigma(s; \theta^\star)\right\|_1 \leq \varepsilon$.

Similar to normal-form Bayesian games, we define the (missed) opportunity and risk below.

Definition 3. *Given a payoff gap $\Delta_{\mathsf{SBG}}(\varepsilon; \pi)$ in (10), the (missed) opportunity of π is $\beta(\pi) := \Delta_{\mathsf{SBG}}(0; \pi)$ and the risk is $\delta(\pi) := \max_{\varepsilon>0} \Delta_{\mathsf{SBG}}(\varepsilon; \pi)$.*

[3] Note that if all the other players use Markov strategies, then the considered player has a best response that is a Markov strategy.

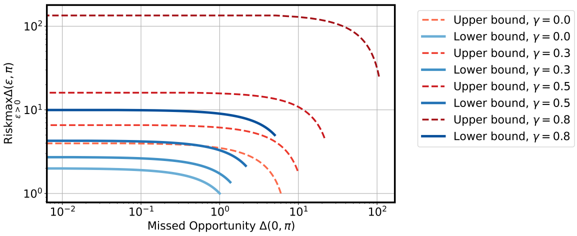


Figure 3: Comparison of upper (Theorem 4.1) and lower (Theorem 4.2) bounds on (missed) opportunity and risk with a varying discount factor $\gamma = 0.0, 0.3, 0.5, 0.8$, with $\nu = 0$ and $r_{\max} = 1$.

Our goal is to characterize a tradeoff between the opportunity and risk for stochastic Bayesian games. Finally, we define the value of the game.

Definition 4. *Given a hypothesis set Θ and a stochastic Bayesian game $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \Theta, \sigma, p, r, \gamma \rangle$, we denote the value of $\mathcal{M}$ by*

$$\nu_\Theta(\mathcal{M}) \coloneqq \max_{\pi \in \Pi} \min_{\theta_{-i} \in \Theta} J(\pi; \theta_{-i}) \quad \textbf{\textit{(value)}}. \tag{11}$$

4.3 Opportunity-Risk Tradeoff for Stochastic Bayesian Games

Based on the definitions above, our first result states an upper bound on the opportunity and risk for stochastic Bayesian games. Proving the bounds in Theorem 4.1 is nontrivial because the players are coupled by a shared state. Consequently, the analysis used in Section 3 for a simple convex combination of the best response to type beliefs θ and a safe strategy in normal-form games cannot be directly applied to the stochastic setting.

Theorem 4.1 (SBG EXISTENCE)**.** *Consider a stochastic Bayesian game $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \Theta, \sigma, p, r, \gamma \rangle$ with $\nu_\Theta(\mathcal{M}) \geq \nu$. For any $0 \leq \lambda \leq 1$, there exists a strategy π whose (missed) opportunity and risk satisfy*

$$\begin{aligned} \beta(\pi) \leq & \frac{1}{1-\lambda\gamma} \left(C_1(\gamma) r_{\max} - \gamma\nu \right) (1-\lambda), \text{ with } C_1(\gamma) \coloneqq \frac{\gamma^2 - 3\gamma + 6}{(1-\gamma)^2}, \\ \delta(\pi) \leq & \frac{1}{1-\lambda\gamma} \left(C_2(\gamma) r_{\max} - \gamma\nu \right) (1+\lambda), \text{ with } C_2(\gamma) \coloneqq \max\left(C_3(\gamma), C_4(\gamma) \right), \\ & C_3(\gamma) \coloneqq \frac{\gamma^2 - 3\gamma + 2}{1-\gamma} \text{ and } C_4(\gamma) \coloneqq \frac{2 - 2\gamma^2 + 1}{(1-\gamma)^2}, \end{aligned}$$

where $r_{\max}$ is defined in Assumption 1.

The proof of Theorem 4.1 is given in Appendix D, which constructs the following strategy.

Given type beliefs θ, we denote a parameterized strategy $\widetilde{\sigma} \coloneqq \sigma(\theta)$, and let $\overline{\sigma} = \sigma(\overline{\theta})$ be the safe strategies of the opponents, with $\overline{\theta}$ being an optimal solution of the minimax optimization in (11). Denote $\widetilde{V} \coloneqq V^{\star,\widetilde{\sigma}}$ and $\overline{V} \coloneqq V^{\star,\overline{\sigma}}$ the optimal value functions with the opponents' strategies being $\widetilde{\sigma}$ and $\overline{\sigma}$ respectively. Unlike the convex combination used to show Theorem 3.1 for normal-form Bayesian games, the strategy constructed for proving Theorem 4.1 is a value-based strategy below

$$\pi(\theta; s) \in \arg\max_{a \in \mathsf{P}_{\mathcal{A}(i)}} a^\top \left(\lambda R_{\widetilde{V}}(s) \widetilde{\sigma}(s) + (1-\lambda) R_{\overline{V}}(s) \overline{\sigma}(s) \right), \tag{12}$$

where λ is a parameter that indicates the level of trusts on type beliefs, and for a fixed state $s \in \mathcal{S}$ and a given value function $V : \mathcal{S} \to \mathbb{R}$, R_V is an $\mathcal{A}(i) \times \prod_{j \neq i} \mathcal{A}(j)$ matrix defined as

$$R_V\left((a_i, a_{-i}); s\right) \coloneqq r\left(s, (a_i, a_{-i})\right) + \gamma \sum_{s' \in \mathcal{S}} p\left(s' | s, (a_i, a_{-i})\right) V(s).$$

As indicated by Theorem 4.1, this design ensures safe and exploitative play, with its efficacy validated through the case studies presented in Section 5. Finally, the following lower bounds can be derived, implying that the upper bounds on $\beta(\pi)$ and $\delta(\pi)$ are tight and the value-based strategy formed in (12) are Pareto optimal up to multiplicative constants. Together, the bounds are illustrated in Figure 3.

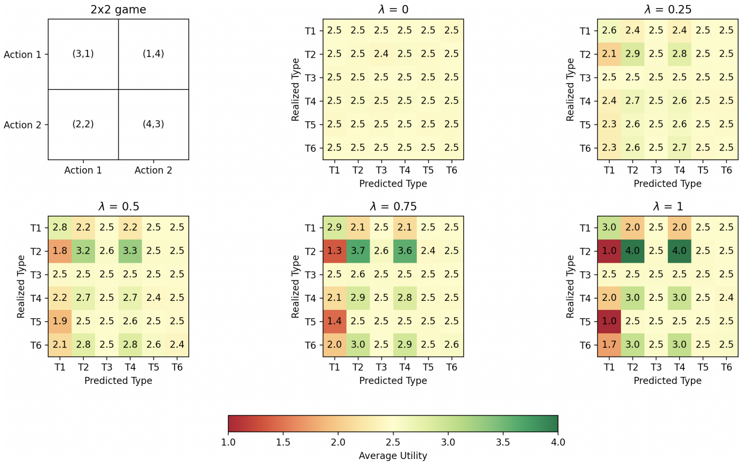


Figure 4: Comparison of average payoff for a player when varying values of λ and 6 potential discrete types for an instantiation of a 2×2 game.

Theorem 4.2 (SBG IMPOSSIBILITY)**.** *There is a stochastic Bayesian game $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \Theta, \sigma, p, r, \gamma \rangle$ whose value satisfies $\nu_\Theta(\mathcal{M}) \geq \nu$ such that for any $0 \leq \lambda \leq 1$, if any strategy π misses at most $\frac{1}{1-\lambda\gamma}\left(r_{\max} - \nu\right)(1-\lambda)$ opportunity, then it incurs at least $\frac{1}{1-\lambda\gamma}\left(r_{\max} - \nu\right)(1+\lambda)$ risk.*

In Theorem 4.2, the bounds on (missed) opportunity and risk differ from those in Theorem 4.1 by multiplicative constants. Theorem 4.2 can be derived by applying Theorem 3.2 for normal form games to a stateless MDP. We relegate the proof of Theorem 4.2 to Appendix E.

5 Case Studies

2×2 **Game.** We provide a set of numerical evaluations to showcase the opportunity-risk tradeoff across a range of game theoretic scenarios. We begin with a focus on 2×2 games and to that end we sample games spanning the topology of 2×2 matrix games as shown in [49]. This topology includes the comprehensive set of 78 strictly ordinal 2×2 games introduced by [50]. We formulate our evaluation as a single state Stochastic Bayesian Game with the state being a specific 2×2 game. This evaluation is helpful in providing a concrete illustration of the tradeoff dynamics on a wide range of games within a particular topology given the comprehensive bench-marking done on the set of 2×2 games.

We make use of 3 broad type classes (Markovian, Leader-Follower-Trigger-Agents, and Co-evoloved Neural Networks) inspired from work by [51]. It is from these classes that we construct our types which are then used in simulations with a player leveraging a strategy that is a convex combination of a safe strategy and the best response give type beliefs or predictions. Figs 4 and 5 shows the tradeoff in one particular game as an agent moves from being fully robust as indicated by $\lambda = 0$ to fully trusting of the advice when $\lambda = 1$. In Appendix F, we provide further information on the particular types as well as illustrations on a couple of other games sampled from the topology of 2×2 games to showcase the range of tradeoffs which exist in this landscape.

Security Game. Extending our empirical study to real world settings, we also provide an evaluation of the opportunity-risk tradeoff using data from wildlife tracking studies. Security games are a clear domain wherein the tradeoff between opportunity and risk is very critical to the application. Building on a large body of work that has sought to understand the implications of game theoretic analysis in the environmental conservation domain, we formulate and explore the opportunity-risk tradeoff using data from real world wildlife movements. In particular, we formulate a green security game motivated by works such as [52] wherein we simulate a defender who is trying to protect an elephant population from attackers who are illegal poachers. The defender and attacker engage a multi-state Stochastic Bayesian Game where the states are constructed from historic elephant movement data sourced from [53]. Our evaluation shows the practicality and wide ranging impact of this investigation. Figure 6

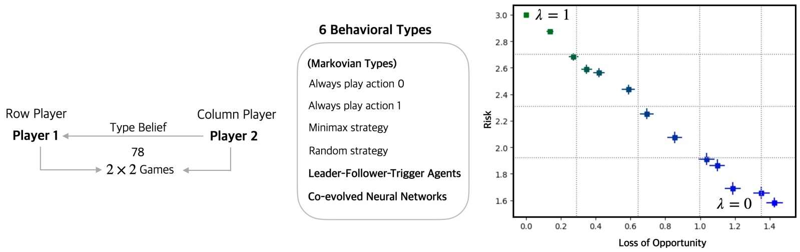


Figure 5: **Left**: 2×2 games considered in our case study; **Right**: Opportunity-risk tradeoff in the evaluation of a 2×2 game using an algorithm that has varying trust of type beliefs in 1,000 random runs. Fully trusting ($\lambda = 1$) and distrusting ($\lambda = 0$) type beliefs yield a best response strategy and a minimax strategy correspondingly.

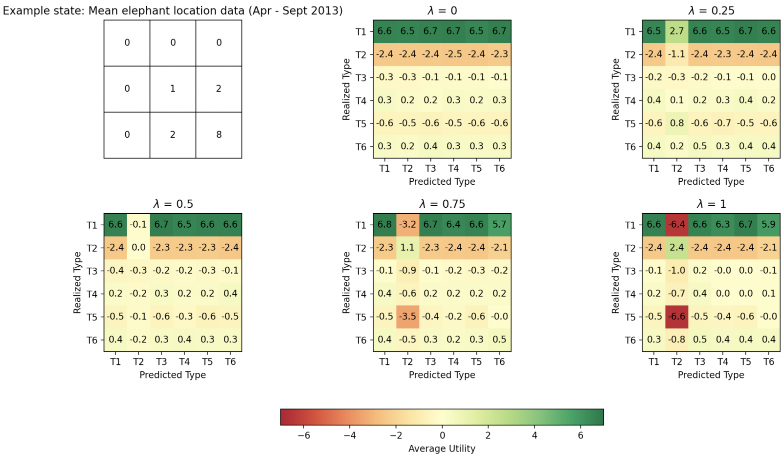


Figure 6: Comparison of average payoff for the defender in a security game protecting an elephant population against illegal poachers when varying values of λ and 6 potential discrete attacker types.

shows the tradeoff for this setting evaluated for each particular type. We include implementation details on this game in Appendix F.

6 Concluding Remarks

In this study, we explored the fundamental limits of safe and exploitative strategies within both normal-form and stochastic Bayesian games, where pre-established type beliefs about opponents are considered. Given that these type beliefs may be inaccurate, relying on them to exploit opponents can result in high-risk strategies. Conversely, not leveraging these beliefs yields overly cautious play, leading to missed opportunities. We have quantified these dynamics by providing upper and lower bounds on the payoff gaps corresponding to different type belief inaccuracies, thereby characterizing the tradeoffs between opportunity and risk. These bounds are consistent up to multiplicative constants.

Limitations and Future Directions. In our current problem setting, the dynamics of a stochastic Bayesian game is assumed to be stationary, aligning with the canonical models in the related literature (for example [51]). To address this limitation, we plan to extend our framework to include time-varying type beliefs, addressing the absence of analysis for MDPs with dynamic transition probabilities and reward structures. Additionally, refining the opportunity and risk bounds to make them tighter would provide more precise strategic insights.

Acknowledgement

We would like to thank all the anonymous reviewers for their helpful comments.

Tongxin Li was supported in part by the National Natural Science Foundation of China (NSFC) under grant No.72301234, the GuangDong Basic and Applied Basic Research Foundation, NSFC under grant No. 62336005, the PengCheng Peacock Supporting Scientific Research Fund Category C (2024TC0024), the GuangDong Provincial Key Laboratory of Mathematical Foundations for Artificial Intelligence (2023B1212010001), and the Shenzhen Key Lab of Crowd Intelligence Empowered Low-Carbon Energy Network (No. ZDSYS20220606100601002).

Adam Wierman was supported in part by the U.S. NSF under grants CNS–2146814, CPS–2136197, CNS–2106403, NGSDI–2105648.

References

[1] Xue Yan, Jiaxian Guo, Xingzhou Lou, Jun Wang, Haifeng Zhang, and Yali Du. An efficient end-to-end training approach for zero-shot human-ai coordination. *Advances in Neural Information Processing Systems*, 36, 2024.

[2] Shiyong Wang, Jiafu Wan, Daqiang Zhang, Di Li, and Chunhua Zhang. Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. *Computer networks*, 101:158–168, 2016.

[3] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[4] Noam Brown and Tuomas Sandholm. Superhuman ai for multiplayer poker. *Science*, 365(6456):885–890, 2019.

[5] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30, 2017.

[6] Stefano V Albrecht and Peter Stone. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, 258:66–95, 2018.

[7] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 21(178):1–51, 2020.

[8] Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, and Tuomas Sandholm. On the convergence of no-regret learning dynamics in time-varying games. *Advances in Neural Information Processing Systems*, 36, 2024.

[9] Simin Li, Jun Guo, Jingqiao Xiu, Ruixiao Xu, Xin Yu, Jiakai Wang, Aishan Liu, Yaodong Yang, and Xianglong Liu. Byzantine robust cooperative multi-agent reinforcement learning as a bayesian game. In *The Twelfth International Conference on Learning Representations*, 2023.

[10] John C Harsanyi. Games with incomplete information played by “bayesian” players, i–iii part i. the basic model. *Management science*, 14(3):159–182, 1967.

[11] John C Harsanyi. Games with incomplete information played by “bayesian” players part ii. bayesian equilibrium points. *Management science*, 14(5):320–334, 1968.

[12] Lawrence Friedman. Optimal bluffing strategies in poker. *Management Science*, 17(12):B–764, 1971.

[13] Norman Zadeh. Computation of optimal poker strategies. *Operations Research*, 25(4):541–562, 1977.

[14] David Milec, Ondřej Kubíček, and Viliam Lisý. Continual depth-limited responses for computing counter-strategies in sequential games. *arXiv preprint arXiv:2112.12594*, 2021.

[15] Stefano V Albrecht, Jacob W Crandall, and Subramanian Ramamoorthy. Belief and truth in hypothesised behaviours. *Artificial Intelligence*, 235:63–94, 2016.

[16] James S Jordan. Bayesian learning in normal form games. *Games and Economic Behavior*, 3(1):60–81, 1991.

[17] Martin Bichler, Max Fichtl, and Matthias Oberlechner. Computing bayes–nash equilibrium strategies in auction games via simultaneous online dual averaging. *Operations Research*, 2023.

[18] Martin Bichler, Maximilian Fichtl, Stefan Heidekrüger, Nils Kohring, and Paul Sutterer. Learning equilibria in symmetric auction games using artificial neural networks. *Nature machine intelligence*, 3(8):687–695, 2021.

[19] Paul Milgrom and John Roberts. Adaptive and sophisticated learning in normal form games. *Games and economic Behavior*, 3(1):82–100, 1991.

[20] Ehud Kalai and Ehud Lehrer. Rational learning leads to nash equilibrium. *Econometrica: Journal of the Econometric Society*, pages 1019–1045, 1993.

[21] John H Nachbar. Prediction, optimization, and learning in repeated games. *Econometrica: Journal of the Econometric Society*, pages 275–309, 1997.

[22] John H Nachbar. Beliefs in repeated games. *Econometrica*, 73(2):459–480, 2005.

[23] Dean P Foster and H Peyton Young. On the impossibility of predicting the behavior of rational agents. *Proceedings of the National Academy of Sciences*, 98(22):12848–12853, 2001.

[24] Eddie Dekel, Drew Fudenberg, and David K Levine. Learning to play bayesian games. *Games and economic behavior*, 46(2):282–303, 2004.

[25] Finnegan Southey, Michael P Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and Chris Rayner. Bayes' bluff: Opponent modelling in poker. *arXiv preprint arXiv:1207.1411*, 2012.

[26] Mohammad Mahdian, Hamid Nazerzadeh, and Amin Saberi. Online optimization with uncertain information. *ACM Transactions on Algorithms (TALG)*, 8(1):1–29, 2012.

[27] Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ml predictions. In *Advances in Neural Information Processing Systems*, pages 9661–9670, 2018.

[28] Dhruv Rohatgi. Near-optimal bounds for online caching with machine learned advice. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1834–1845. SIAM, 2020.

[29] Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. *Journal of the ACM (JACM)*, 68(4):1–25, 2021.

[30] Sungjin Im, Ravi Kumar, Aditya Petety, and Manish Purohit. Parsimonious learning-augmented caching. In *International Conference on Machine Learning*, pages 9588–9601. PMLR, 2022.

[31] Antonios Antoniadis, Christian Coester, Marek Elias, Adam Polak, and Bertrand Simon. Online metric algorithms with untrusted predictions. In *International Conference on Machine Learning*, pages 345–355. PMLR, 2020.

[32] Nicolas Christianson, Tinashe Handina, and Adam Wierman. Chasing convex bodies and functions with black-box advice. In *Conference on Learning Theory*, pages 867–908. PMLR, 2022.

[33] Pengfei Li, Jianyi Yang, Adam Wierman, and Shaolei Ren. Robust learning for smoothed online convex optimization with feedback delay. *Advances in Neural Information Processing Systems*, 36, 2024.

[34] Tongxin Li, Yue Chen, Bo Sun, Adam Wierman, and Steven Low. Information aggregation for constrained online control. *ACM SIGMETRICS Performance Evaluation Review*, 49(1):7–8, 2021.

[35] Tongxin Li, Ruixiao Yang, Guannan Qu, Guanya Shi, Chenkai Yu, Adam Wierman, and Steven Low. Robustness and consistency in linear quadratic control with untrusted predictions. *ACM SIGMETRICS Performance Evaluation Review*, 50(1):107–108, 2022.

[36] Yiheng Lin, Yang Hu, Guannan Qu, Tongxin Li, and Adam Wierman. Bounded-regret mpc via perturbation analysis: Prediction error, constraints, and nonlinearity. *Advances in Neural Information Processing Systems*, 35:36174–36187, 2022.

[37] Tongxin Li, Ruixiao Yang, Guannan Qu, Yiheng Lin, Adam Wierman, and Steven H Low. Certifying black-box policies with stability for nonlinear control. *IEEE Open Journal of Control Systems*, 2:49–62, 2023.

[38] Noah Golowich and Ankur Moitra. Can q-learning be improved with advice? In *Conference on Learning Theory*, pages 4548–4619. PMLR, 2022.

[39] Tongxin Li, Yiheng Lin, Shaolei Ren, and Adam Wierman. Beyond black-box advice: Learning-augmented algorithms for mdps with q-value predictions. *Advances in Neural Information Processing Systems*, 36, 2024.

[40] Jianyi Yang, Pengfei Li, Tongxin Li, Adam Wierman, and Shaolei Ren. Anytime-competitive reinforcement learning with policy prior. *Advances in Neural Information Processing Systems*, 36, 2024.

[41] Tongxin Li, Bo Sun, Yue Chen, Zixin Ye, Steven H Low, and Adam Wierman. Learning-based predictive control via real-time aggregate flexibility. *IEEE Transactions on Smart Grid*, 12(6):4897–4913, 2021.

[42] Nicolas Christianson, Christopher Yeh, Tongxin Li, Mahdi Torabi Rad, Azarang Golmohammadi, and Adam Wierman. Robustifying machine-learned algorithms for efficient grid operation. In *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*, 2022.

[43] Tongxin Li. *Learning-Augmented Control and Decision-Making: Theory and Applications in Smart Grids*. PhD thesis, California Institute of Technology, 2023.

[44] Tongxin Li and Chenxi Sun. Out-of-distribution-aware electric vehicle charging. *IEEE Transactions on Transportation Electrification*, 2024.

[45] Tiancheng Jin, Longbo Huang, and Haipeng Luo. The best of both worlds: stochastic and adversarial episodic mdps with unknown transition. *Advances in Neural Information Processing Systems*, 34:20491–20502, 2021.

[46] Idan Amir, Guy Azov, Tomer Koren, and Roi Livni. Better best of both worlds bounds for bandits with switching costs. *Advances in neural information processing systems*, 35:15800–15810, 2022.

[47] Stefano V Albrecht and Subramanian Ramamoorthy. A game-theoretic model and best-response learning method for ad hoc coordination in multiagent systems. *arXiv preprint arXiv:1506.01170*, 2015.

[48] Lloyd S Shapley. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.

[49] Bryan Bruns. Names for games: Locating 2 × 2 games. *Games*, 6(4):495–520, October 2015.

[50] Anatol Rapoport and Melvin Guyer. A taxonomy of 2 × 2 games. *General Systems: Yearbook of the Society for General Systems Research*, 11:203–214, 1966.

[51] Stefano V. Albrecht, Jacob W. Crandall, and Subramanian Ramamoorthyc. Belief and truth in hypothesised behaviours. *arXiv preprint arXiv:1507.07688*, 2015.

[52] Fei Fang, Peter Stone, and Milind Tambe. When security games go green: Designing defender strategies to prevent poaching and illegal fishing. *IJCAI*, 2015.

[53] Chamaille-Jammes. African elephant (migration) chamaillé-jammes hwange np, 2009-2017.

[54] Satinder P Singh and Richard C Yee. An upper bound on the loss from approximate optimal-value functions. *Machine Learning*, 16:227–233, 1994.

[55] Jacob W. Crandall. Towards minimizing disappointment in repeated games. *Journal of Artificial Intelligence Research (JAIR)*, 49:111–142, 2014.

[56] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. The MIT Press, 1992.

Broader Impacts. The implications of our research extend beyond theoretical interests and have practical significance in fields such as economics, cybersecurity, and strategic planning, where decision-making under uncertainty is crucial. By improving the understanding of how predictive information can be used safely and effectively in competitive environments, our work supports the development of more robust strategies in these areas. This can lead to better risk management practices and enhance the ability of systems to make informed decisions even when faced with unreliable or incomplete information. However, there are potential negative impacts, such as the risk of misuse of predictive models that may lead to biased or unfair decisions if the underlying data or assumptions are flawed. Future research could focus on reducing these risks, ensuring AI systems remain reliable and safe.

A Proof of Theorem 3.1

We first consider bounding the opportunity, focusing on the case when the predicted belief contains no error, i.e., $y^\star = \mathbb{E}_\rho[y]$. By the definition of $\pi(\rho)$, for any payoff matrix A with $\mu_\Theta(A) \leq \mu$ and $\nu_\Theta(A) \geq \nu$, we get

$$\begin{aligned}\Delta_{\mathsf{NFG}}(0;\pi) = &\max_{y^\star\in\Theta}\left(\max_{x\in\mathsf{P}_a} x^\top A y^\star - \pi(\rho) A y^\star\right)\\ \leq &(1-\lambda)\left(\max_{x\in\mathsf{P}_a}\max_{y^\star\in\Theta} x^\top A y^\star - \min_{y^\star\in\Theta}\overline{x}^\top A y^\star\right), \qquad (13)\end{aligned}$$

where we have maximized the two terms $\max_{x\in\mathsf{P}_a} x^\top A y^\star$ and $\overline{x}^\top A y^\star$ separately over $y^\star \in \Theta$ to obtain (13). Therefore, noting the definition of the safe strategy $\overline{x}$ in (7),

$$\Delta_{\mathsf{NFG}}(0;\pi) \leq (1-\lambda)\left(\max_{x\in\mathsf{P}_a}\max_{y^\star\in\Theta} x^\top A y^\star - \max_{x\in\mathsf{P}_a}\min_{y^\star\in\Theta} x^\top A y^\star\right) \leq (1-\lambda)(\mu-\nu). \qquad (14)$$

Now, we consider bounding the risk. By the definition of the following mixed strategy $\pi(\rho)$ used by Player 1, for any payoff matrix A with $\mu_\Theta(A) \leq \mu$, $\nu_\Theta(A) \geq \nu$, and $\|A\|_{\max} \leq \alpha$,

$$\begin{aligned}\Delta_{\mathsf{NFG}}(\varepsilon;\pi) = &\max_{d(\rho,y^\star)\leq\varepsilon}\left(\max_{x\in\mathsf{P}_a} x^\top A y^\star - \pi(\rho)^\top A y^\star\right)\\ \leq &\max_{d(\rho,y^\star)\leq\varepsilon}\left(\max_{x\in\mathsf{P}_a} x^\top A\mathbb{E}_\rho[y] - \lambda\widetilde{x}^\top A y^\star\right) - (1-\lambda)\min_{y^\star\in\Theta}\overline{x}^\top A y^\star. \qquad (15)\end{aligned}$$

In (15), we replace $\max_{x\in\mathsf{P}_a} x^\top A y^\star$ by $\max_{x\in\mathsf{P}_a} x^\top A\mathbb{E}_\rho[y]$ since $y^\star \in \Theta$ and there exists a feasible ρ with $\mathbb{E}_\rho[y] = y^\star$ that always satisfies the constraint $d(\rho, y^\star) \leq \varepsilon$.

Since $\widetilde{x}$ is a best response strategy given ρ, it holds that $\max_{x\in\mathsf{P}_a} x^\top A\mathbb{E}_\rho[y] = \widetilde{x}^\top A\mathbb{E}_\rho[y]$. Decomposing $\max_{x\in\mathsf{P}_a} x^\top A y^\star = (1-\lambda)\max_{x\in\mathsf{P}_a} x^\top A y^\star + \lambda\max_{x\in\mathsf{P}_a} x^\top A y^\star$ and maximizing the two terms $(1-\lambda)\max_{x\in\mathsf{P}_a} x^\top A y^\star$ and $\lambda\max_{x\in\mathsf{P}_a} x^\top A y^\star$ over $d(\rho, y^\star) \leq \varepsilon$ respectively, $\Delta(\varepsilon;\alpha,\pi)$ can be further bounded from above by

$$\max_{A:\|A\|_{\max}\leq\alpha}\left((1-\lambda)\max_{x\in\mathsf{P}_a, y^\star\in\Theta} x^\top A y^\star + \lambda\max_{d(\rho,y^\star)\leq\varepsilon}\left(\widetilde{x}^\top A(\mathbb{E}_\rho[y]-y^\star)\right) - (1-\lambda)\nu\right). \qquad (16)$$

Since $d(\rho, y^\star) \coloneqq \|\mathbb{E}_\rho[y] - y^\star\|_1 \leq \varepsilon$, (16) yields

$$\Delta_{\mathsf{NFG}}(\varepsilon;\pi) \leq (1-\lambda)(\mu-\nu) + \lambda\mu\varepsilon. \qquad (17)$$

Maximizing over $\varepsilon \leq \eta(\Theta)$ for (17), we conclude the theorem.

B Proof of Theorem 3.2

We first suppose a mixed strategy $\pi : \mathsf{P}_\Theta \to \mathsf{P}_a$ for Player 1 misses at most $(1-\lambda)(\mu-\nu)$ opportunity, given a belief of types $\rho \in \mathsf{P}_\Theta$.

Let y' and y'' be two mixed strategies in Θ that achieve $\kappa(\Theta) \geq 0$ (select arbitrary strategies if there are multiple to break the tie), i.e.,

$$\kappa(\Theta) = \sum_{i:y_i' \leq y_i''} y_i'' - \sum_{i:y_i' > y_i''} y_i'' \text{ subject to } \sum_{i:y_i' \leq y_i''} y_i' < \sum_{i:y_i' > y_i''} y_i'.$$

Let $\mathcal{I}_+ := \{i \in [b] : y_i' > y_i''\}$ and $\mathcal{I}_- := \{i \in [b] : y_i' \leq y_i''\}$ be two indices of actions with non-negative and positive coordinates of z where $[b] := \{1, \ldots, b\}$. Consider a payoff matrix A_μ with the following form. Suppose without loss of generality that a is even; otherwise, we append an all-zero row to A_μ. Let $\beta := \mu - \nu$. For the i-th column of A_μ, we set it as $(\beta, -\beta, \ldots, -\beta)^\top$ if $i \in \mathcal{I}_+$ and $(-\beta, \beta, \ldots, \beta)^\top$ if $i \in \mathcal{I}_-$. Formally, we define:

$$A = A_\mu = \begin{bmatrix} \cdots & \beta & -\beta & \cdots \\ \cdots & -\beta & \beta & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \underbrace{-\beta}_{i \in \mathcal{I}_+} & \underbrace{\beta}_{i+1 \in \mathcal{I}_-} & \cdots \end{bmatrix} + A_\nu,$$

where $(A_\nu)_{ij} = \nu$ for all $i \in [a]$ and $j \in [b]$. Clearly, with this A, $\mu_\Theta(A) \leq \mu$ and $\nu_\Theta(A) \geq \nu$.

Then, by definition, setting $\varepsilon = 0$ with $\mathbb{E}_\rho[y] = y^\star$ (the type belief contains no error), the payoff gap satisfies

$$\Delta_{\mathsf{NFG}}(0; \pi) = \max_{y^\star \in \Theta} \left(\max_{x \in \mathsf{P}_a} x^\top A y^\star - \pi(\rho) A y^\star \right) \leq (1 - \lambda)(\mu - \nu). \tag{18}$$

Now, plugging in y' into (18) above. This implies

$$\max_{y^\star \in \Theta} \left(\max_{x \in \mathsf{P}_a} x^\top A y^\star - \pi(\rho)^\top A y^\star \right) \geq \max_{x \in \mathsf{P}_a} x^\top A y' - \pi(\rho)^\top A y' = \mu - \pi(\rho)^\top A y'. \tag{19}$$

For notational convenience, denote $\digamma(\Theta) := \sum_{i:y_i' > y_i''} y_i' - \sum_{i:y_i' \leq y_i''} y_i'$. Combining (18) and (19), $\pi(\rho)^\top A y' = \pi(\rho)^\top (A_\mu + A_\nu) y' \geq (1 - \lambda)\nu + \lambda\mu$. Thus, $\pi(\rho)^\top A_\mu y' \geq \lambda(\mu - \nu)$. Equivalently, $\digamma(\Theta)\,(\mu - \nu)\,\pi(\rho)^\top f \geq \lambda\,(\mu - \nu)\,$, where $f := (1, -1, 1, \ldots, 1, -1)^\top$. By the construction of A, we get

$$\pi(\rho)^\top A_\mu y'' = -\left(\pi(\rho)^\top f\right)(\mu - \nu) \cdot \left(\sum_{i:y_i' \leq y_i''} y_i'' - \sum_{i:y_i' > y_i''} y_i'' \right) \leq -\lambda\,(\mu - \nu) \cdot \frac{\kappa(\Theta)}{\digamma(\Theta)}.$$

Therefore, since $0 < \digamma(\Theta) \leq 1$,

$$\pi(\rho)^\top (A_\mu + A_\nu) y'' = \nu + \pi(\rho)^\top A_\mu y'' \leq \nu - \lambda\,(\mu - \nu)\,\kappa(\Theta) = -\lambda\mu\kappa(\Theta) + (1 + \lambda\kappa(\Theta))\nu.$$

Hence, we get

$$\begin{aligned} \max_{y^\star \in \Theta} \left(\max_{x \in \mathsf{P}_a} x^\top A y^\star - \pi(\rho)^\top A y^\star \right) &\geq \max_{x \in \mathsf{P}_a} x^\top A y'' - \pi(\rho)^\top A y'' \\ &\geq \kappa(\Theta)\,(1 + \lambda)\,\mu - (1 + \lambda\kappa(\Theta))\nu \\ &\geq (\kappa(\Theta)\mu - \nu)\,(1 + \lambda)\,. \end{aligned}$$

C Proof of Corollary 3.1

If $\Theta = \mathsf{P}_b$, then $\kappa(\Theta) = 1$. Furthermore, by definition, we get

$$\mu_A\left(\mathsf{P}_b\right) := \max_{y \in \mathsf{P}_b} \max_{x \in \mathsf{P}_a} \left| x^\top A y \right| = \|A\|_{\max}, \quad \nu_A\left(\mathsf{P}_b\right) := \min_{y \in \mathsf{P}_b} \max_{x \in \mathsf{P}_a} x^\top A y = 0,$$

where the equality on the RHS holds since the game is fair. Applying Theorem 3.1 and 3.2, we obtain the corollary.

D Proof of Theorem 4.1

Given type beliefs, we denote strategies $\widetilde{\sigma} := \sigma(\theta)$, $\sigma := \sigma(\theta^{\star})$, and let $\overline{\sigma} = \sigma(\overline{\theta})$ be the safe strategies of the opponents, with $\overline{\theta}$ being an optimal solution of the minimax optimization in (11). Denoting $\widetilde{V} := V^{\star,\widetilde{\sigma}}$, $\overline{V} := V^{\star,\overline{\sigma}}$, $V^{\star} := V^{\star,\sigma}$ the value functions with the opponents' strategies being $\widetilde{\sigma}$, $\overline{\sigma}$, and σ, respectively. Given the strategy π defined in (12), we first state two useful bounds.

Lemma 1. *Given the strategy defined in* (12)*, the following bounds on the reward function hold:*

$$\begin{aligned}\textsc{Bound 1}: \quad \mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi,\sigma}[r] \leq\; & \gamma\lambda\left(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[V^{\star}(s)\big] - \mathbb{E}_{p,\pi^{\star},\widetilde{\sigma}}\big[V^{\star}(s)\big]\right) + 2\lambda\left(\varepsilon r_{\max} + \gamma\eta\right)\\ & + \gamma(1-\lambda)\left(\mathbb{E}_{p,\pi,\overline{\sigma}}\big[\overline{V}(s)\big] - \mathbb{E}_{p,\pi^{\star},\overline{\sigma}}\big[\overline{V}(s)\big]\right)\\ & + (1-\lambda)\left(\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi^{\star},\overline{\sigma}}[r]\right) + (1-\lambda)\left(\mathbb{E}_{\pi,\overline{\sigma}}[r] - \mathbb{E}_{\pi,\sigma}[r]\right), \qquad (20)\end{aligned}$$

where $\eta := \max_{s\in\mathcal{S}}\left|V^{\star}(s) - \widetilde{V}(s)\right|$*, and*

$$\begin{aligned}\textsc{Bound 2}: \; \mathbb{E}_{\overline{\pi},\overline{\sigma}}[r] - \mathbb{E}_{\pi,\overline{\sigma}}[r] \leq\; & \gamma\left(\mathbb{E}_{p,\pi,\overline{\sigma}}\big[\overline{V}(s)\big] - \mathbb{E}_{p,\overline{\pi},\overline{\sigma}}\big[\overline{V}(s)\big]\right)\\ & + \frac{\lambda}{1-\lambda}\left(\left(\mathbb{E}_{\pi,\widetilde{\sigma}}[r] - \mathbb{E}_{\overline{\pi},\widetilde{\sigma}}[r]\right) + \gamma\left(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[\widetilde{V}(s)\big] - \mathbb{E}_{p,\overline{\pi},\widetilde{\sigma}}\big[\widetilde{V}(s)\big]\right)\right). \qquad (21)\end{aligned}$$

Proof of Lemma 1. First, we prove Bound 1 in (20).

Based on the definition of the policy π in (12), it follows that for any fixed state $s \in \mathcal{S}$,

$$\begin{aligned}&\lambda\mathbb{E}_{\pi^{\star},\widetilde{\sigma}}[r] + (1-\lambda)\mathbb{E}_{\pi^{\star},\overline{\sigma}}[r] + \gamma\lambda\mathbb{E}_{p,\pi^{\star},\widetilde{\sigma}}\big[\widetilde{V}(s)\big] + \gamma(1-\lambda)\mathbb{E}_{p,\pi^{\star},\overline{\sigma}}\big[\overline{V}(s)\big]\\ \leq & \lambda\mathbb{E}_{\pi,\widetilde{\sigma}}[r] + (1-\lambda)\mathbb{E}_{\pi,\overline{\sigma}}[r] + \gamma\lambda\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[\widetilde{V}(s)\big] + \gamma(1-\lambda)\mathbb{E}_{p,\pi,\overline{\sigma}}\big[\overline{V}(s)\big]. \qquad (22)\end{aligned}$$

Furthermore, considering the difference between the expected rewards corresponding to implementing $\pi^{\star}$ and the strategy π defined in (12), since by Assumption 1, $|r| \leq r_{\max}$ for all $t \in [T]$ and any strategy $\pi \in \Pi$,

$$\left|\mathbb{E}_{\pi,\sigma}[r] - \mathbb{E}_{\pi,\widetilde{\sigma}}[r]\right| \leq \max_{s\in\mathcal{S}}\left\|\sigma(s) - \widetilde{\sigma}(s)\right\|_1 r_{\max} \leq \varepsilon r_{\max}. \qquad (23)$$

Therefore, using (23),

$$\begin{aligned}\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi,\sigma}[r] =& \lambda\left(\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi,\sigma}[r]\right) + (1-\lambda)\left(\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi,\sigma}[r]\right)\\ \leq & \lambda\left(\mathbb{E}_{\pi^{\star},\widetilde{\sigma}}[r] - \mathbb{E}_{\pi,\widetilde{\sigma}}[r]\right) + (1-\lambda)\left(\mathbb{E}_{\pi^{\star},\overline{\sigma}}[r] - \mathbb{E}_{\pi,\overline{\sigma}}[r]\right) + 2\lambda\varepsilon r_{\max}\\ & + (1-\lambda)\left(\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi^{\star},\overline{\sigma}}[r]\right)\\ & + (1-\lambda)\left(\mathbb{E}_{\pi,\overline{\sigma}}[r] - \mathbb{E}_{\pi,\sigma}[r]\right)\\ =& \left(\lambda\mathbb{E}_{\pi^{\star},\widetilde{\sigma}}[r] + (1-\lambda)\mathbb{E}_{\pi^{\star},\overline{\sigma}}[r]\right) - \left(\lambda\mathbb{E}_{\pi,\widetilde{\sigma}}[r] + (1-\lambda)\mathbb{E}_{\pi,\overline{\sigma}}[r]\right)\\ & + (1-\lambda)\left(\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi^{\star},\overline{\sigma}}[r]\right)\\ & + (1-\lambda)\left(\mathbb{E}_{\pi,\overline{\sigma}}[r] - \mathbb{E}_{\pi,\sigma}[r]\right) + 2\lambda\varepsilon r_{\max}. \qquad (24)\end{aligned}$$

Combining (22) and (24), we obtain

$$\begin{aligned}&\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi,\sigma}[r] - 2\lambda\varepsilon r_{\max}\\ \leq & \gamma\left(\lambda\left(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[\widetilde{V}(s)\big] - \mathbb{E}_{p,\pi^{\star},\widetilde{\sigma}}\big[\widetilde{V}(s)\big]\right) + (1-\lambda)\left(\mathbb{E}_{p,\pi,\overline{\sigma}}\big[\overline{V}(s)\big] - \mathbb{E}_{p,\pi^{\star},\overline{\sigma}}\big[\overline{V}(s)\big]\right)\right)\\ & + (1-\lambda)\left(\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi^{\star},\overline{\sigma}}[r]\right) + (1-\lambda)\left(\mathbb{E}_{\pi,\overline{\sigma}}[r] - \mathbb{E}_{\pi,\sigma}[r]\right). \qquad (25)\end{aligned}$$

By definition,

$$\left|\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[V^{\star}(s)\big] - \mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[\widetilde{V}(s)\big]\right| \leq \sum_{s'\in\mathcal{S}}\sum_{a,a'\in\mathcal{A}} p(s'|s,(a,a'))\pi_t(s,a)\widetilde{\sigma}_t(s,a')\eta \leq \eta,$$

and similarly, $\left|\mathbb{E}_{p,\pi^{\star},\widetilde{\sigma}}\big[V^{\star}(s)\big] - \mathbb{E}_{p,\pi^{\star},\widetilde{\sigma}}\big[\widetilde{V}(s)\big]\right| \leq \eta$, thus, (25) yields

$$\begin{aligned}\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi,\sigma}[r] \leq & \gamma\lambda\left(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[V^{\star}(s)\big] - \mathbb{E}_{p,\pi^{\star},\widetilde{\sigma}}\big[V^{\star}(s)\big]\right) + 2\lambda\left(\varepsilon r_{\max} + \gamma\eta\right)\\ & + \gamma(1-\lambda)\left(\mathbb{E}_{p,\pi,\overline{\sigma}}\big[\overline{V}(s)\big] - \mathbb{E}_{p,\pi^{\star},\overline{\sigma}}\big[\overline{V}(s)\big]\right)\\ & + (1-\lambda)\left(\mathbb{E}_{\pi^{\star},\sigma}[r] - \mathbb{E}_{\pi^{\star},\overline{\sigma}}[r]\right) + (1-\lambda)\left(\mathbb{E}_{\pi,\overline{\sigma}}[r] - \mathbb{E}_{\pi,\sigma}[r]\right).\end{aligned}$$

Next, we show BOUND 2 in (21).

By the definition of the policy π in (12), it follows that for any fixed state $s \in \mathcal{S}$, similarly we obtain

$$\begin{aligned}&\lambda\mathbb{E}_{\overline{\pi},\widetilde{\sigma}}\left[r\right] + (1-\lambda)\mathbb{E}_{\overline{\pi},\overline{\sigma}}\left[r\right] + \gamma\lambda\mathbb{E}_{p,\overline{\pi},\widetilde{\sigma}}\big[\widetilde{V}(s)\big] + \gamma(1-\lambda)\mathbb{E}_{p,\overline{\pi},\overline{\sigma}}\big[\overline{V}(s)\big]\\ \leq&\lambda\mathbb{E}_{\pi,\widetilde{\sigma}}\left[r\right] + (1-\lambda)\mathbb{E}_{\pi,\overline{\sigma}}\left[r\right] + \gamma\lambda\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[\widetilde{V}(s)\big] + \gamma(1-\lambda)\mathbb{E}_{p,\pi,\overline{\sigma}}\big[\overline{V}(s)\big].\end{aligned} \tag{26}$$

Rearranging the terms in (26),

$$\begin{aligned}\mathbb{E}_{\overline{\pi},\overline{\sigma}}\left[r\right] - \mathbb{E}_{\pi,\overline{\sigma}}\left[r\right] \leq& \gamma\left(\mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s)\right] - \mathbb{E}_{p,\overline{\pi},\overline{\sigma}}\left[\overline{V}(s)\right]\right)\\ &+\frac{\lambda}{1-\lambda}\Big(\left(\mathbb{E}_{\pi,\widetilde{\sigma}}\left[r\right] - \mathbb{E}_{\overline{\pi},\widetilde{\sigma}}\left[r\right]\right) + \gamma\left(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[\widetilde{V}(s)\big] - \mathbb{E}_{p,\overline{\pi},\widetilde{\sigma}}\big[\widetilde{V}(s)\big]\right)\Big).\end{aligned}$$

□

Using Lemma 1, the following lemma provides a bound on $\Delta_{\mathsf{SBG}}(\varepsilon;\pi)$, from which bounds on both risk and (missed) opportunity can be directly derived.

Lemma 2. *The payoff gap for the policy* π *defined in* (12) *satisfies*

$$\Delta_{\mathsf{SBG}}(\varepsilon;\pi) \leq \frac{r_{\max}}{1-\lambda\gamma}\left(\frac{2\lambda\varepsilon}{1-\gamma} + \min\left\{\frac{2(1-\lambda)}{1-\gamma}, \frac{2\lambda}{(1-\gamma)^2}\right\} + \frac{(1-\lambda)(2-\gamma)}{1-\gamma}\right) + \frac{\gamma\left(2\lambda\eta - \nu\right)}{1-\lambda\gamma}.$$

Proof. By definition of the payoff gap (see Definition 2) and the value function in (9), we get

$$\Delta_{\mathsf{SBG}}(\varepsilon;\pi) = \sup_{d(\theta,\theta^\star)\leq\varepsilon} \max_{s\in\mathcal{S}} \Big(V^{\star,\sigma}(s) - V^{\pi(\theta),\sigma}(s)\Big).$$

Our first goal is to derive an upper bound on $\Delta_{\mathsf{SBG}}(\varepsilon;\pi)$, aligning with the classic loss bound of approximate value functions [54]. However, the policy considered in our context is not directly maximizing the approximate value function, but a mixed strategy defined by an optimization as in (12). For notational convenience, we write for $s \in \mathcal{S}$,

$$\mathbb{E}_{p,\pi,\sigma}\left[V(s)\right] \coloneqq \sum_{a,a'\in\mathcal{A}}\sum_{s'\in\mathcal{S}} p(s'|s,(a,a'))\pi(s,a)\sigma(s,a')V(s').$$

For notational simplicity, we denote $V^\pi(s) \coloneqq V^{\pi,\sigma}(s)$. For any state $s \in \mathcal{S}$ and strategies $(\sigma,\widetilde{\sigma})$ satisfying $d\left(\theta,\theta^\star\right) = \max_{s\in\mathcal{S}}\left\|\sigma(s;\theta) - \sigma(s;\theta^\star)\right\|_1 \leq \varepsilon$, applying BOUND 1 in Lemma 1, the Bellman equations corresponding to $\pi^\star$ and π imply

$$\begin{aligned}V^\star(s) - V^\pi(s) =&\mathbb{E}_{\pi^\star,\sigma}[r] - \mathbb{E}_{\pi,\sigma}[r] + \gamma\left(\mathbb{E}_{p,\pi^\star,\sigma}\big[V^\star(s)\big] - \mathbb{E}_{p,\pi,\sigma}\big[V^\pi(s)\big]\right)\\ \leq&\,\lambda\gamma\Big(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[V^\star(s)\big] - \mathbb{E}_{p,\pi^\star,\widetilde{\sigma}}\big[V^\star(s)\big] + \mathbb{E}_{p,\pi^\star,\sigma}\big[V^\star(s)\big] - \mathbb{E}_{p,\pi,\sigma}\big[V^\pi(s)\big]\Big)\\ &+\gamma(1-\lambda)\left(\mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s)\right] - \mathbb{E}_{p,\pi,\sigma}\big[V^\pi(s)\big] + \mathbb{E}_{p,\pi^\star,\sigma}\big[V^\star(s)\big] - \mathbb{E}_{p,\pi^\star,\overline{\sigma}}\big[\overline{V}(s)\big]\right)\\ &+(1-\lambda)\left(\mathbb{E}_{\pi,\overline{\sigma}}\left[r\right] - \mathbb{E}_{\pi,\sigma}\left[r\right]\right) + (1-\lambda)\left(\mathbb{E}_{\pi^\star,\sigma}\left[r\right] - \mathbb{E}_{\pi^\star,\overline{\sigma}}\left[r\right]\right)\\ &+2\lambda\left(\varepsilon r_{\max} + \gamma\eta\right).\end{aligned} \tag{27}$$

Applying the Bellman equations, for any $s \in \mathcal{S}$,

$$\begin{aligned}V^\pi(s) =&\mathbb{E}_{a\sim\pi(s),a'\sim\sigma}(s)\left[\left(r+\gamma\mathbb{P}V^\pi\right)(s,(a,a'))\right]\\ =&\mathbb{E}_{\pi,\sigma}\left[r\right] + \gamma\mathbb{E}_{p,\pi,\sigma}\left[V^\pi(s)\right].\end{aligned} \tag{28}$$

Similarly, it also holds that

$$V^\star(s) = \mathbb{E}_{\pi^\star,\sigma}\left[r\right] + \gamma\mathbb{E}_{p,\pi^\star,\sigma}\left[V^\star(s)\right]. \tag{29}$$

Plugging (28), and (29) into (27),

$$\begin{aligned}V^\star(s) - V^\pi(s) \leq&\lambda\gamma\Big(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\big[V^\star(s)\big] - \mathbb{E}_{p,\pi^\star,\widetilde{\sigma}}\big[V^\star(s)\big] + \mathbb{E}_{p,\pi^\star,\sigma}\big[V^\star(s)\big] - \mathbb{E}_{p,\pi,\sigma}\big[V^\pi(s)\big]\Big)\\ &+(1-\lambda)\Big(\,\Big(\left(\mathbb{E}_{\pi,\overline{\sigma}}\left[r\right] + \gamma\mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s)\right]\right) - V^\pi(s)\Big)\,\Big) &(30)\\ &+(1-\lambda)\Big(V^\star(s) - \left(\mathbb{E}_{\pi^\star,\overline{\sigma}}\left[r\right] + \gamma\mathbb{E}_{p,\pi^\star,\overline{\sigma}}\left[\overline{V}(s)\right]\right)\,\Big) &(31)\\ &+2\lambda\left(\varepsilon r_{\max} + \gamma\eta\right). &(32)\end{aligned}$$

Let $\overline{V}^{\pi}(s) := V^{\pi,\overline{\sigma}}(s)$ for any $s \in \mathcal{S}$. For the term in (30), we always have $V^{\pi}(s) \geq \overline{V}^{\pi}(s)$ (c.f. [48]), thus,

$$\begin{aligned}\left(\mathbb{E}_{\pi,\overline{\sigma}}\left[r\right] + \gamma\mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s)\right]\right) - V^{\pi}(s) \leq & \gamma\left(\mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s)\right] - \mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}^{\pi}(s)\right]\right)\\ = & \gamma\mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s) - \overline{V}^{\pi}(s)\right].\end{aligned}$$

Furthermore, the Bellman equation implies

$$\overline{V}(s) - \overline{V}^{\pi}(s) = \left(\mathbb{E}_{\overline{\pi},\overline{\sigma}}\left[r\right] - \mathbb{E}_{\pi,\overline{\sigma}}\left[r\right]\right) + \gamma\left(\mathbb{E}_{p,\overline{\pi},\overline{\sigma}}\left[\overline{V}(s)\right] - \mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}^{\pi}(s)\right]\right). \tag{33}$$

Applying BOUND 2 in Lemma 1,

$$\begin{aligned}\overline{V}(s) - \overline{V}^{\pi}(s) \leq & \gamma\left(\mathbb{E}_{p,\overline{\pi},\overline{\sigma}}\left[\overline{V}(s)\right] - \mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}^{\pi}(s)\right] + \mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s)\right] - \mathbb{E}_{p,\overline{\pi},\overline{\sigma}}\left[\overline{V}(s)\right]\right)\\ & + \frac{\lambda}{1-\lambda}\left(\left(\mathbb{E}_{\pi,\widetilde{\sigma}}\left[r\right] - \mathbb{E}_{\overline{\pi},\widetilde{\sigma}}\left[r\right]\right) + \gamma\left(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\left[\widetilde{V}(s)\right] - \mathbb{E}_{p,\overline{\pi},\widetilde{\sigma}}\left[\widetilde{V}(s)\right]\right)\right)\\ \leq & \gamma\left(\mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s)\right] - \mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}^{\pi}(s)\right]\right)\\ & + \frac{2\lambda}{1-\lambda}\left(\left(1 + \frac{\gamma}{1-\gamma}\right)\right) r_{\max}.\end{aligned} \tag{34}$$

Since above holds for any state $s \in \mathcal{S}$, let s' be a state such that the value gap $V^{\star}(s) - V^{\pi}(s)$ is maximized. Therefore, (34) implies

$$\overline{V}(s) - \overline{V}^{\pi}(s) \leq \frac{2}{(1-\gamma)^2}\frac{\lambda}{1-\lambda} r_{\max}. \tag{35}$$

Then, for the term in (31), by definition $V^{\star}(s) \leq r_{\max}/(1-\gamma)$, and

$$\mathbb{E}_{\pi^{\star},\overline{\sigma}}\left[r\right] + \gamma\mathbb{E}_{p,\pi^{\star},\overline{\sigma}}\left[\overline{V}(s)\right] \geq \gamma\nu - r_{\max}. \tag{36}$$

Combing the bounds in (35) and (36) with (32), we conclude that

$$\begin{aligned}V^{\star}(s) - V^{\pi}(s) \leq & \lambda\gamma\Big(\mathbb{E}_{p,\pi,\widetilde{\sigma}}\left[V^{\star}(s)\right] - \mathbb{E}_{p,\pi^{\star},\widetilde{\sigma}}\left[V^{\star}(s)\right] + \mathbb{E}_{p,\pi^{\star},\sigma}\left[V^{\star}(s)\right] - \mathbb{E}_{p,\pi,\sigma}\left[V^{\pi}(s)\right]\Big)\\ & + \frac{2}{(1-\gamma)^2}\lambda r_{\max} + (1-\lambda)\left(\frac{2-\gamma}{1-\gamma} r_{\max} - \gamma\nu\right) + 2\lambda\left(\varepsilon r_{\max} + \gamma\eta\right).\end{aligned} \tag{37}$$

Now, noting that by definition we have

$$\begin{aligned}&\left|\mathbb{E}_{p,\pi,\sigma}\left[V^{\star}\right] - \mathbb{E}_{p,\pi,\widetilde{\sigma}}\left[V^{\star}\right]\right|\\ \leq & \sum_{a,a'\in\mathcal{A},s'\in\mathcal{S}} p(s'|s,(a,a'))\pi(s,a)\left|\sigma(s,a') - \widetilde{\sigma}(s,a')\right| V^{\star}(s') \leq \varepsilon\frac{r_{\max}}{1-\gamma},\end{aligned}$$

and similarly,

$$\left|\mathbb{E}_{p,\pi^{\star},\sigma}\left[V^{\star}\right] - \mathbb{E}_{p,\pi^{\star},\widetilde{\sigma}}\left[V^{\star}\right]\right| \leq \varepsilon\frac{r_{\max}}{1-\gamma}.$$

Therefore, rearranging the terms and simplifying above, (37) becomes

$$\begin{aligned}V^{\star}(s) - V^{\pi}(s) \leq & \frac{2\lambda}{(1-\gamma)^2} r_{\max} + (1-\lambda)\left(\frac{2-\gamma}{1-\gamma} r_{\max} - \gamma\nu\right) + 2\lambda\left(\varepsilon r_{\max} + \gamma\eta\right)\\ & + \frac{2\gamma\lambda r_{\max}}{1-\gamma}\varepsilon + \gamma\lambda\left(\mathbb{E}_{p,\pi,\sigma}\left[V^{\star}\right] - \mathbb{E}_{p,\pi,\sigma}\left[V^{\pi}\right]\right).\end{aligned} \tag{38}$$

Since above holds for any state $s \in \mathcal{S}$, let s^{*} be a state such that the value gap $V^{\star}(s) - V^{\pi}(s)$ is maximized. Thus,

$$\begin{aligned}&\mathbb{E}_{p,\pi,\sigma}\left[V^{\star}(s^{*})\right] - \mathbb{E}_{p,\pi,\sigma}\left[V^{\pi}(s^{*})\right]\\ = & \sum_{a,a'\in\mathcal{A}}\sum_{s'\in\mathcal{S}} p\left(s'|s^{*},(a,a')\right)\pi\left(s^{*},a\right)\sigma\left(s^{*},a'\right)\left(V^{\star}(s') - V^{\pi}(s')\right) \leq V^{\star}(s^{*}) - V^{\pi}(s^{*}).\end{aligned}$$

Continuing from (38),

$$\begin{aligned}&(V^\star(s^*) - V^\pi(s^*)) - \lambda\gamma\left(V^\star(s^*) - V^\pi(s^*)\right)\\ \le& 2r_{\max}\left(\frac{\lambda}{(1-\gamma)^2} + \frac{\gamma\lambda}{1-\gamma}\varepsilon + \lambda\varepsilon\right) + 2\lambda\gamma\eta + (1-\lambda)\left(\frac{2-\gamma}{1-\gamma}r_{\max} - \gamma\nu\right)\\ =& 2r_{\max}\left(\frac{\lambda}{(1-\gamma)^2} + \frac{\lambda\varepsilon}{1-\gamma}\right) + 2\lambda\gamma\eta + (1-\lambda)\left(\frac{2-\gamma}{1-\gamma}r_{\max} - \gamma\nu\right).\end{aligned}$$

Therefore, rearranging the terms we obtain

$$\Delta_{\mathsf{SBG}}(\varepsilon;\pi) \le \frac{r_{\max}}{1-\lambda\gamma}\left(\frac{2\lambda\varepsilon}{1-\gamma} + \frac{2\lambda}{(1-\gamma)^2} + \frac{(1-\lambda)(2-\gamma)}{1-\gamma}\right) + \frac{\gamma\left(2\lambda\eta - (1-\lambda)\nu\right)}{1-\lambda\gamma}.$$

Now, the terms in (30) can be bounded alternatively as

$$(1-\lambda)\Big(\left(\left(\mathbb{E}_{\pi,\overline{\sigma}}\left[r\right] + \gamma\mathbb{E}_{p,\pi,\overline{\sigma}}\left[\overline{V}(s)\right]\right) - V^\pi(s)\right)\Big) \le 2\frac{1-\lambda}{1-\gamma}r_{\max}.$$

Using this and following the same steps, we obtain

$$\Delta_{\mathsf{SBG}}(\varepsilon;\pi) \le \frac{r_{\max}}{1-\lambda\gamma}\left(\frac{2\lambda\varepsilon}{1-\gamma} + \frac{2(1-\lambda)}{1-\gamma} + \frac{(1-\lambda)(2-\gamma)}{1-\gamma}\right) + \frac{\gamma\left(2\lambda\eta - (1-\lambda)\nu\right)}{1-\lambda\gamma}.$$

□

Part I: Proof of Risk Bound

Since $d\left(\theta,\theta^\star\right) \coloneqq \max_{s\in\mathcal{S}}\left\|\sigma(s;\theta) - \sigma(s;\theta^\star)\right\|_1 \le 2$, setting the worst $\varepsilon = 2$, and noticing that $\eta \le \frac{2r_{\max}}{1-\gamma}$,

$$\Delta_{\mathsf{SBG}}(\varepsilon;\pi) \le \frac{r_{\max}}{1-\lambda\gamma}\left(\frac{\left(4(1-\gamma^2)+2\right)\lambda}{(1-\gamma)^2} + \frac{(1-\lambda)(2-\gamma)}{1-\gamma}\right) - \frac{\gamma(1-\lambda)\nu}{1-\lambda\gamma}.$$

Rearranging above leads to the result.

Part II: Proof of (Missed) Opportunity Bound

Finally, setting $\varepsilon = \eta = 0$, we obtain

$$\Delta_{\mathsf{SBG}}(0;\pi) \le \frac{1}{1-\lambda\gamma}\left(r_{\max}\left(\frac{\gamma^2 - 3\gamma + 6}{(1-\gamma)^2}\right) - \gamma\nu\right)(1-\lambda).$$

E Proof of Theorem 4.2

We prove the theorem by considering a stateless MDP, which reduces a stochastic Bayesian game $\mathcal{M} = \langle\mathcal{S},\mathcal{A},\Theta,\sigma,p,r,\gamma\rangle$ to a normal-form game in Theorem 3.2.

Let $\mathcal{L} \coloneqq \left\{1,\ldots,\prod_{j\neq i}|\mathcal{A}(j)|\right\}$. We fix a strategy kernel σ such that

$$\kappa(\Theta) \coloneqq \max_{s\in\mathcal{S},\theta,\theta'\in\Theta}\left(\sum_{l\in\mathcal{L}:\sigma_l(s;\theta)=0}\sigma_l(s;\theta') - \sum_{l\in\mathcal{L}:\sigma_l(s;\theta)>0}\sigma_l(s;\theta')\right) = 1.$$

Therefore, the construction of the payoff matrix $A \in \mathbb{R}^{|\mathcal{A}(i)|\times\prod_{j\neq i}|\mathcal{A}(j)|}$ in the proof of Theorem 3.2 with $\mu_\Theta(A) \le \frac{r_{\max}}{1-\gamma}$ and $\nu_\Theta(A) \ge \nu$ yields that for any $0 \le \lambda \le 1$, if any mixed strategy π misses $\frac{1}{1-\lambda\gamma}\left(r_{\max} - \nu\right)(1-\lambda)$ opportunity, then since

$$\frac{1}{1-\lambda\gamma}\left(r_{\max} - \nu\right)(1-\lambda) \le \left(\frac{r_{\max}}{1-\gamma} - \nu\right)(1-\lambda),$$

the strategy π has at least $\left(\frac{r_{\max}}{1-\gamma} - \nu\right)(1+\lambda)$ risk. Furthermore, since

$$\left(\frac{r_{\max}}{1-\gamma} - \nu\right)(1+\lambda) \ge \left(\frac{r_{\max}}{1-\lambda\gamma} - \frac{\nu}{1-\lambda\gamma}\right)(1+\lambda),$$

π has at least $\frac{1}{1-\lambda\gamma}\left(r_{\max} - \nu\right)(1+\lambda)$ risk.

F Details on Experiments

F.1 Details on 2×2 Game simulations

F.1.1 Game definition

- We define a Stochastic Bayesian game with one state: a particular 2×2 game sampled from the topology of 2×2 games.
- The payoffs for each player are stipulated by the 2×2 game. We set the time horizon for the game be 1,000 giving us an empirical evaluation of the expected payoff for each strategy

F.1.2 Type definitions

As mentioned we have 3 classes of types and below we provide a description of the general classes as well as the specific types we used in our evaluation.

1. **Markovian Types:** These are types whose strategy only depends on the current state. We made use of 4 types of Markovian strategies
   - **Type 1:** Always play action 0
   - **Type 2:** Always play action 1
   - **Type 3:** The minimax strategy for the player
   - **Type 4:** Returns a random strategy
2. **Leader- Follower-Trigger Agents:** Inspired by [55], these agents have a preferred sequence of play they seek to enforce. Importantly they have access to history of play and when the other player does not play according to their preferred sequence the alter their strategy by engaging in “punishing” behavior such as playing the minimax strategy or by resetting to a previous action. In our case, we evaluated against simple versions of such agents, wherein our agent had a preferred mixed strategy and when the empirical observed strategy over a preset number of previous plays from the opponent did not match their preferred strategy, they “punished” the opponent by playing a minimax strategy. In particular, we had an agent look at the previous 4 actions of the opponent and if they selected action 1 more than twice, they chose to play a minimax strategy for the next round.
3. **Co-evolved Neural Networks:** Inspired by work in [51] we use ideas of genetic programming [56], to generate agents from neural networks. We randomly initialized 10 neural networks with a single hidden layer for both the row and column player. All networks would take as input the previous 4 actions of both players and the corresponding state information. We sample randomly from populations of the row and column players and simulate the game. After simulation we calculate a fitness score based on average payoffs for each agent and a similarity score so as to ensure diversity in the models. We “evolve” the populations by selecting random portions of both populations to mutate whilst also having cross-over between members of the populations selected by fitness (this is done for both populations hence “co-evolve”). We then proceed to create new populations using the most fit agents from a previous generation (i.e., we take the top 50% of a population pre-evolution and 50% post evolution and then create a new population if the average fitness of this new constructed population is greater than the average fitness of the previous population.)

F.1.3 Experimental illustration of tightness of bounds

We have included additional adversarial examples to demonstrate the tightness of the bounds presented in Theorems 3.1 and 3.2. In particular, we consider the zero-sum Matching Pennies (MP) and an Adjusted Matching Pennies (AMP) as our two examples. We assume the hypothesis set Θ for Player 2 contains 6 behavioral types:

- **Markovian Types**:
  - **Type 1**: Always play action 0
  - **Type 2**: Always play action 1
  - **Type 3**: Minimax strategy
  - **Type 4**: Random strategy

- **Type 5**: Leader-Follower-Trigger Agents
- **Type 6**: Co-evolved Neural Networks

The payoff matrices for the MP and AMP, respectively are

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \text{and } \begin{bmatrix} 1.2 & -0.8 \\ -0.8 & 1.2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{bmatrix}.$$

It's worth mentioning that the AMP payoff matrix constructed above is an adversarial example that follows the same construction of the adversarial payoff matrix A used in the proof of Theorem 3.2. Given the considered Θ, we know that $\kappa(\Theta) = 1$ and $\eta(\Theta) = 2$. The upper and lower bounds in Theorem 3.1 and 3.2 therefore read:

$$\begin{aligned} &\textit{Upper Bound: } x = (1-\lambda)(\mu-\nu),\ y = (1-\lambda)(\mu-\nu) + 2\lambda\mu, \lambda \in [0,1], \\ &\textit{Lower Bound: } x = (1-\lambda)(\mu-\nu),\ y = (1+\lambda)(\mu-\nu), \lambda \in [0,1], \end{aligned}$$

where for MP, $\mu = 1$, $\nu = 0$; for AMP, $\mu = 1.2$, $\nu = 0.2$.

Besides the bounds above, we also plot simulated risk and opportunity values in the figures using an algorithm that has varying trust of type beliefs in 100 and 1,000 runs. The algorithm is presented with a type prediction and takes a convex combination of the best response to the type prediction and the minimax strategy with the trust parameter λ determining how much to weigh the best response. It is the same as π, the mixed strategy used to prove Theorem 3.1. Fully trusting ($\lambda = 1$) and distrusting ($\lambda = 0$) type beliefs yield a best response strategy and a minimax strategy correspondingly. The agent then samples from their resulting mixed strategy an action to play while the opponent also samples from whatever mixed strategy they are using an action to play.

We sample this interaction for the number of runs and gather empirical payoff information, which we use to plot the opportunity risk tradeoff. The variance of the risk and opportunity values decreases as the number of runs increases, as illustrated by the error bars in the attached figures. Moreover, we include as plots the bounds on the opportunity risk tradeoff and show the tightness or looseness of these bounds in two games we pick. Please note that some of the simulated risk and opportunity values fall outside the bounds. This is because the bounds are applicable only to expected payoff gaps, and individual simulations may deviate from these expected values.

In the Fig 7, the top two figures display the results from 100 runs, while the bottom two figures present the results from 1,000 runs. The left two figures correspond to the MP example, while the right two figures are for the AMP example, demonstrating the gap between the lower and upper bounds. In particular, in an AMP game, we see that the upper bound is loose, and the empirical opportunity-risk tradeoff matches the lower bound we derive. In future work, we will investigate if this gap can be closed. This 'adversarial' example illustrates the looseness in the upper bound we derive. In the canonical Matching Pennies game, we find that the lower and upper bounds are tight with the empirical tradeoff matching these theoretical bounds, coinciding with Corollary 3.1.

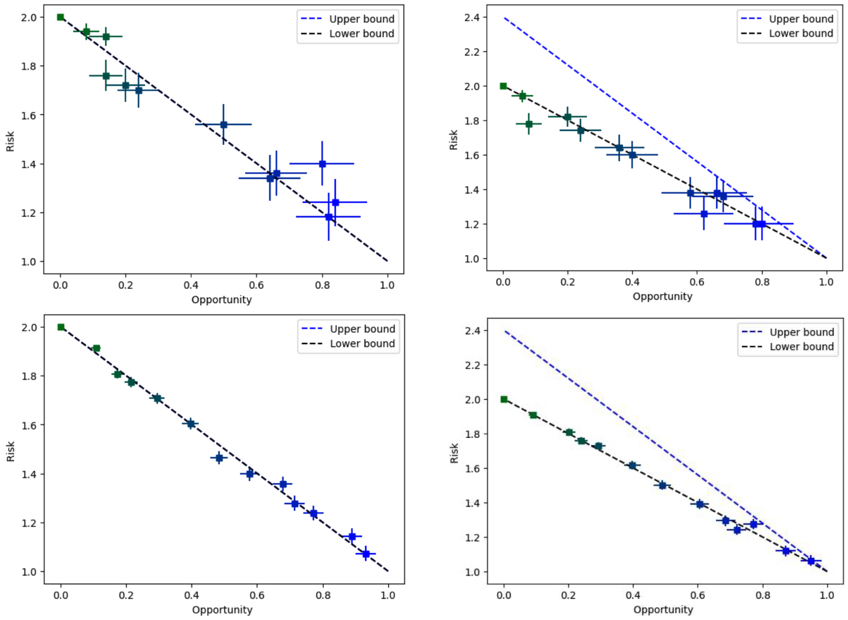


Figure 7: Opportunity-risk tradeoff for Matching Pennies (MP) and Adjusted Matching Pennies (AMP) games over 100 (top) and 1,000 (bottom) runs. **Left**: Matching Pennis. **Right**: Adjusted Matching Pennies. As the number of runs increases, the variance of both risk and opportunity values decreases.

F.2 Details on Green Security Game simulations

F.2.1 Data description

The data is from a study tracking geo-location information of 32 African elephants in the Hwange National Park in Zimbabwe, Africa. The data was collected from 2009 to 2017 and includes time stamp information as well as longitudinal and latitudinal information of the elephants

F.2.2 Data processing

We divided each year into two based on the seasons in Zimbabwe which would affect the location of the elephants as they migrate given changes in the environment. A year was divided into a “Rainy” season which runs from October to March and a “Dry” season which runs from April to September.

We divided the area covered in the dataset into 9 locations and thus had a 3×3 grid which served as a surveillance area. For each elephant in the season, we calculated its mean location and so for each of the seasons we had a mean location for the elephants. It is important to note, given migratory nature of elephants across national borders, each season does not necessarily have all 32 elephants as they move as a result of a wide range of factors (e.g., availability of water). Some seasons, also notably, do not have any recorded elephant presence in the area under surveillance. We do not see this as a limitation as the dynamic environment presents changes in the game which the defender has to take into account.

F.2.3 Green Security Game

Each of the seasonal elephant information presents us with a state for our Stochastic Bayesian Game. We define the game in the following manner:

- **State:** We have 16 states from the seasonal information we get from the data
- **Actions:** Each player (attacker or defender) has 9 available actions each corresponding to a selection of an area in the 3×3 grid defined above.
- **Payoffs:** Let n be the number of elephants recorded in a particular grid square:

- If the attacker and defender select the same grid square, the defender gets n while the attacker gets -2
- If the attacker and defender select different grid squares, let n_{att} be the number of elephants in the grid square selected by the attacker. The defender gets payoff $-n_{att}$ while the attacker gets payoff n_{att}

These payoffs were designed to model the asymmetric nature of the defending task. An offender often will get a fixed penalty as stipulated by law, whilst the defender always depends on the number of elephants the attacker has access to.

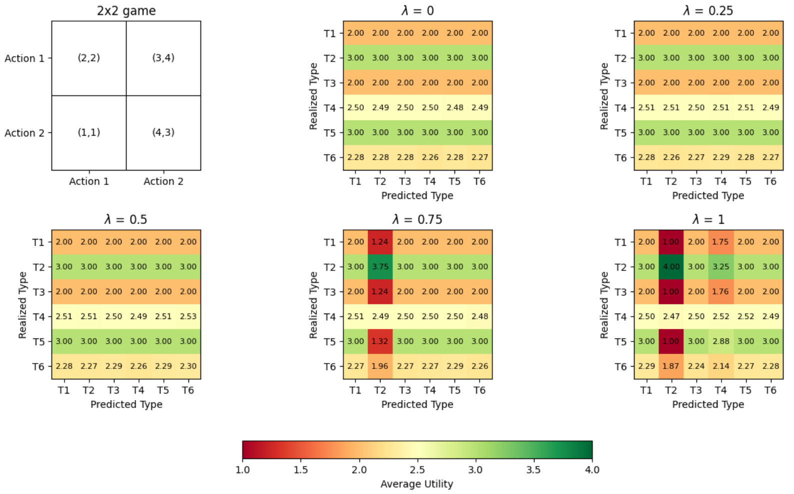
- **Transitions:** To take into account the effect of changes in weather whilst also bringing in some stochasticity, we made it such that there is some transition probability between any "Rainy" state to any "Dry" state and vice-versa. We do however, slightly, increase the probability of transition between adjacent historical states, to reflect historical data. We have the probability of transitioning between any two "Rainy" or "Dry" states to be zero. Our transition probabilities do not depend on the actions taken by the agents.

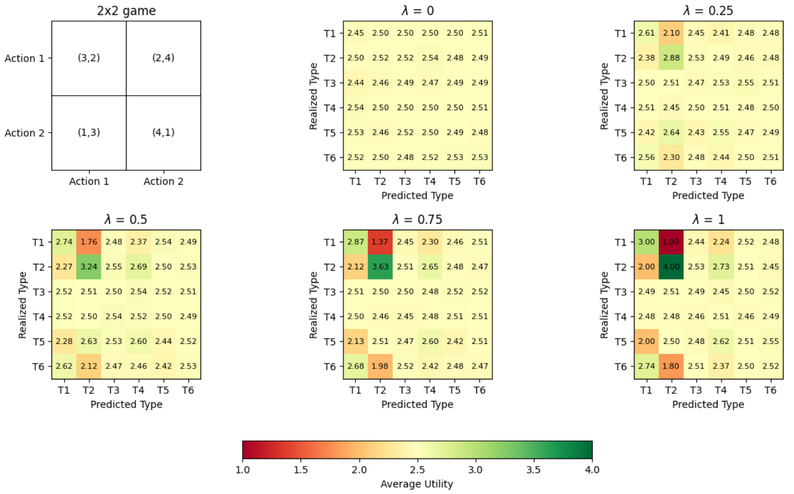
F.2.4 Type definitions

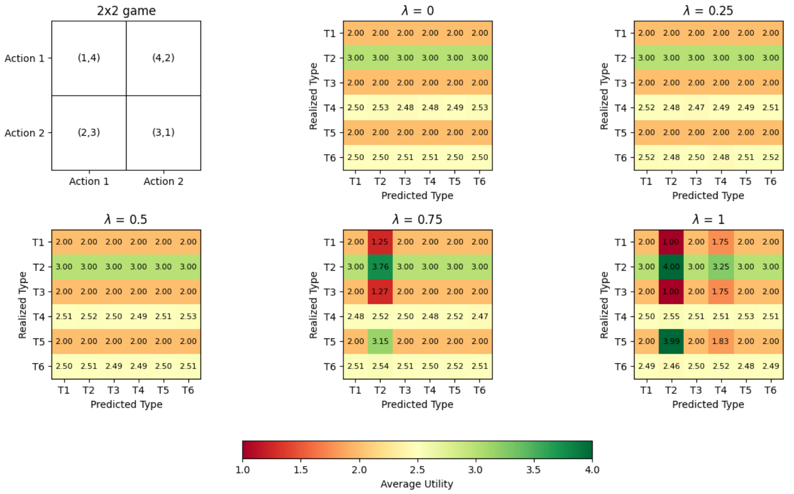
We make use of the same types as in the 2×2 game simulations. We make adjustments to the Markovian types in that Types 1 and 2 now select the grid with the highest population of elephants and second highest population of elephants, respectively. The Leader-Follower-Trigger Agents now looks to see if the other player is selecting the grid with the highest number of elephants for more than 50% of the historic play in which case they turn to play their minimax strategy

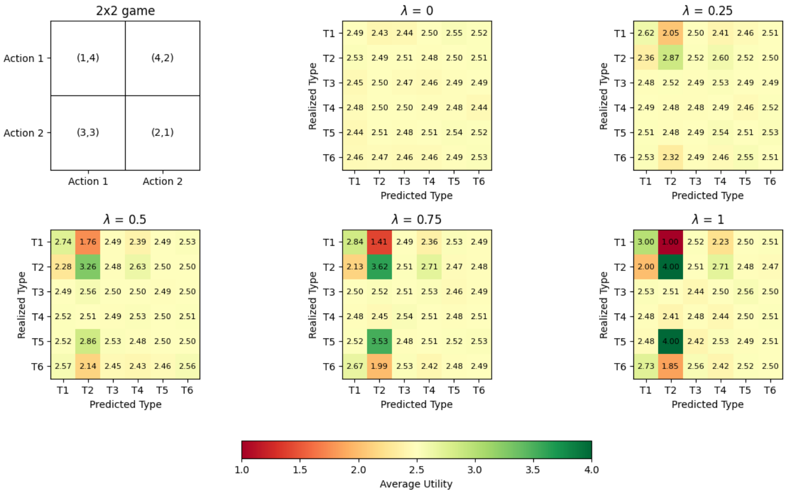
F.3 Additional 2×2 game evaluations

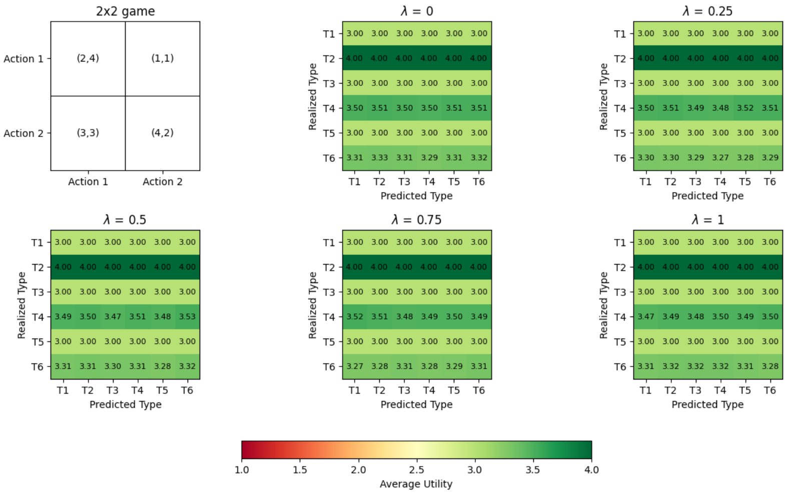
We include as an illustration as well as for completeness a couple of other 2×2 games we also evaluated against. This is helpful as it shows the variety of tradeoffs that exist within the topology of 2×2 games. In particular, we see some games exhibiting gradation as the agent moves from fully robust to fully trusting whilst in others there does not exist such a tradeoff because of the existence of a dominant strategy for the row player regardless of the type of the column player (e.g., the last game we show in this section). We note that the games provided in this file do not exhaust the entire topology of 2×2 games. They are however added to show the range of tradeoffs that could exist in games.


2x2 game
Action 1
Action 2
(2,2)
(3,4)
(1,1)
(4,3)
Action 1
Action 2
λ = 0
T1 2.00 2.00 2.00 2.00 2.00 2.00
T2 3.00 3.00 3.00 3.00 3.00 3.00
T3 2.00 2.00 2.00 2.00 2.00 2.00
T4 2.50 2.49 2.50 2.50 2.48 2.49
T5 3.00 3.00 3.00 3.00 3.00 3.00
T6 2.28 2.28 2.28 2.26 2.28 2.27
λ = 0.25
T1 2.00 2.00 2.00 2.00 2.00 2.00
T2 3.00 3.00 3.00 3.00 3.00 3.00
T3 2.00 2.00 2.00 2.00 2.00 2.00
T4 2.51 2.51 2.50 2.51 2.51 2.49
T5 3.00 3.00 3.00 3.00 3.00 3.00
T6 2.28 2.26 2.27 2.29 2.28 2.27
λ = 0.5
T1 2.00 2.00 2.00 2.00 2.00 2.00
T2 3.00 3.00 3.00 3.00 3.00 3.00
T3 2.00 2.00 2.00 2.00 2.00 2.00
T4 2.51 2.51 2.50 2.49 2.51 2.53
T5 3.00 3.00 3.00 3.00 3.00 3.00
T6 2.28 2.27 2.29 2.26 2.29 2.30
λ = 0.75
T1 2.00 1.24 2.00 2.00 2.00 2.00
T2 3.00 3.75 3.00 3.00 3.00 3.00
T3 2.00 1.24 2.00 2.00 2.00 2.00
T4 2.51 2.49 2.50 2.50 2.50 2.48
T5 3.00 1.32 3.00 3.00 3.00 3.00
T6 2.27 1.96 2.27 2.27 2.29 2.26
λ = 1
T1 2.00 1.00 2.00 1.75 2.00 2.00
T2 3.00 4.00 3.00 3.25 3.00 3.00
T3 2.00 1.00 2.00 1.76 2.00 2.00
T4 2.50 2.47 2.50 2.52 2.52 2.49
T5 3.00 1.00 3.00 2.88 3.00 3.00
T6 2.29 1.87 2.24 2.14 2.27 2.28
Realized Type
T1 T2 T3 T4 T5 T6
Predicted Type
1.0 1.5 2.0 2.5 3.0 3.5 4.0
Average Utility


2x2 game
Action 1
Action 2
(3,2)
(2,4)
(1,3)
(4,1)
Action 1
Action 2
λ = 0
T1 2.45 2.50 2.50 2.50 2.50 2.51
T2 2.50 2.52 2.52 2.54 2.48 2.49
T3 2.44 2.46 2.49 2.47 2.49 2.49
T4 2.54 2.50 2.50 2.50 2.50 2.51
T5 2.53 2.46 2.52 2.50 2.49 2.48
T6 2.52 2.50 2.48 2.52 2.53 2.53
λ = 0.25
T1 2.61 2.10 2.45 2.41 2.48 2.48
T2 2.38 2.88 2.53 2.49 2.46 2.48
T3 2.50 2.51 2.47 2.53 2.55 2.51
T4 2.51 2.45 2.50 2.51 2.48 2.50
T5 2.42 2.64 2.43 2.55 2.47 2.49
T6 2.56 2.30 2.48 2.44 2.50 2.51
λ = 0.5
T1 2.74 1.76 2.48 2.37 2.54 2.49
T2 2.27 3.24 2.55 2.69 2.50 2.53
T3 2.52 2.51 2.50 2.54 2.52 2.51
T4 2.52 2.50 2.54 2.52 2.50 2.49
T5 2.28 2.63 2.53 2.60 2.44 2.52
T6 2.62 2.12 2.47 2.46 2.42 2.53
λ = 0.75
T1 2.87 1.37 2.45 2.30 2.46 2.51
T2 2.12 3.63 2.51 2.65 2.48 2.47
T3 2.51 2.50 2.50 2.48 2.52 2.52
T4 2.50 2.46 2.45 2.48 2.51 2.51
T5 2.13 2.51 2.47 2.60 2.42 2.51
T6 2.68 1.98 2.52 2.42 2.48 2.47
λ = 1
T1 3.00 1.00 2.44 2.24 2.52 2.48
T2 2.00 4.00 2.53 2.73 2.51 2.45
T3 2.49 2.51 2.49 2.45 2.50 2.52
T4 2.48 2.48 2.46 2.51 2.46 2.49
T5 2.00 2.50 2.48 2.62 2.51 2.55
T6 2.74 1.80 2.51 2.37 2.50 2.52
Realized Type
T1 T2 T3 T4 T5 T6
Predicted Type
1.0 1.5 2.0 2.5 3.0 3.5 4.0
Average Utility


2x2 game
Action 1
Action 2
(1,4)
(4,2)
(2,3)
(3,1)
Action 1
Action 2
λ = 0
T1 2.00 2.00 2.00 2.00 2.00 2.00
T2 3.00 3.00 3.00 3.00 3.00 3.00
T3 2.00 2.00 2.00 2.00 2.00 2.00
T4 2.50 2.53 2.48 2.48 2.49 2.53
T5 2.00 2.00 2.00 2.00 2.00 2.00
T6 2.50 2.50 2.51 2.51 2.50 2.50
λ = 0.25
T1 2.00 2.00 2.00 2.00 2.00 2.00
T2 3.00 3.00 3.00 3.00 3.00 3.00
T3 2.00 2.00 2.00 2.00 2.00 2.00
T4 2.52 2.48 2.47 2.49 2.49 2.51
T5 2.00 2.00 2.00 2.00 2.00 2.00
T6 2.52 2.48 2.50 2.48 2.51 2.52
λ = 0.5
T1 2.00 2.00 2.00 2.00 2.00 2.00
T2 3.00 3.00 3.00 3.00 3.00 3.00
T3 2.00 2.00 2.00 2.00 2.00 2.00
T4 2.51 2.52 2.50 2.49 2.51 2.53
T5 2.00 2.00 2.00 2.00 2.00 2.00
T6 2.50 2.51 2.49 2.49 2.50 2.51
λ = 0.75
T1 2.00 1.25 2.00 2.00 2.00 2.00
T2 3.00 3.76 3.00 3.00 3.00 3.00
T3 2.00 1.27 2.00 2.00 2.00 2.00
T4 2.48 2.52 2.50 2.48 2.52 2.47
T5 2.00 3.15 2.00 2.00 2.00 2.00
T6 2.51 2.54 2.51 2.50 2.52 2.51
λ = 1
T1 2.00 1.00 2.00 1.75 2.00 2.00
T2 3.00 4.00 3.00 3.25 3.00 3.00
T3 2.00 1.00 2.00 1.75 2.00 2.00
T4 2.50 2.55 2.51 2.51 2.53 2.51
T5 2.00 3.99 2.00 1.83 2.00 2.00
T6 2.49 2.46 2.50 2.52 2.48 2.49
Realized Type
T1 T2 T3 T4 T5 T6
Predicted Type
1.0 1.5 2.0 2.5 3.0 3.5 4.0
Average Utility





NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The claims made in the abstract and introduction align with our theoretical and experimental results.

   Guidelines:

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

2. **Limitations**

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

Answer: [Yes]

Justification: The limitations of this work are discussed in our last concluding remark section.

Guidelines:

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

3. **Theory Assumptions and Proofs**

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

   Answer: [Yes]

   Justification: We have clearly stated our model assumptions for the derived results to hold.

   Guidelines:

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

4. **Experimental Result Reproducibility**

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

   Answer: [Yes]

   Justification: We have detailed the settings and parameters used in the experiments, and we will further release our code after the anonymous review stage.

Guidelines:

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

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

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

   Answer: [NA]

   Justification: Our contributions focus on the theoretical side. The experimental results shown in the submitted manuscript do not depend on private datasets and can be reproduced using the provided settings and parameters.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.

- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

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

   Answer: [Yes]

   Justification: We have clearly stated our experimental settings and details to reproduce the results.

   Guidelines:

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

7. **Experiment Statistical Significance**

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

   Answer: [Yes]

   Justification: When there is randomness in our experiments, we characterize the variability.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

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

   Answer: [Yes]

   Justification: Our contributions focus on the theory side, and the experimental setup is sufficiently basic that it does not require intense computing resources such as GPUs.

   Guidelines:

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

9. **Code Of Ethics**

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

   Answer: [Yes]

   Justification: The research conducted conform with the NeurIPS Code of Ethics.

   Guidelines:

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

10. **Broader Impacts**

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

    Answer: [Yes]

    Justification: We have discussed both positive and negative societal impacts in our concluding remark section.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

Answer: [NA]

Justification: The paper poses no such risks.

Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [NA]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

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

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

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

14. **Crowdsourcing and Research with Human Subjects**

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

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

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

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

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

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

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