Game Redesign in No-Regret Game Playing

Jerry Zhu

SILO Dec. 1, 2021

Joint work with Yuzhe Ma (CS->Microsoft), Young Wu (CS)

Prisoner's Dilemma

ℓ_1(**mum, mum**)

Player 2

Player 1

	mum	fink
mum	2, 2	5, 1
fink	1, 5	4, 4

Prisoner's Dilemma

a_i **strictly dominated by** $a_i' : \forall a_{-i} : \ell_i(a_i, a_{-i}) > \ell_i(a_i', a_{-i})$

	mum	fink
~~mum~~	~~2, 2~~	~~5, 1~~
fink	1, 5	4, 4

Prisoner's Dilemma

	mum	fink
~~mum~~	~~2, 2~~	~~5, 1~~
fink	~~1, 5~~	4, 4

dominant strategy equilibrium

Prisoner's Dilemma

	mum	fink
mum	2, 2	5, 1
fink	1, 5	4, 4

also (pure) Nash equilibrium

$$\forall i, a_i : \ell_i(a_i^*, a_{-i}^*) \leq \ell_i(a_i, a_{-i}^*)$$

No-Regret Game Playing

---

for $t = 1, 2, \ldots, T$ **do**

Players form action profile $a^t = (a_1^t, \ldots, a_M^t)$, where $a_i^t \sim \pi_i^t, \forall i \in [M]$.

Player i observes the loss $\ell_i(a^t)$ and updates policy π_i^t.

end for

---

e.g. EXP3.P

No-Regret Game Playing

Regret $R_i^T = \sum_{t=1}^{T} \ell_i^t(a_i^t, a_{-i}^t) - \min_{b \in A_i} \sum_{t=1}^{T} \ell_i^t(b, a_{-i}^t)$

$\alpha-$**No-Regret player:** $\mathbb{E}[R_i^T] = O(T^\alpha)$

e.g. EXP3.P $\alpha = 1/2$

Approximate Nash equilibrium (two-player zero-sum)

Approximate coarse correlated equilibrium (general-sum)

No-Regret Game Playing

	mum	fink
mum	2, 2	5, 1
fink	1, 5	4, 4

will get here, bummer

“I’m gonna make them an offer they can’t refuse.” — game redesigner

Redesigned Prisoner's Dilemma

	mum	fink
mum	2, 2	1.5, 2.5
fink	2.5, 1.5	4, 4

Volunteer's Dilemma

$M = 3$ **players**

		Number of other volunteers		
		0	1	2
Player i	volunteer	0	0	0
	not volunteer	10	-1	-1

Nash has free-riders.

Won’t it be nice to make everyone volunteer?

		Number of other volunteers		
		0	1	2
Player i	volunteer	0	0	0
	not volunteer	10	-1	-1

Game Redesign Goals

1. Force players to choose a target joint action $a^\dagger$ in *T-o(T)* rounds
2. Only incur *o(T)* cumulative design cost $\sum_{t=1}^{T} C(\ell^0, \ell^t, a^t)$

Game Redesign Protocol

Original game

target joint action

Designer knows ℓ^o, $a^\dagger$, M, $\mathcal{A}_i, \ldots, \mathcal{A}_M$, and player no-regret rate α

for $t = 1, 2, \ldots, T$ **do**

Designer prepares new loss function ℓ^t.

Players form action profile $a^t = (a_1^t, \ldots, a_M^t)$, where $a_i^t \sim \pi_i^t, \forall i \in [M]$.

Player i observes the new loss $\ell_i^t(a^t)$ and updates policy π_i^t.

Designer incurs cost $C(\ell^o, \ell^t, a^t)$.

end for

Are Players Suspicious of ℓ^t ?

No	$\ell_i^t(a) \in \mathbb{R}$
A little	$\ell_i^t(a) \in [L, U]$
Somewhat	$\ell_i^t(a) \in \mathscr{L}$
Very	?

Design Cost

$$C(\ell^0, \ell^t, a^t) := \|\ell^0(a^t) - \ell^t(a^t)\|_1$$

	mum	fink
mum	2, 2	5, 1
fink	1, 5	4, 4

	mum	fink
mum	2, 2	1.5, 2.5
fink	2.5, 1.5	4, 4

$$(5 - 1.5) + (2.5 - 1)$$

Game Redesign Goals (Recap)

1. Force target $a^\dagger$ in *T-o(T)* rounds
2. *o(T)* cumulative design cost $\sum_{t=1}^{T} \|\ell^0(a^t) - \ell^t(a^t)\|_1$

Main Idea

1. Make $a^\dagger$ the dominant strategy equilibrium
2. Don't ever change $\ell^0(a^\dagger)$

Main Idea

Easier when

$$\ell_i^0(a^\dagger) < U$$

1. Make $a^\dagger$ the dominant strategy equilibrium
2. Don't ever change $\ell^0(a^\dagger)$

Make other actions look worse!

Algorithm 1: Interior Design

$$\exists \rho > 0 : \ell_i^0(a^\dagger) \in [L + \rho, U - \rho]$$

Input: the target action profile $a^\dagger$; the original game ℓ^o.
Output: a time-invariant game ℓ constructed as follows:

$$\forall i, a, \ell_i(a) = \begin{cases} \ell_i^o(a^\dagger) - (1 - \frac{d(a)}{M})\rho & \text{if } a_i = a_i^\dagger, \\ \ell_i^o(a^\dagger) + \frac{d(a)}{M}\rho & \text{if } a_i \neq a_i^\dagger, \end{cases}$$

where $d(a) = \sum_{j=1}^{M} \mathbb{1}\left[a_j = a_j^\dagger\right]$.

Algorithm 1: Interior Design

Optional postprocessing for general-sum games:

$$\forall i, a, \ell_i(a) = \begin{cases} \min\{\ell_i^o(a^\dagger) - (1 - \frac{d(a)}{M})\rho, \ell^o(a)\} & \text{if } a_i = a_i^\dagger \\ \max\{\ell_i^o(a^\dagger) + \frac{d(a)}{M}\rho, \ell^o(a)\} & \text{if } a_i \neq a_i^\dagger \end{cases}$$

Prisoner's Dilemma

ℓ^0

	mum	fink
mum	2, 2	5, 1
fink	1, 5	4, 4

ℓ

	mum	fink
mum	2, 2	1.5, 2.5
fink	2.5, 1.5	4, 4

Volunteer's Dilemma

ℓ^0

Player i

		Number of other volunteers: 0	1	2
Player i	volunteer	0	0	0
	not volunteer	10	-1	-1

ℓ

		Number of other volunteers: 0	1	2
Player i	volunteer	$-2/3$	$-1/3$	0
	not volunteer	10	$1/3$	$2/3$

Volunteer’s Dilemma

Covid: Greece to fine over-60s who refuse Covid-19 vaccine

1 day ago

BBC

ℓ

		Number of other volunteers		
		0	1	2
Player i	volunteer	$-2/3$	$-1/3$	0
	not volunteer	10	$1/3$	$2/3$

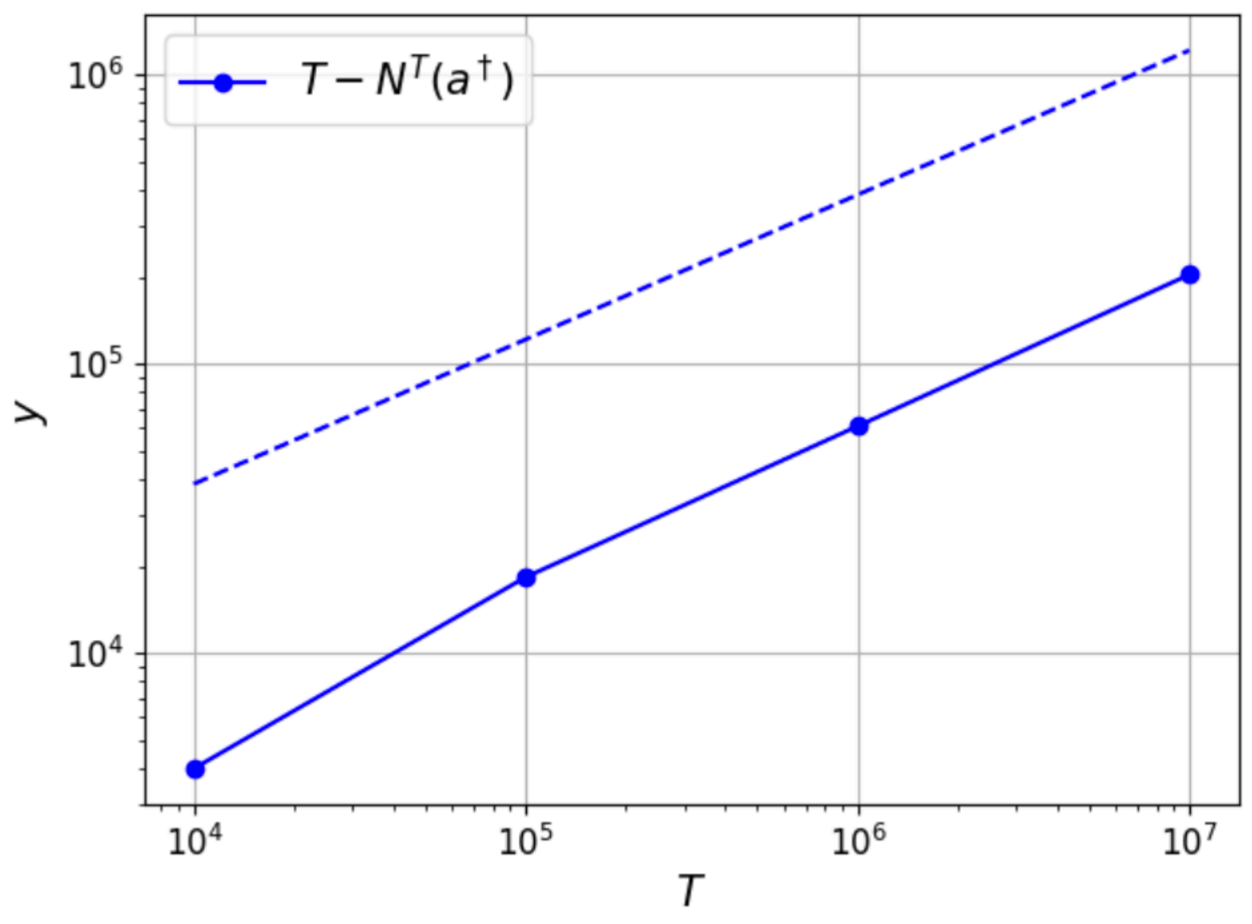
Interior Design Guarantees

$$\mathbb{E}[\sum_{t=1}^{T} 1(a^t = a^\dagger)] = T - O(MT^\alpha)$$

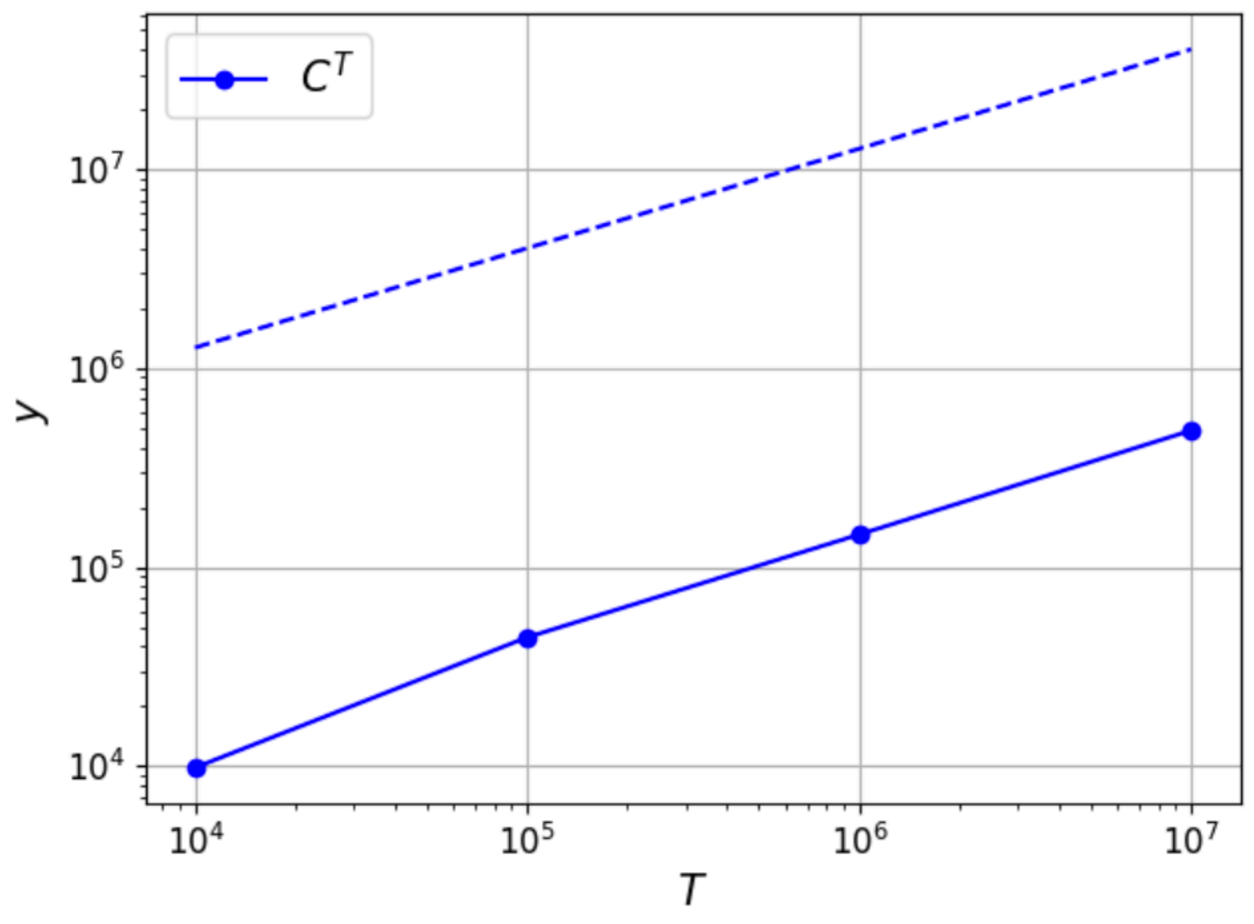
$$\mathbb{E}[\sum_{t=1}^{T} \|\ell^0(a^t) - \ell(a^t)\|_1] = O(M^2 T^\alpha)$$

Volunteer's Dilemma (3 EXP3.P players)

T	10^4	10^5	10^6	10^7
Target	60%	82%	94%	98%
Per-round Cost	0.98	0.44	0.15	0.05



(a) Number of rounds with $a^t \neq a^\dagger$ grows sublinearly



(b) The cumulative design cost grows sublinearly too

Non-target play, cumulative cost

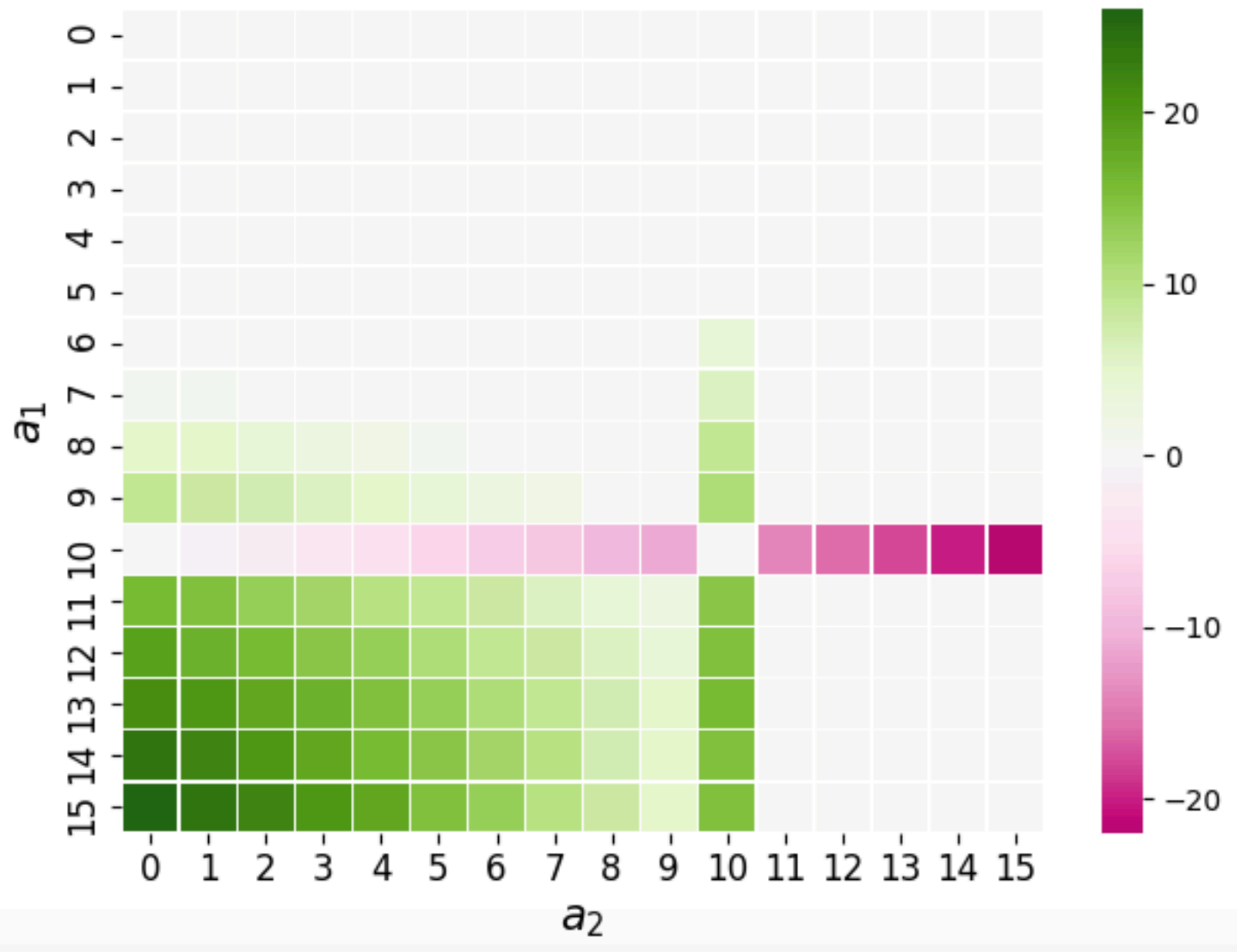
$$O(\sqrt{T})$$

Tragedy of the Commons

- Two farmers
- Each grace {0, 1, …, 15} sheep
- Price per sheep $p(a) = \sqrt{30 - a_1 - a_2}$
- Loss $-p(a)a_i$
- Nash equilibrium: $a^* = (12,12)$
- Suboptimal social welfare $-p(a^*)(a_1^* + a_2^*) \approx -59$

Redesigned Commons

- social welfare optimizer $a^\dagger = (10{,}10)$ $\quad -p(a^\dagger)(a_1^\dagger + a_2^\dagger) \approx -63$



$\ell_1(a) - \ell_1^0(a)$

T	10^4	10^5	10^6	10^7
Target	41%	77%	92%	98%
Cost	9.4	4.2	1.4	0.5

Main Idea (Revisited)

What if

$$\ell_i^0(a^\dagger) = U?$$

1. Make $a^\dagger$ the dominant strategy equilibrium
2. Don't ever change $\ell^0(a^\dagger)$

Cannot make other actions look worse!

Algorithm 2: Boundary Design

Works for any $\ell_i^0(a^\dagger)$: boundary or interior.

any interior vector

Output: a time-varying game with loss ℓ^t.

1: Use v in place of $\ell^o(a^\dagger)$ in (2) and apply the interior design 1. Call the resulting time-invariant game the "source game" $\underline{\ell}$.

2: Define a "destination game" $\overline{\ell}$ where $\overline{\ell}(a) = \ell^o(a^\dagger), \forall a$.

3: Interpolate the source and destination games:

$$\ell^t = w_t \underline{\ell} + (1 - w_t)\overline{\ell} \tag{8}$$

where

$$w_t = t^{\alpha+\epsilon-1} \tag{9}$$

$\epsilon \in (0, 1-\alpha)$: Slower decay than player regret

Rock-Paper-Scissors

$v = (0,0)$
$\epsilon = 0.3$

	R	P	S
R	$-0.5, 0.5$	$0, 0$	$-0.5, 0.5$
P	$0, 0$	$0.5, -0.5$	$0, 0$
S	$0, 0$	$0.5, -0.5$	$0, 0$

(a) $\ell^t (t = 1)$.

	R	P	S
R	$0.62, -0.62$	$0.75, -0.75$	$0.62, -0.62$
P	$0.75, -0.75$	$0.87, -0.87$	$0.75, -0.75$
S	$0.75, -0.75$	$0.87, -0.87$	$0.75, -0.75$

(b) $\ell^t (t = 10^3)$.

	R	P	S
R	$0.94, -0.94$	$0.96, -0.96$	$0.94, -0.94$
P	$0.96, -0.96$	$0.98, -0.98$	$0.96, -0.96$
S	$0.96, -0.96$	$0.98, -0.98$	$0.96, -0.96$

(c) $\ell^t (t = 10^7)$.

Boundary Design Guarantees

$$\mathbb{E}[\sum_{t=1}^{T} 1(a^t = a^\dagger)] = T - O(MT^{1-\epsilon})$$

$$\mathbb{E}[\sum_{t=1}^{T} \|\ell^0(a^t) - \ell(a^t)\|_1] = O(M^2T^{1-\epsilon} + MT^{\alpha+\epsilon})$$

Are Players Suspicious of ℓ^t ?

No	$\ell_i^t(a) \in \mathbb{R}$
A little	$\ell_i^t(a) \in [L, U]$
Somewhat	$\ell_i^t(a) \in \mathscr{L}$
Very	?

Algorithm 3: Discrete Design

$$\widehat{\ell}_i^{\,t}(a) \sim \mathrm{Ber}\left(\frac{U - \ell_i^t(a)}{U - L}, \frac{\ell_i^t(a) - L}{U - L}\right)$$

Rock-Paper-Scissors

	R	P	S
R	1, 1	1, 1	−1, 1
P	−1, −1	1, −1	−1, −1
S	−1, 1	−1, −1	−1, −1

(a) $\hat{\ell}^t (t = 1)$.

	R	P	S
R	1, −1	1, 1	−1, −1
P	1, −1	1, −1	1, −1
S	1, −1	1, −1	1, 1

(b) $\hat{\ell}^t (t = 10^3)$.

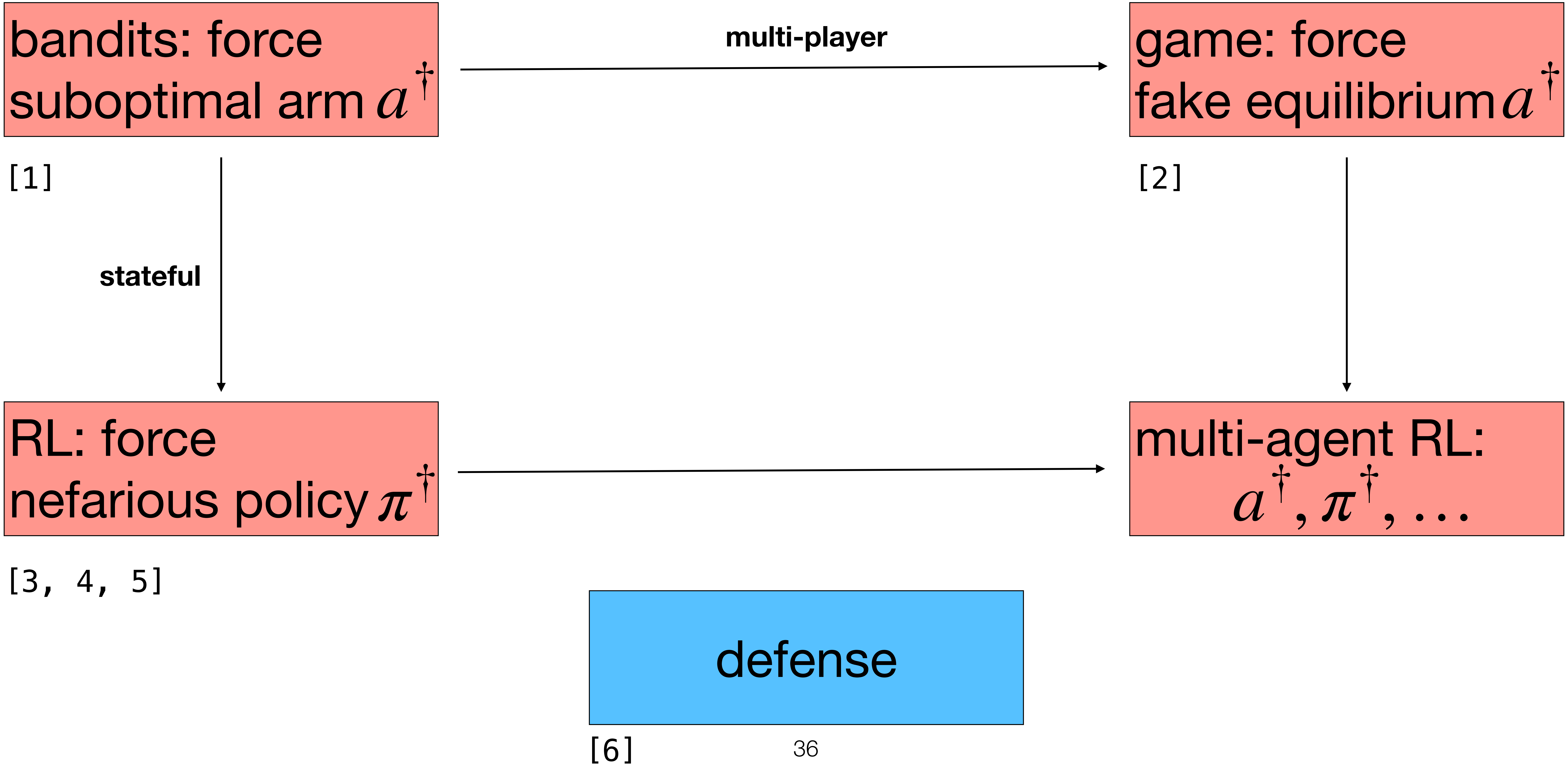
	R	P	S
R	1, −1	1, −1	1, −1
P	1, −1	1, −1	1, −1
S	1, −1	1, −1	1, −1

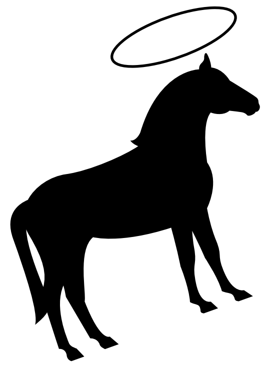
(c) $\hat{\ell}^t (t = 10^7)$.

T	10^4	10^5	10^6	10^7
Target	35%	59%	75%	88%
Per-round Cost	1.7	1.2	0.79	0.41

(almost the same performance as boundary design)

Related “Sequential Adversarial Attack” Problems





[1] Adversarial attacks on stochastic bandits. Kwang-Sung Jun, Lihong Li, Yuzhe Ma, and Xiaojin Zhu. NeurIPS 2018.

[2] Game Redesign in No-regret Game Playing. Yuzhe Ma, Young Wu, Xiaojin Zhu. https://arxiv.org/abs/2110.11763. 2021

[3] Policy poisoning in batch reinforcement learning and control. Yuzhe Ma, Xuezhou Zhang, Wen Sun, and Xiaojin Zhu. NeurIPS 2019.

[4] Online Data Poisoning Attacks. Xuezhou Zhang, Xiaojin Zhu, and Laurent Lessard. L4DC 2020.

[5] Adaptive reward-poisoning attacks against reinforcement learning. Xuezhou Zhang, Yuzhe Ma, Adish Singla, and Xiaojin Zhu. ICML 2020.

[6] Robust policy gradient against strong data corruption. Xuezhou Zhang, Yiding Chen, Xiaojin Zhu, and Wen Sun. ICML 2021.