R2-B2: Recursive Reasoning-Based Bayesian Optimization for No-Regret Learning in Games

Zhongxiang Dai [1]   Yizhou Chen [1]   Bryan Kian Hsiang Low [1]   Patrick Jaillet [2]   Teck-Hua Ho [3]

[1] Department of Computer Science, National University of Singapore
[2] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology
[3] NUS Business School, National University of Singapore

Overview

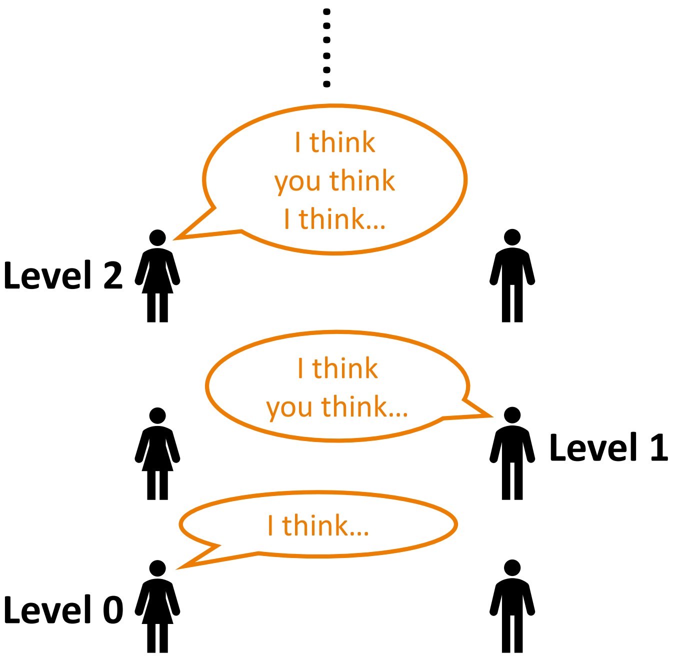
- **Problem**:
  - Repeated games between boundedly rational, self-interested agents, with unknown, complex and costly-to-evaluate payoff functions.

Adversarial Machine Learning (ML)

- **Solution**:
  - **R2-B2**: Recursive Reasoning + Bayesian Optimization

Model the reasoning process in interactions between agents

Principled efficient strategies for action selection



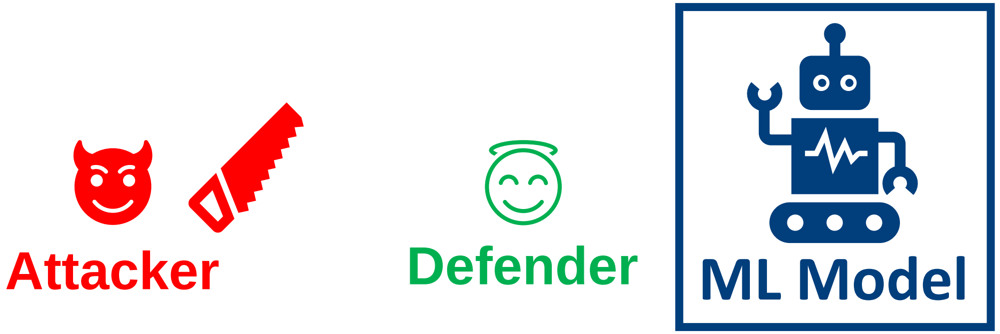
Cognitive hierarchy model of games

- **Theoretical results:**
  - No-regret strategies for different levels of reasoning
  - Improved convergence for level-$k \geq 2$ reasoning
- **Empirical results:**
  - Adversarial ML, and multi-agent reinforcement learning

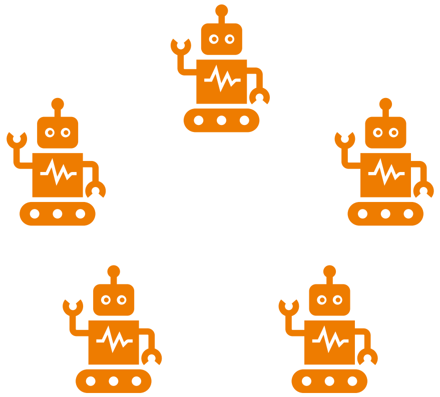
https://en.wikipedia.org/wiki/R2-D2

Introduction

- Some real-world machine learning (ML) tasks can be modelled as **repeated games** between **boundedly rational, self-interested agents**, with **unknown, complex and costly-to-evaluate payoff functions.**



Adversarial Machine Learning (ML)



Multi-Agent Reinforcement Learning (MARL)

Introduction

- How do we derive an efficient strategy for these games?
  - The payoffs of different actions of each agent are usually **correlated**

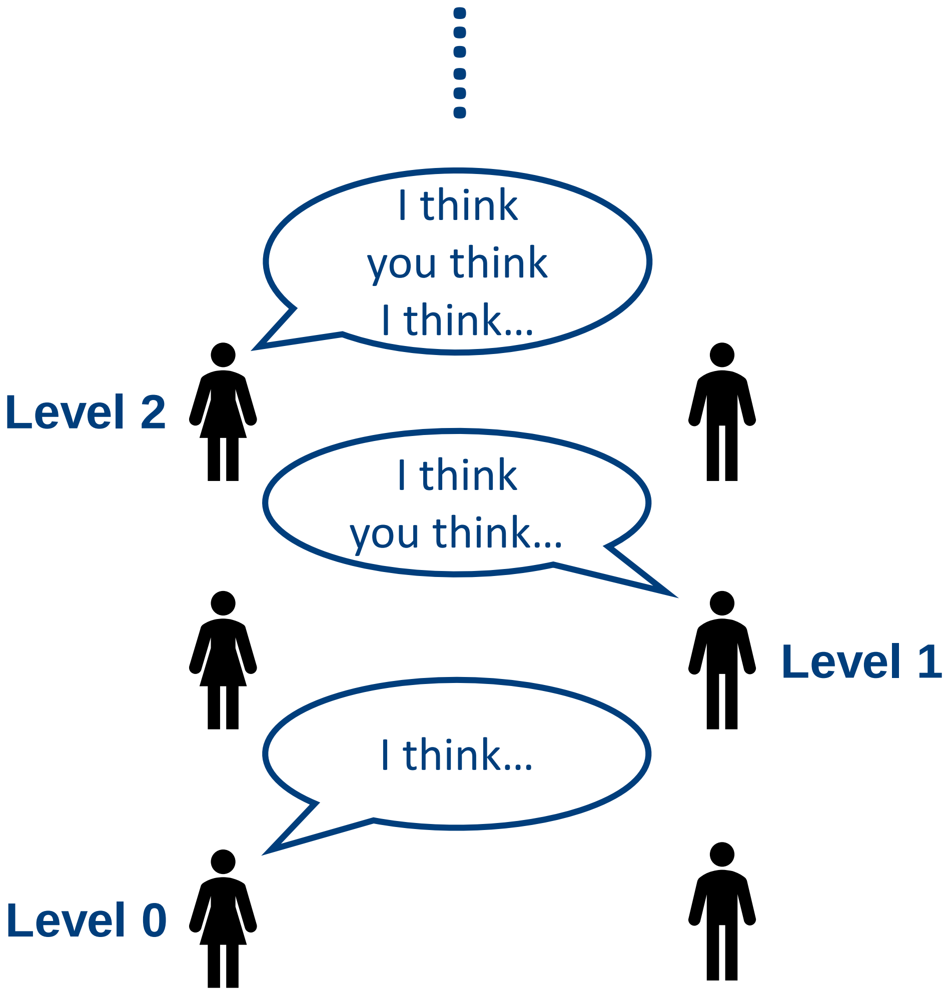
  - Predict the payoff function using **Gaussian processes** (GP)
  - Select actions using **Bayesian optimization** (BO)

- How do we account for **interactions between agents** in a principled way?

Introduction

- The **cognitive hierarchy model of games** (Camerer et al., 2004) models the **recursive reasoning** process between **humans,** i.e., boundedly rational, self-interested agents.

- Every agent is associated with a level of reasoning k (**cognitive limit**):
  - **Level-0 Agent**: randomizes action
  - **Level-$k \geq 1$ Agent**: best-responds to lower-level agents



Introduction

- We introduce ***R2-B2***:

***R**ecursive **R**easoning-**B**ased **B**ayesian optimization*, to help agents perform effectively in these games through the **recursive reasoning** formalism

- **Repeated games** with **simultaneous moves** and **perfect monitoring**

- **Generally applicable**:
  - Constant-sum games (e.g., adversarial ML)
  - General-sum games (e.g., MARL)
  - Common-payoff games

https://en.wikipedia.org/wiki/R2-D2

Recursive Reasoning-Based Bayesian Optimization (R2-B2)

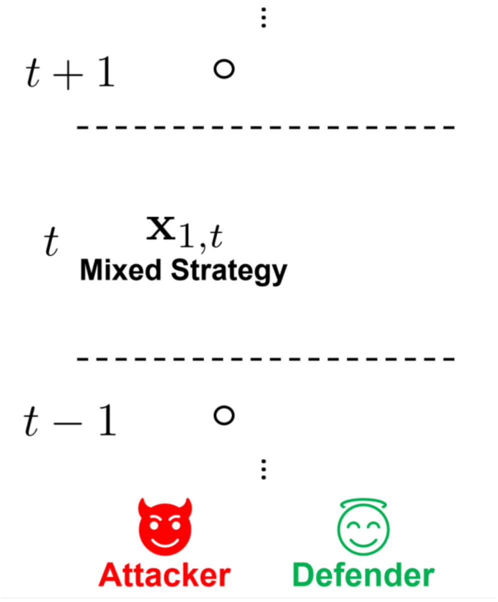
- We focus on the view of **Attacker (A)**, playing against **Defender (D)**
- Can be extended to games with ≥ 2 agents

Algorithm 1 R2-B2 for attacker $\mathcal{A}$'s level-k reasoning

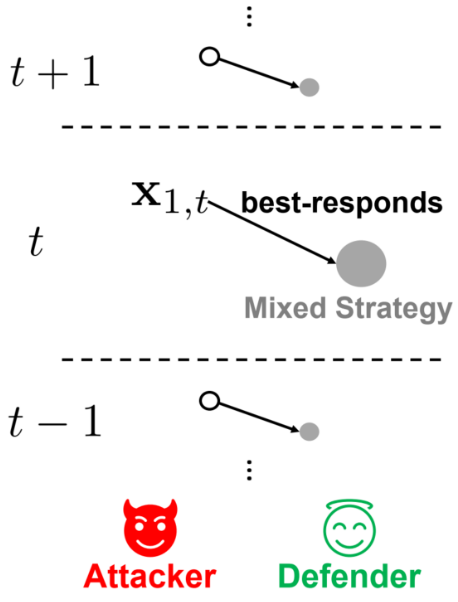
for $t = 1, 2, \ldots, T$ **do**
    Select input action $\mathbf{x}_{1,t}$ using its level-k strategy (while defender $\mathcal{D}$ selects input action $\mathbf{x}_{2,t}$)
    Observe noisy payoff $y_{1,t} = f_1(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}) + \epsilon_1$
    Update GP posterior belief using $\langle(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}), y_{1,t}\rangle$

Recursive Reasoning-Based Bayesian Optimization (R2-B2)

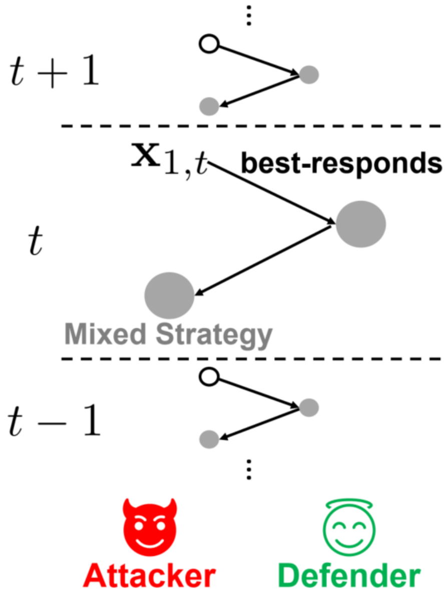
- **Level-0**: randomized action selection (mixed strategy)
- **Level-$k \geq 1$**: best-responds to level-$(k-1)$ agents



Leve-0 Strategy



Leve-1 Strategy



Leve-2 Strategy

Recursive Reasoning-Based Bayesian Optimization (R2-B2)

Level-$k = 0$ Strategy

- **Require no knowledge about opponent's strategy**
- Mixed strategy
- Any strategy, including **existing baselines**, can be considered as level-0

- Some reasonable choices:
  - Random search
  - EXP3 for adversarial linear bandit
  - **GP-MW** (Sessa et al., 2019); sublinear upper bound on the regret:

$$R_{1,T} = \mathcal{O}(\sqrt{T \log |\mathcal{X}_1|} + \sqrt{T \log(2/\delta)} + \sqrt{T \beta_T \gamma_T})$$

Recursive Reasoning-Based Bayesian Optimization (R2-B2)

Level-$k = 1$ Strategy

Attacker's level-1 action

GP-UCB acquisition function

$$\mathbf{x}_{1,t}^{1} \triangleq \arg\max_{\mathbf{x}_1 \in \mathcal{X}_1} \mathbb{E}_{\mathbf{x}_{2,t}^{0} \sim \mathcal{P}_{2,t}^{0}} [\alpha_{1,t}(\mathbf{x}_1, \mathbf{x}_{2,t}^{0})]$$

Opponent's level-0 mixed strategy

- Sublinear upper bound on the expected regret:

$$\mathbb{E}[R_{1,T}] \leq \sqrt{C_1 T \beta_T \gamma_T}$$

- Holds for **any** opponent's level-0 strategy
- **Opponent may not even perform recursive reasoning**

Recursive Reasoning-Based Bayesian Optimization (R2-B2)

Level-$k \geq 2$ Strategy

- Sublinear upper bound on the regret:

$$R_{1,T} \leq \sqrt{C_1 T \beta_T \gamma_T}$$

- **Converges faster** than level-0 strategy using GP-MW
- Higher level of reasoning $\Rightarrow$ more computational cost

⬇

- Agents **favour reasoning at lower levels**

⬇

- Cognitive hierarchy model: **humans usually reason at a level ≤ 2**

Attacker's level-k action

Defender's level-$(k-1)$ action

$$\mathbf{x}_{1,t}^{k} \triangleq \arg\max_{\mathbf{x}_1 \in \mathcal{X}_1} \alpha_{1,t}(\mathbf{x}_1, \mathbf{x}_{2,t}^{k-1})$$

$$\mathbf{x}_{2,t}^{k-1} \triangleq \arg\max_{\mathbf{x}_2 \in \mathcal{X}_2} \alpha_{2,t}(\mathbf{x}_{1,t}^{k-2}, \mathbf{x}_2)$$

......

Compute recursively until level 1

Recursive Reasoning-Based Bayesian Optimization (R2-B2)

R2-B2-Lite for Level-1 Reasoning

- R2-B2-Lite for level-1 reasoning:
  - **Better computational efficiency**
  - **Worse convergence guarantee**

- Firstly sample an action from opponent's level-0 strategy: $\widetilde{\mathbf{x}}^0_{2,t}$
- Then select

$$\mathbf{x}^1_{1,t} \triangleq \arg\max_{\mathbf{x}_1 \in \mathcal{X}_1} \alpha_{1,t}(\mathbf{x}_1, \widetilde{\mathbf{x}}^0_{2,t})$$

More accurate action sampling

- Theoretical insights:
  - Benefits if opponent's level-0 strategy has **smaller variance**
  - Asymptotically no-regret **if the variance of opponent's level-0 strategy → 0**

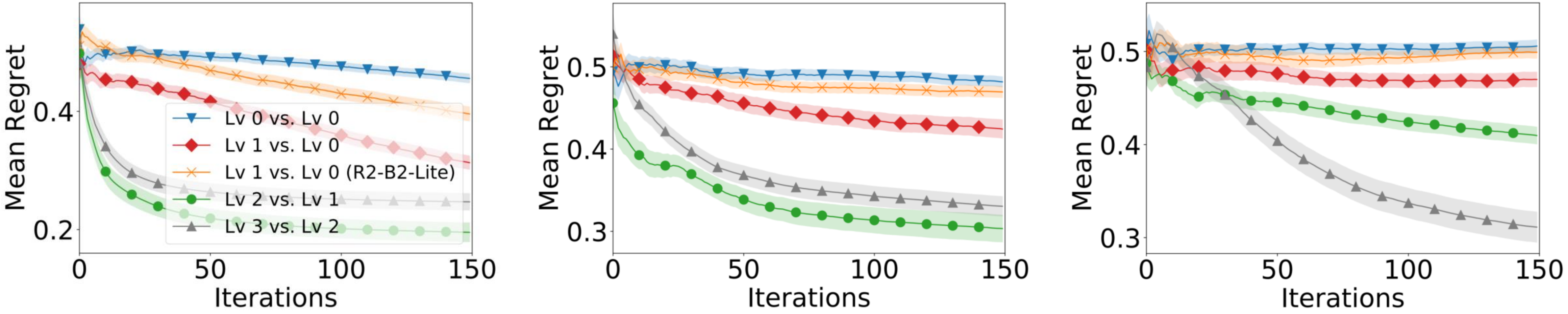
Exploration ⇒ Exploitation

Experiments and Discussion

Synthetic Games (2 agents)

- GP-MW level-0 strategy
- **Reasoning at one level higher than opponent** gives better performance
- Our level-1 agent outperforms the baseline of GP-MW (red vs blue)
- Effect of **incorrect thinking about opponent's level of reasoning**

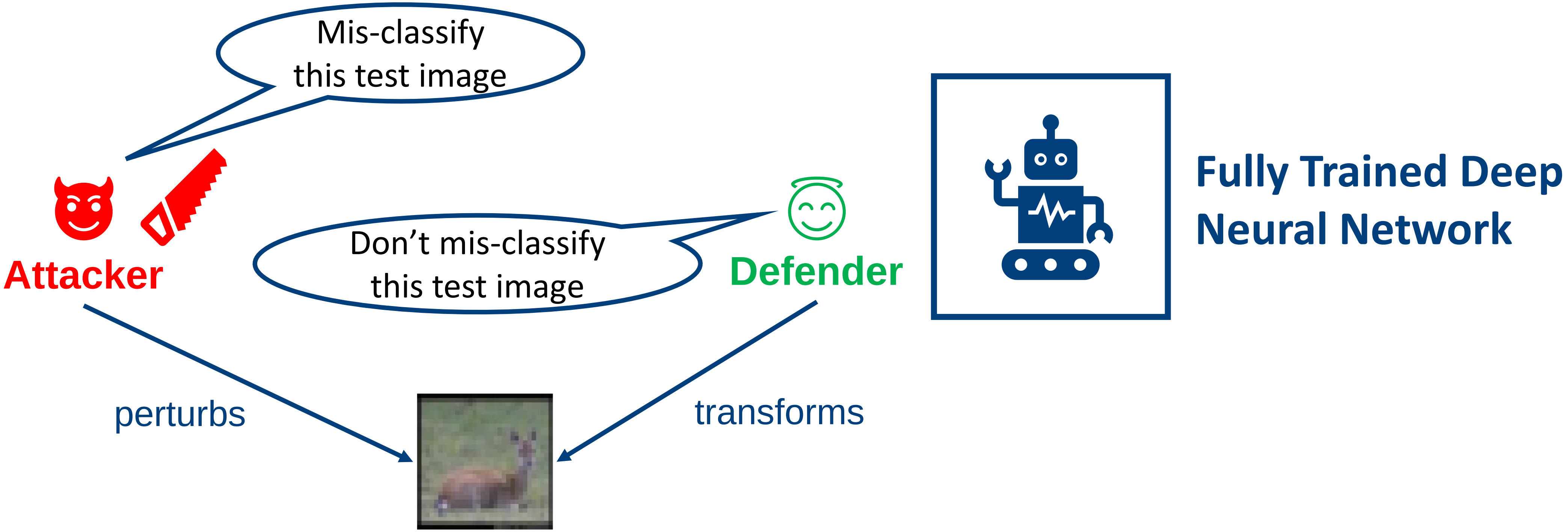
Mean regret of **agent 1 (**legends**: level of agent 1 vs. agent 2)**



Common-payoff General-sum Constant-sum

Experiments and Discussion

Adversarial Machine Learning (ML)



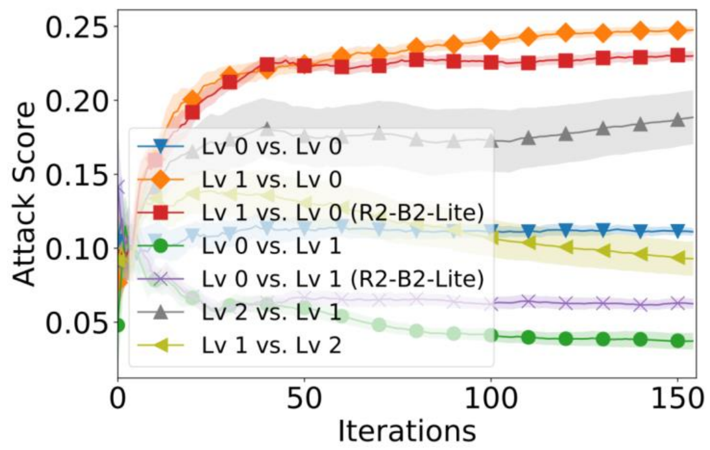
Experiments and Discussion

Adversarial Machine Learning (ML)

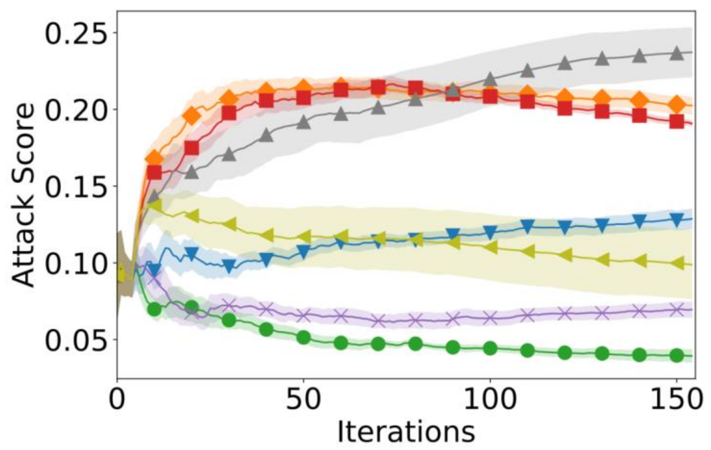
- When **attacker** reasons at one level higher than **defender** $\Rightarrow$ **higher attack scores, more successful attacks**
- The same applies to the defender

Table 1. Average number of successful attacks by $\mathcal{A}$ over 150 iterations in adversarial ML for MNIST and CIFAR-10 datasets where the levels of reasoning are in the form of $\mathcal{A}$ vs. $\mathcal{D}$.

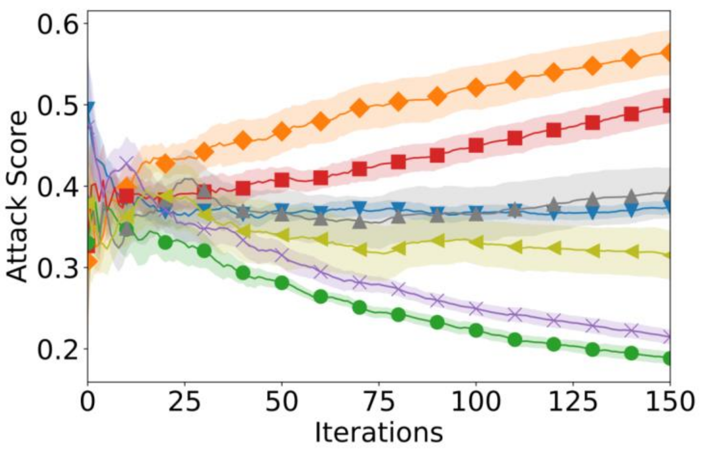
Levels of reasoning	MNIST (random)	MNIST (GP-MW)	CIFAR-10
0 vs. 0	2.6	4.3	70.1
1 vs. 0	12.8	6.0	113.1
1 vs. 0 (R2-B2-Lite)	10.2	6.8	99.7
0 vs. 1	0.8	0.4	25.2
0 vs. 1 (R2-B2-Lite)	1.8	1.0	29.7
2 vs. 1	3.0	5.2	70.9
1 vs. 2	0.9	0.4	54.0



MNIST, random search



MNIST, GP-MW



CIFAR-10, random search

Experiments and Discussion

Adversarial Machine Learning (ML)

- Play our **level-1 defender** against state-of-the-art black-box adversarial attacker, **Parsimonious**, used as **level-0 strategy**

- Among 70 CIFAR-10 images
  - ***Completely prevent any successful attacks*** for 53 images
  - ***Requires*** ≥ 3.5 ***times more queries*** for 10 other images

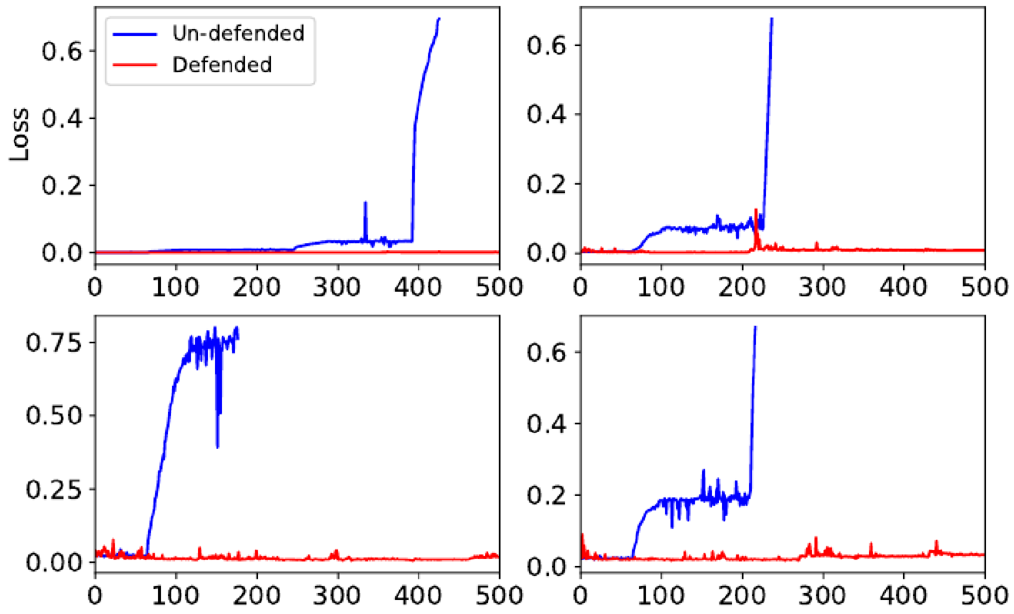


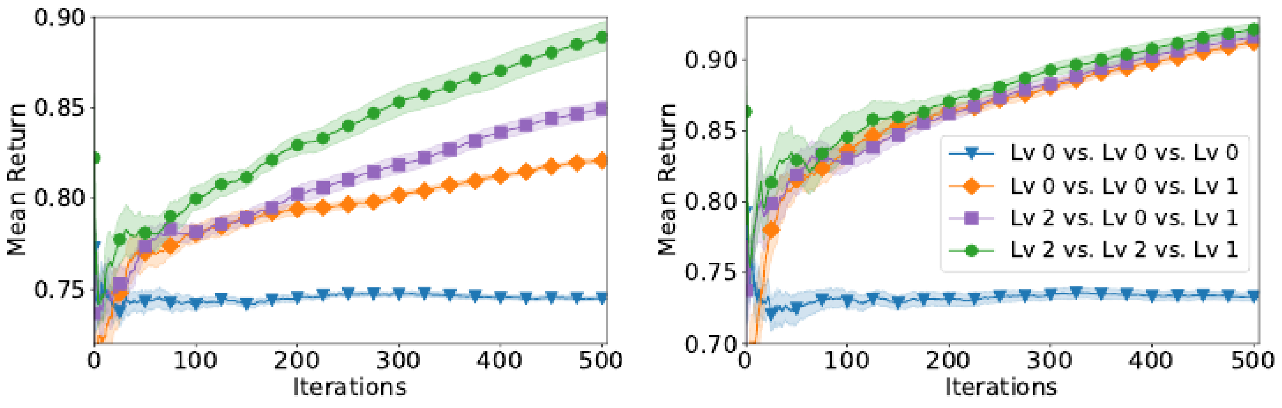
Figure 3. Loss incurred by Parsimonious with and without our level-1 R2-B2 defender on 4 randomly selected images that are successfully attacked by Parsimonious.

Experiments and Discussion

Multi-Agent Reinforcement Learning (MARL)

- Predator-pray game: **2 predators vs 1 prey**
- **General-sum game**

- Prey at level 1 ⇒ better return for prey
- 1 predator at one level higher ⇒ better return for predators
- 2 predators at one level higher ⇒ even better return for predators



(a) predators (b) prey

Conclusion and Future Work

- We introduce **R2-B2**, the first **recursive reasoning** formalism of BO to model the reasoning process in the interactions between **boundedly rational, self-interested agents** with **unknown, complex, and costly-to-evaluate payoff functions** in **repeated games**

- Future works:
  - Extend R2-B2 to allow a level-k agent to best-respond to an agent whose **reasoning level follows a distribution** such as Poisson distribution (Camerer et al., 2004)
  - Investigate connection of R2-B2 with other game-theoretic solution concepts such as **Nash equilibrium**