Strategic Interactive Decision-Making

Thomas Kleine Buening

The Alan Turing Institute & University of Oxford

December 5, 2024

INTENTION

REDUCE COBRA POPULATION

INTENTION

REDUCE COBRA POPULATION

ACTION

A BOUNTY FOR DEAD COBRAS!

INTENTION

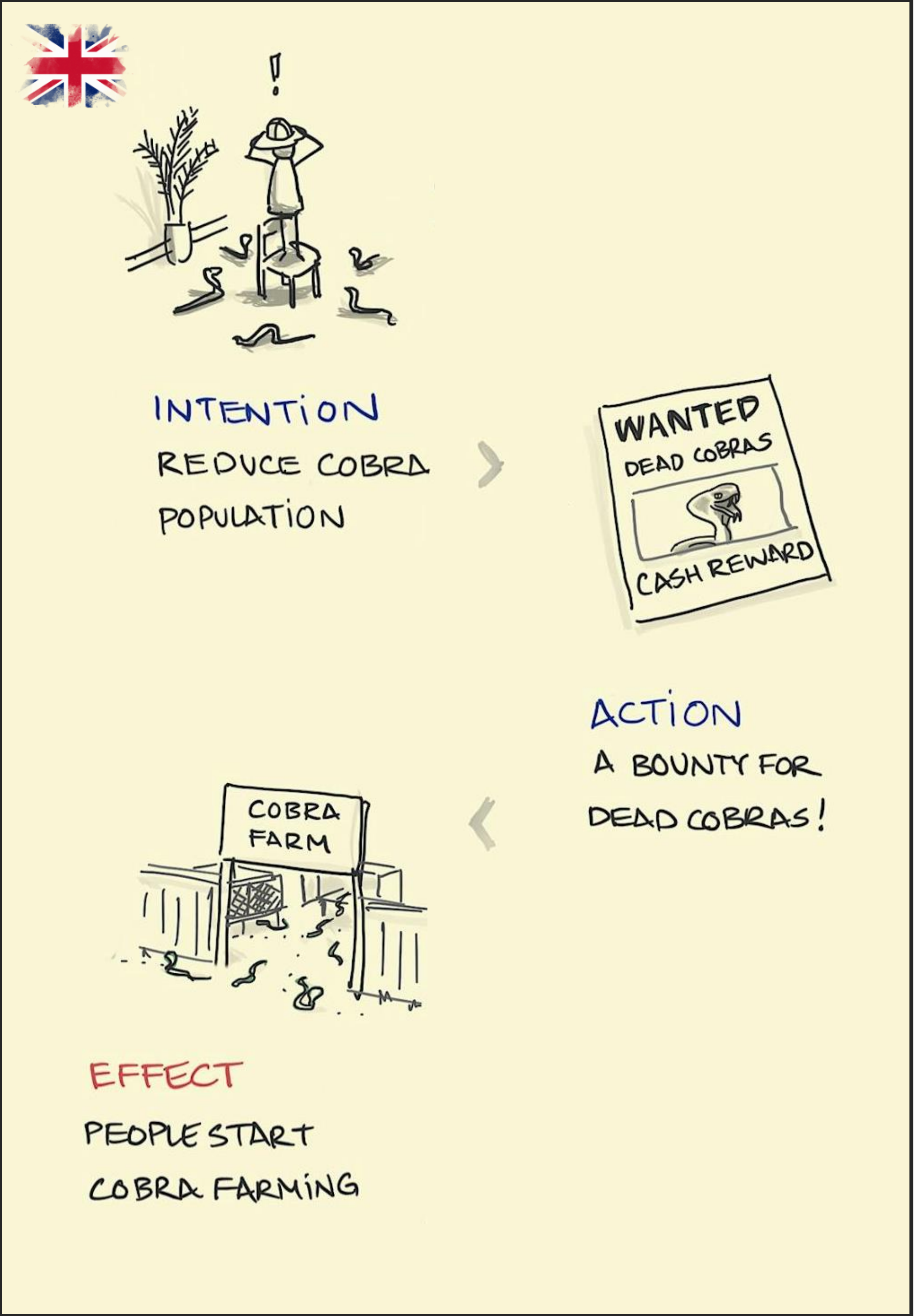
REDUCE COBRA POPULATION

ACTION

A BOUNTY FOR DEAD COBRAS!

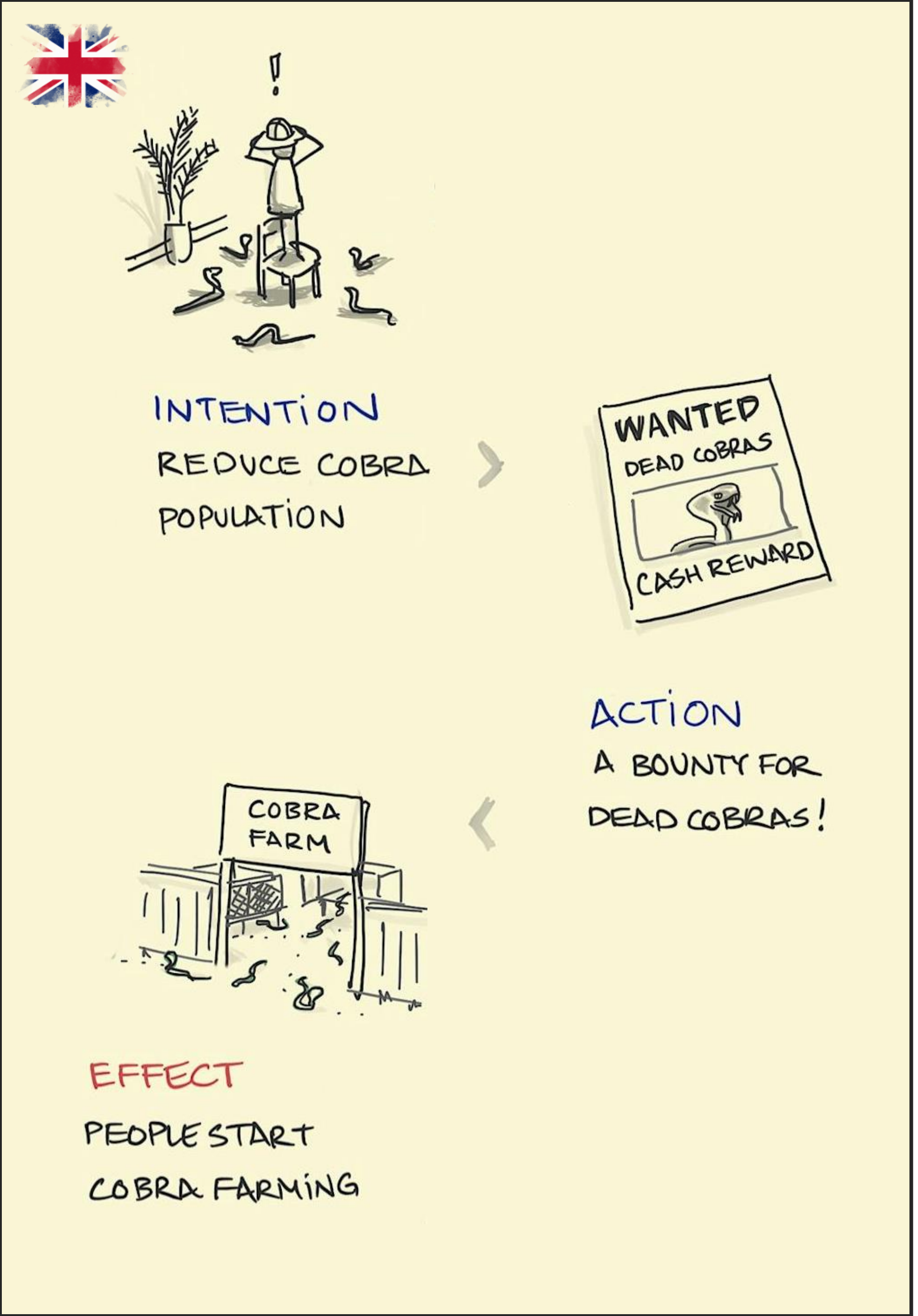
EFFECT

PEOPLE START COBRA FARMING



"Anything that can go wrong will go wrong.

Murphy's Law



"Any system that can be gamed will be gamed.

W. Brian Arthur

"Any system that can be gamed will be gamed.

W. Brian Arthur

"Any system that can be gamed will be gamed.

W. Brian Arthur

"Any system that can be gamed will be gamed.

W. Brian Arthur

"Any system that can be gamed will be gamed.

W. Brian Arthur

"Any system that can be gamed will be gamed.

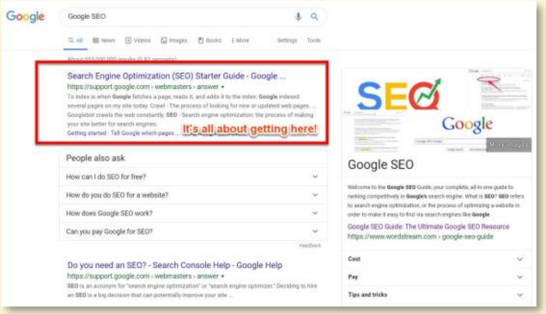
W. Brian Arthur

"Any system that can be gamed will be gamed.

W. Brian Arthur

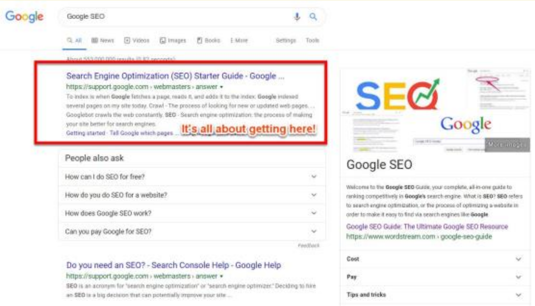
"Any system that can be gamed will be gamed.

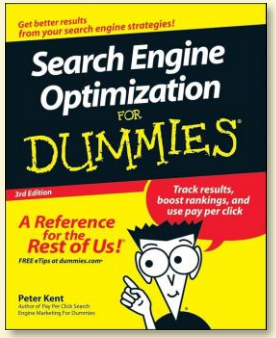
W. Brian Arthur



"Any system that can be gamed will be gamed.

W. Brian Arthur





"Any system that can be gamed will be gamed.

W. Brian Arthur

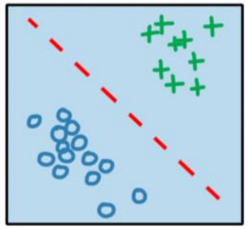
We want to design ***ML algorithms*** *that are* ***(1) robust*** *against* ***strategic behavior***

and ***(2) incentivize agent behavior*** *that is aligned with the* ***system's goals****.*

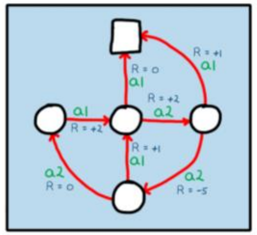
*We want to design **ML algorithms** that are **(1) robust** against **strategic behavior***

*<u>and</u> **(2) incentivize agent behavior** that is aligned with the **system's goals**.*

Supervised Learning



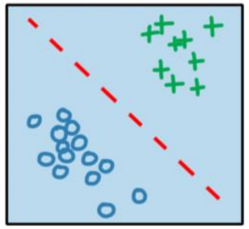
Reinforcement Learning



We want to design ***ML algorithms*** *that are* ***(1) robust*** *against* ***strategic behavior***

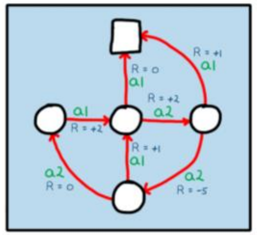
and ***(2) incentivize agent behavior*** *that is aligned with the* ***system's goals****.*

Supervised Learning



Adversarial Robustness

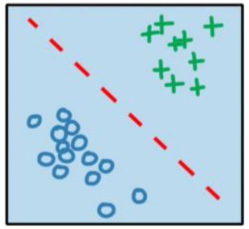
Reinforcement Learning



Adversarial Robustness

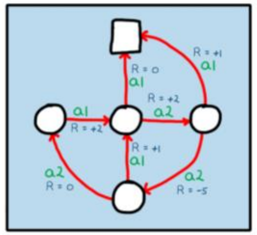
*We want to design **ML algorithms** that are **(1) robust** against **strategic behavior** and **(2) incentivize agent behavior** that is aligned with the **system's goals**.*

Supervised Learning



Adversarial Robustness

Reinforcement Learning

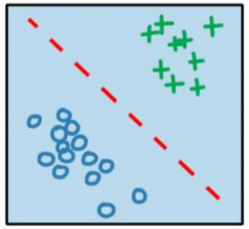


Adversarial Robustness

We want to design ***ML algorithms*** *that are* ***(1) robust*** *against* ***strategic behavior***

<u>and</u> ***(2) incentivize agent behavior*** *that is aligned with the* ***system's goals****.*

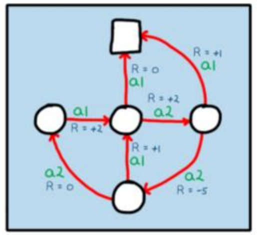
Supervised Learning



Adversarial Robustness

<u>Strategic Classification</u>

Reinforcement Learning

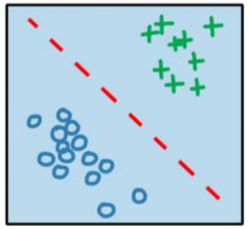


Adversarial Robustness

Corruption-Robust RL

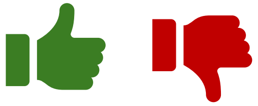
*We want to design **ML algorithms** that are **(1)** **robust** against **strategic behavior** and **(2)** **incentivize agent behavior** that is aligned with the **system's goals**.*

Supervised Learning



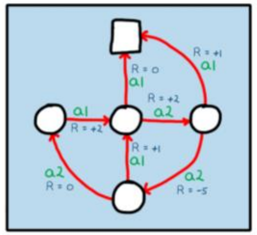
Adversarial Robustness

Strategic Classification



(1) ***(2)***

Reinforcement Learning

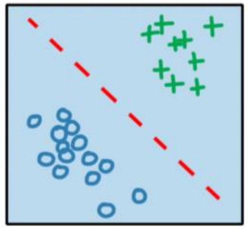


Adversarial Robustness

Corruption-Robust RL

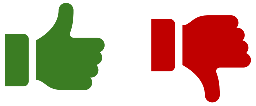
*We want to design **ML algorithms** that are **(1)** **robust** against **strategic behavior** <u>and</u> **(2)** **incentivize** **agent behavior** that is aligned with the **system's goals**.*

Supervised Learning



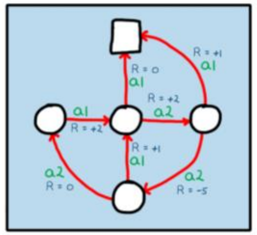
Adversarial Robustness

<u>Strategic Classification</u>



(1) ***(2)***

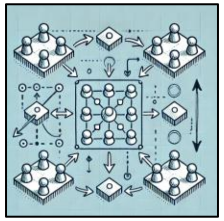
Reinforcement Learning



Adversarial Robustness

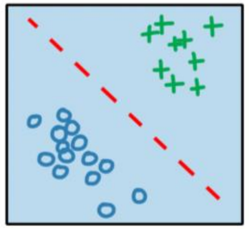
Corruption-Robust RL

Mechanism Design



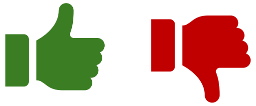
We want to design ***ML algorithms*** *that are* ***(1) robust*** *against* ***strategic behavior*** *and* ***(2) incentivize agent behavior*** *that is aligned with the* ***system's goals***.

Supervised Learning



Adversarial Robustness

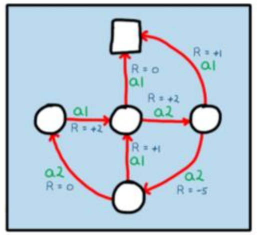
Strategic Classification



(1) ***(2)***

Strategic Interactive Decision-Making

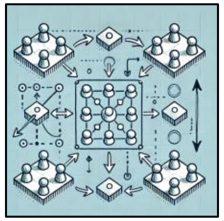
Reinforcement Learning



Adversarial Robustness

Corruption-Robust RL

Mechanism Design



Strategic Linear Contextual Bandits

joint work with Aadirupa Saha, Christos Dimitrakakis, Haifeng Xu

recommends the channels' content

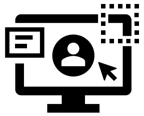
maximizing individual exposure / profit

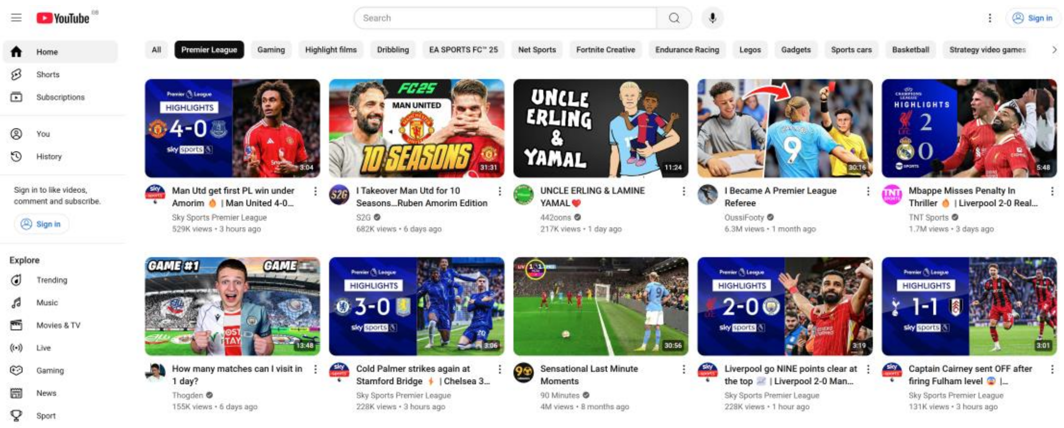
maximizing platform performance

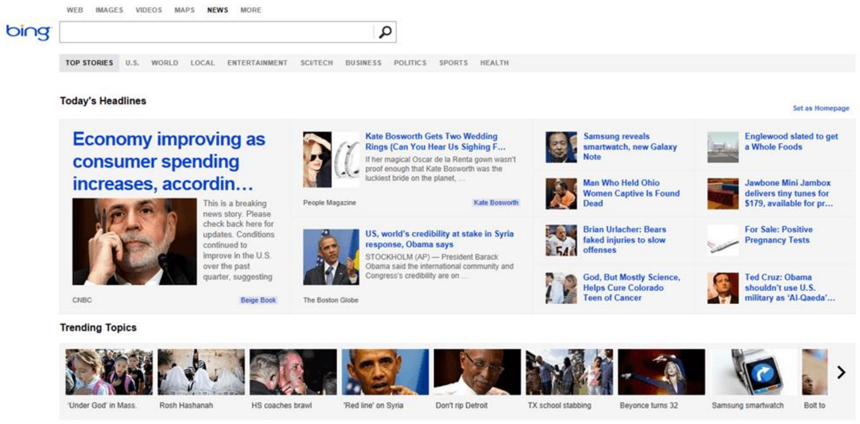
(1) make good recommendations
(2) incentivize good content / truthfulness

report content and data strategically

Linear Contextual Bandits

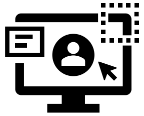


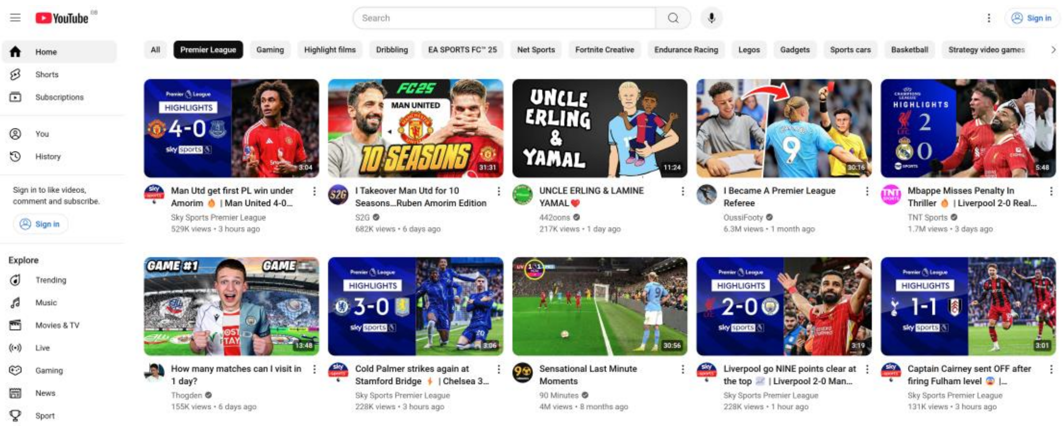


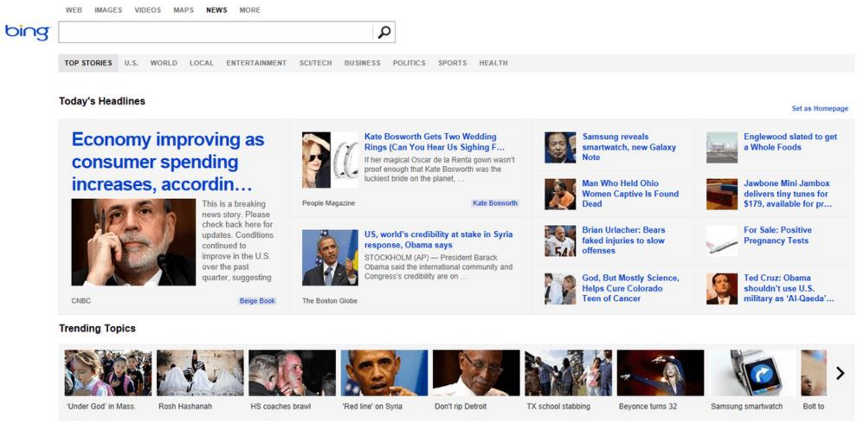


Linear Contextual Bandits

- T rounds, K arms ()



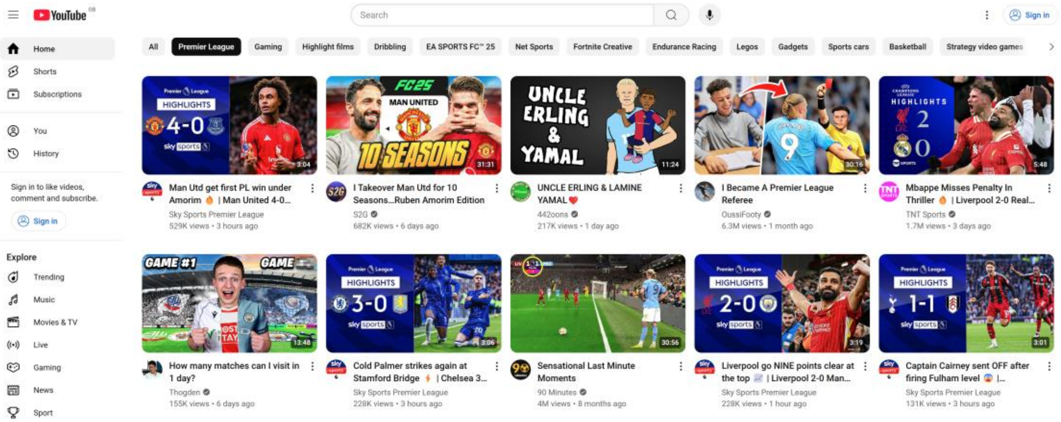


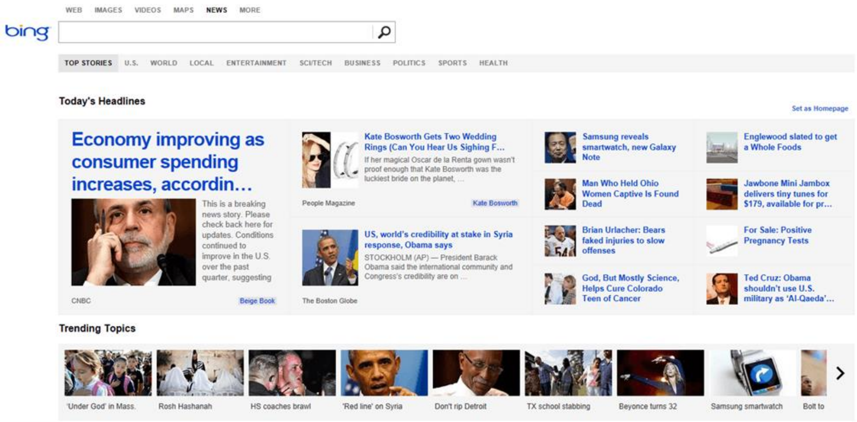


Linear Contextual Bandits

- T rounds, K arms

For $t = 1, \ldots, T$:



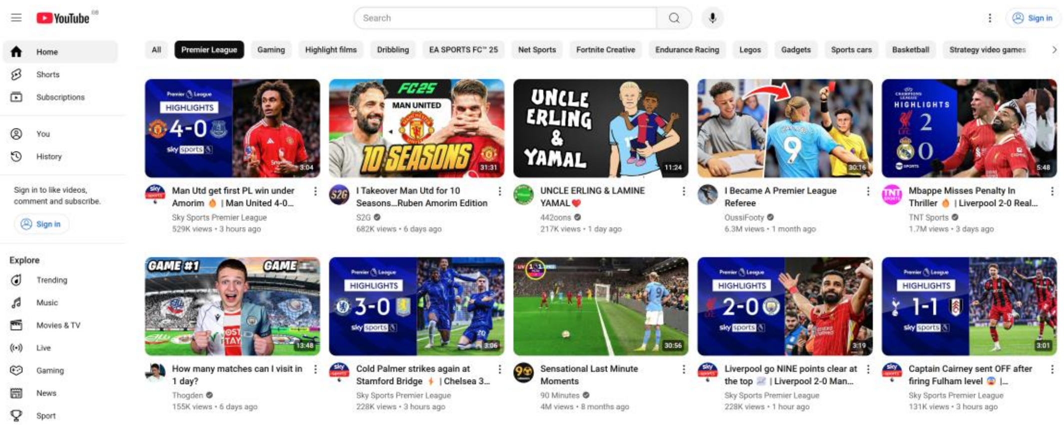


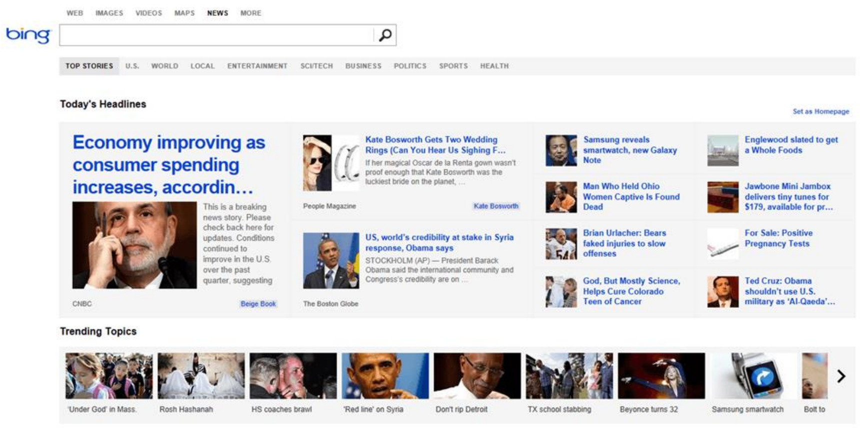
Linear Contextual Bandits

- T rounds, K arms

For $t = 1, \ldots, T$:

1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$





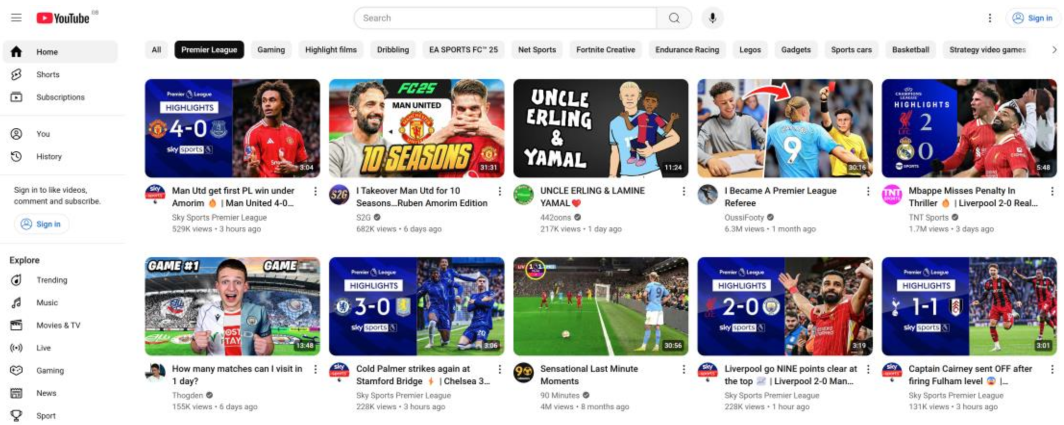
Linear Contextual Bandits

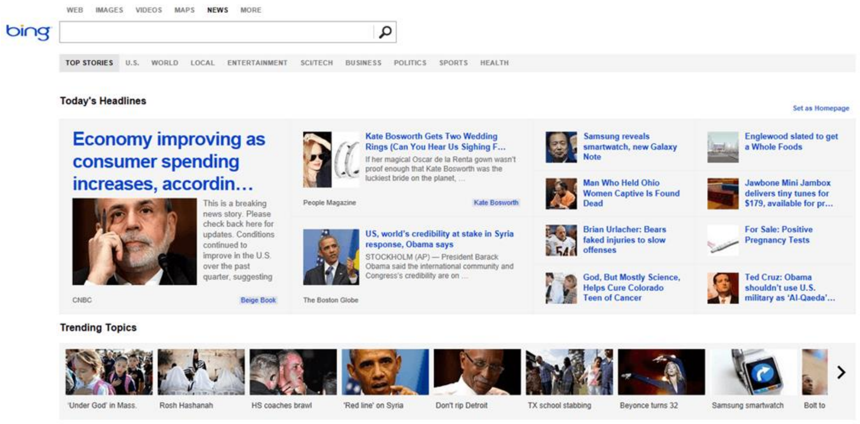
- T rounds, K arms ( )

For $t = 1, \ldots, T$:

1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$ $, \ldots ,$ $x^*_{t,K} =$ $\in \mathbb{R}^d$

2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) :=$





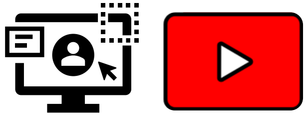
Linear Contextual Bandits

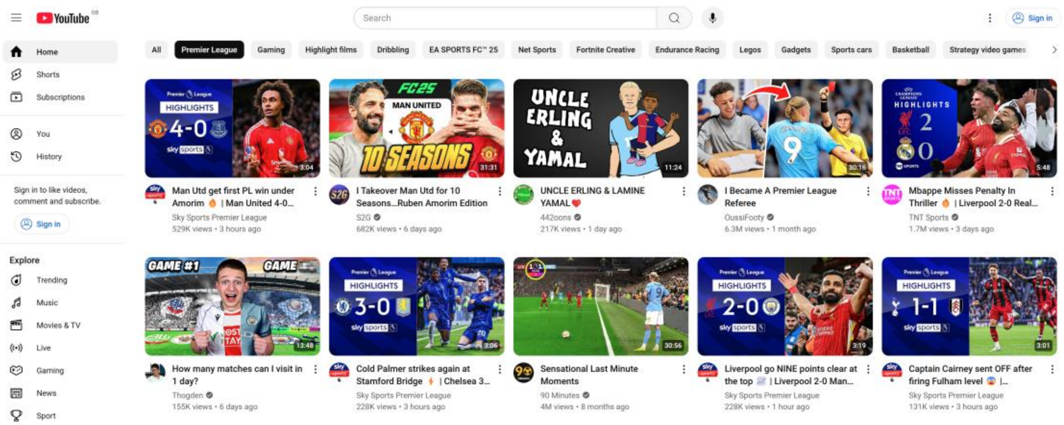
- T rounds, K arms ( )

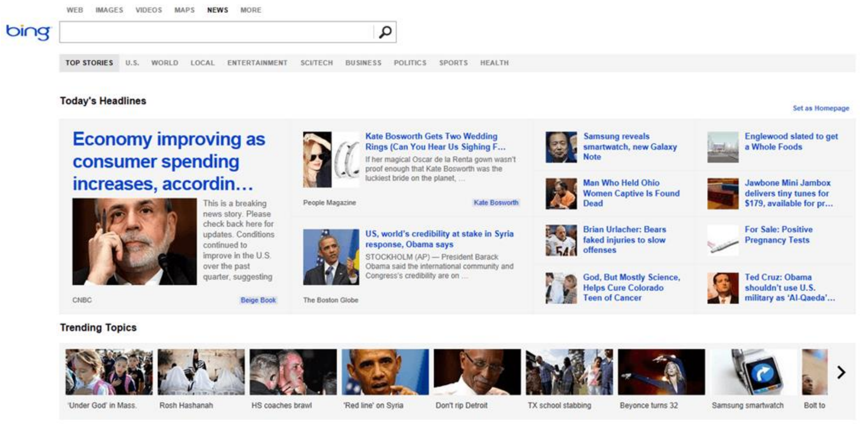
For $t = 1, \dots, T$:

1) **Algorithm** observes **arm-specific** contexts $x_{t,1}^* =$, … , $x_{t,K}^* =$ $\in \mathbb{R}^d$

2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r_t^*(i_t) :=$





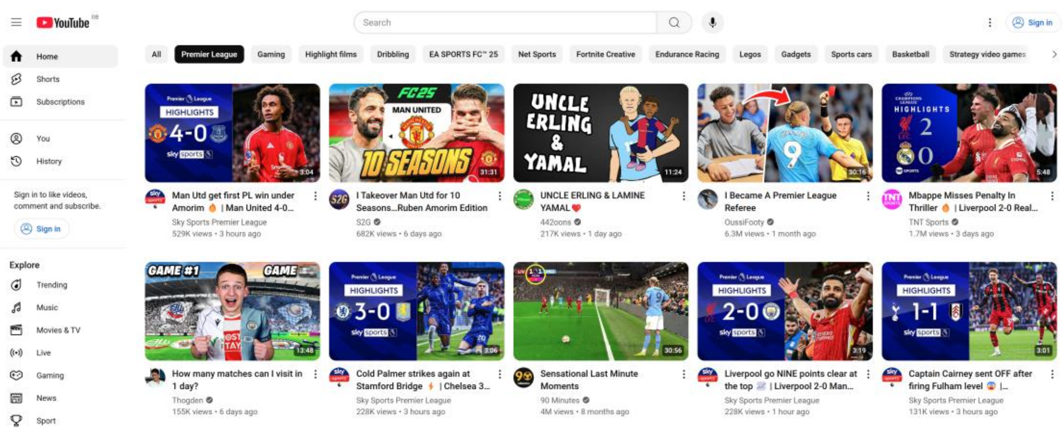


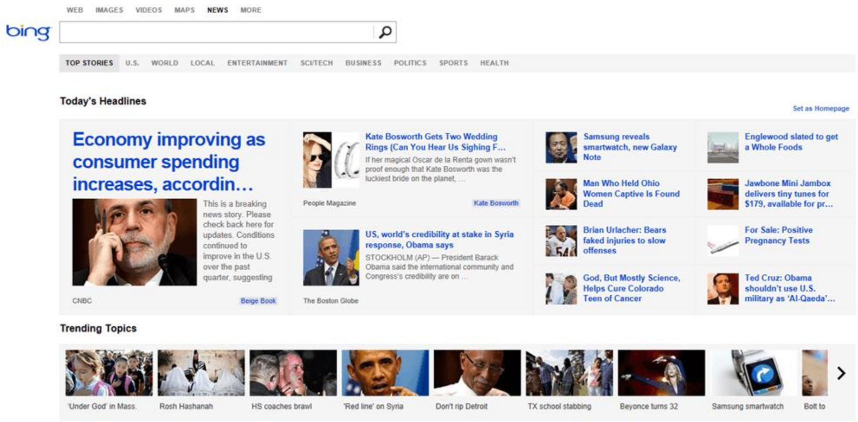
Linear Contextual Bandits

- T rounds, K arms

For $t = 1, \ldots, T$:

1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$
2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) :=$





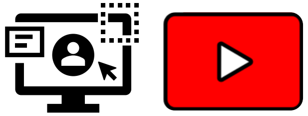
Linear Contextual Bandits

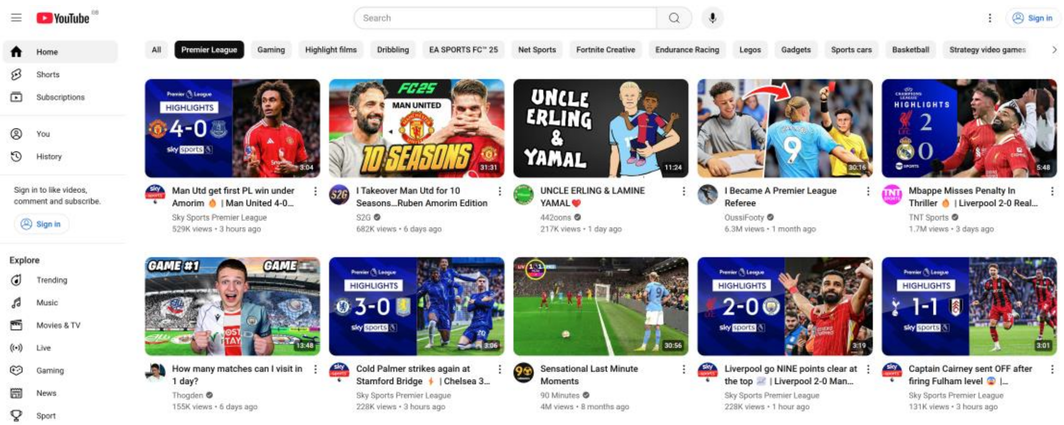
- T rounds, K arms ( )

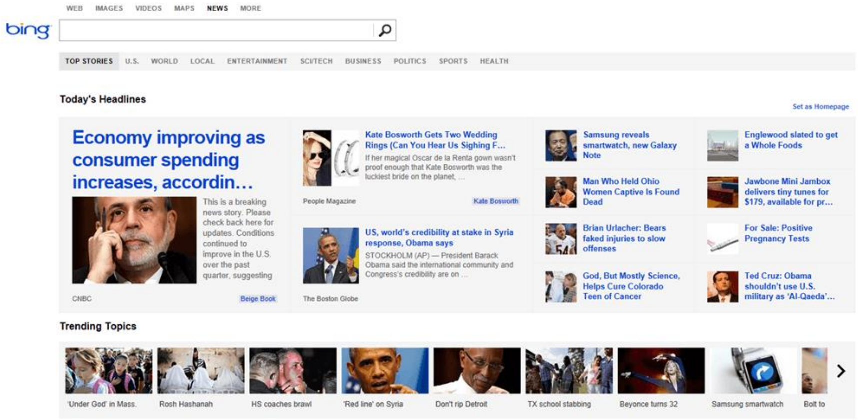
For $t = 1, \ldots, T$:

1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$

2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$







Linear Contextual Bandits

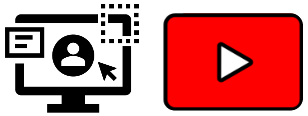
- T rounds, K arms ( )

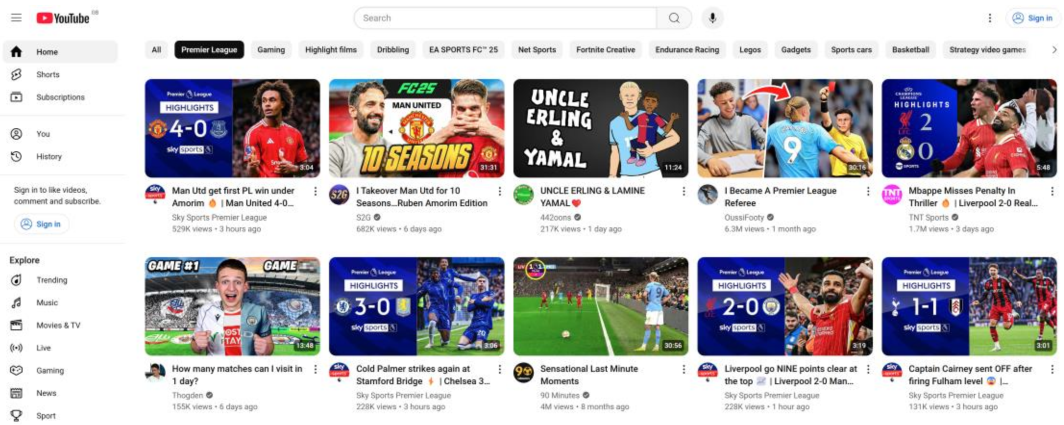
For $t = 1, \dots, T$:

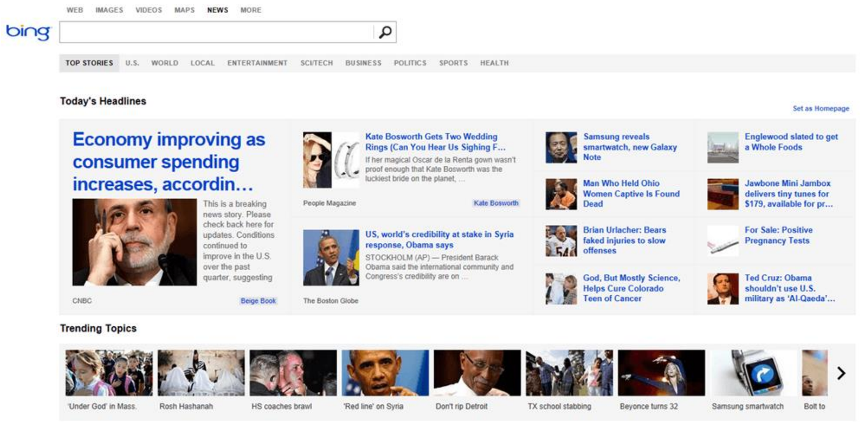
1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$

2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

unknown







Linear Contextual Bandits

- T rounds, K arms ( )

For $t = 1, \dots, T$:

1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$
2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

unknown

Linear Contextual Bandits

- T rounds, K arms ( )

For $t = 1, \dots, T$:

1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$
2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

unknown

Algorithm maximizes cumulative reward

$$\sum_{t=1}^{T} r_t(i_t)$$

Linear Contextual Bandits

- T rounds, K arms ( )

For $t = 1, \dots, T$:

1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$

2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r_t^*(i_t) := \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

unknown

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Linear Contextual Bandits

- T rounds, K arms ( )

For $t = 1, \ldots, T$:

private information

1) **Algorithm** observes **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$

2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) := \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

unknown

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Linear Contextual Bandits

- T rounds, K arms ( )

For $t = 1, \dots, T$:

1) **Algorithm** ~~observes~~ **arm-specific** contexts $x^*_{t,1} =$, … , $x^*_{t,K} =$ $\in \mathbb{R}^d$

private information

2) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

unknown

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Strategic Linear Contextual Bandits

arm = strategic agent 

For $t = 1, \ldots, T$:

1) Every **arm i** $\in [K]$ **privately** observes its context $x^*_{t,i} =$ $\in \mathbb{R}^d$

2) Every **arm i** $\in [K]$ reports a **gamed** context $x_{t,i} =$ $\in \mathbb{R}^d$ to the **Algorithm**

3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

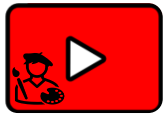
Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Strategic Linear Contextual Bandits

arm = strategic agent 

For $t = 1, \dots, T$:

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$
2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed** context $x_{t,i} =$ $\in \mathbb{R}^d$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

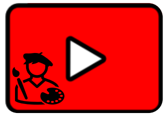
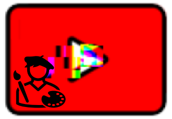
Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Strategic Linear Contextual Bandits

arm = strategic agent 

For $t = 1, \dots, T$:

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$
2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

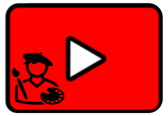
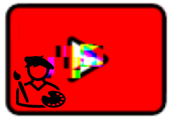
Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Strategic Linear Contextual Bandits

arm = strategic agent 

For $t = 1, \ldots, T$:

1) Every **arm i** $\in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$
2) Every **arm i** $\in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

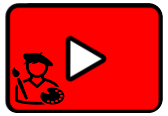
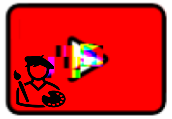
Every Arm i maximizes its **#selections**

$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \ldots, T$:

1) Every **arm *i*** $\in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$
2) Every **arm *i*** $\in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) := \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

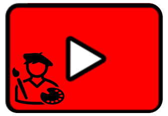
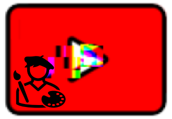
Every Arm *i* maximizes its **#selections**

$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \dots, T$: ↙ *repeated interaction*

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$
2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) := \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

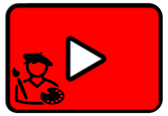
Every Arm $\boldsymbol{i}$ maximizes its **#selections**

$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

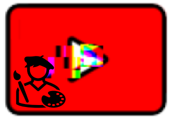
Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \dots, T$: ↙ ***repeated interaction***

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes its context $x^*_{t,i}$ =  $\in \mathbb{R}^d$

unbounded manipulation

2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed** context $x_{t,i}$ =  $\in \mathbb{R}^d$ to the **Algorithm**

3) **Algorithm** plays arm i_t =  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

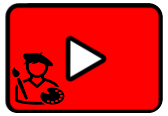
Every Arm $\boldsymbol{i}$ maximizes its **#selections**

$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

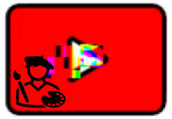
Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \ldots, T$: ↙ *repeated interaction*

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$

unbounded manipulation

2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**

3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

↗ *multiple competing agents*

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Every Arm $\boldsymbol{i}$ maximizes its **#selections**

$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \ldots, T$: ↙ *repeated interaction*

1) Every **arm *i*** $\in [K]$ **privately** observes its context $x^*_{t,i} =$ $\in \mathbb{R}^d$

unbounded manipulation

2) Every **arm *i*** $\in [K]$ reports a **gamed** context $x_{t,i} =$ $\in \mathbb{R}^d$ to the **Algorithm**

3) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

↗ *multiple competing agents*

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Every Arm *i* maximizes its **#selections**

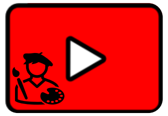
$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

far-sighted agents

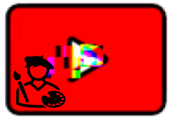
Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \ldots, T$: ↙ *repeated interaction*

1) Every **arm i** $\in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$

unbounded manipulation

2) Every **arm i** $\in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**

3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

↗ *multiple competing agents* 

↗ *unknown environment* 

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Every Arm i maximizes its **#selections**

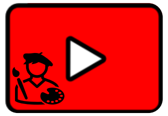
$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

far-sighted agents

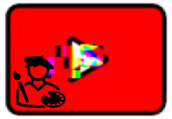
Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \ldots, T$: ↙ ***repeated interaction***

1) Every **arm *i*** $\in [K]$ **privately** observes its context $x_{t,i}^* =$  $\in \mathbb{R}^d$

unbounded manipulation

2) Every **arm *i*** $\in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**

3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r_t^*(i_t) := \langle \theta^*, x_{t,i_t}^* \rangle + \eta_t$

↗ ***multiple competing agents***

↗ ***unknown environment***

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x_{t,i}^* \rangle - \langle \theta^*, x_{t,i_t}^* \rangle\right]$$

Every Arm *i* maximizes its **#selections**

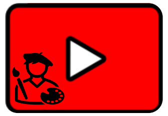
$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

far-sighted agents

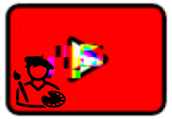
Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**: **arm strategies** $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \ldots, T$: ↙ *repeated interaction*

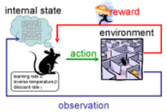
1) Every **arm i** $\in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$

unbounded manipulation

2) Every **arm i** $\in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**

3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

multiple competing agents 

known environment 

Algorithm minimizes expected regret

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \max_{i \in [K]} \langle \theta^*, x^*_{t,i} \rangle - \langle \theta^*, x^*_{t,i_t} \rangle\right]$$

Every Arm i maximizes its **#selections**

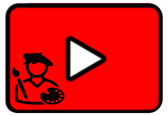
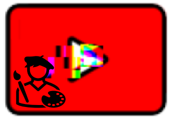
$$\mathbb{E}\left[\sum_{t=1}^{T} 1(i_t = i)\right]$$

far-sighted agents

Strategic Linear Contextual Bandits

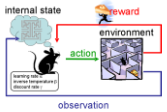
Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \ldots, T$:

1) Every **arm i** $\in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$
2) Every **arm i** $\in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

↗

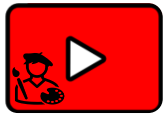
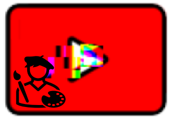
known *environment*



Strategic Linear Contextual Bandits

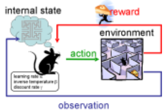
Arms respond in **Equilibrium**:
arm strategies $\in$ $\mathbf{NE}$(**Algorithm**)

For $t = 1, \dots, T$:

1) Every **arm i** $\in [K]$ **privately** observes its context $x^*_{t,i} =$  $\in \mathbb{R}^d$
2) Every **arm i** $\in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

↗

known *environment*

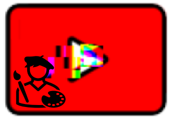


- Let's swap features $x \in \mathbb{R}^d$ for the implied **avg reward μ** $\coloneqq \langle \theta^*, x \rangle \in \mathbb{R}$

Strategic Linear Contextual Bandits

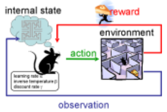
Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\mathrm{Algorithm})$

For $t = 1, \ldots, T$:

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes **true avg reward** $\boldsymbol{\mu}^*_{t,i} =$  $\in \mathbb{R}$
2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed** context $x_{t,i} =$  $\in \mathbb{R}^d$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r_t^*(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

↗

known *environment*



- Let's swap features $x \in \mathbb{R}^d$ for the implied **avg reward** $\boldsymbol{\mu} \coloneqq \langle \theta^*, x \rangle \in \mathbb{R}$

Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\mathbf{Algorithm})$

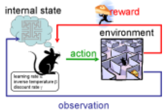
For $t = 1, \ldots, T$:

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes **true avg reward** $\boldsymbol{\mu}^*_{t,i}$ =  $\in \mathbb{R}$

2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed** **value** $\boldsymbol{\mu}_{t,i}$ =  $\in \mathbb{R}$ to the **Algorithm**

3) **Algorithm** plays arm i_t =  $\in [K]$ and receives reward $r_t^*(i_t) \coloneqq \langle \theta^*, x^*_{t,i_t} \rangle + \eta_t$

↗ *known environment*



- Let's swap features $x \in \mathbb{R}^d$ for the implied **avg reward** $\boldsymbol{\mu} \coloneqq \langle \theta^*, x \rangle \in \mathbb{R}$

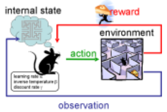
Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \ldots, T$:

1) Every **arm i** $\in [K]$ **privately** observes **true avg reward $\mu^*_{t,i}$** $=$  $\in \mathbb{R}$
2) Every **arm i** $\in [K]$ reports a **gamed value $\mu_{t,i}$** $=$  $\in \mathbb{R}$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) := \mu^*_{t,i_t} + \eta_t$

known *environment* ↗



- Let's swap features $x \in \mathbb{R}^d$ for the implied **avg reward μ** $:= \langle \theta^*, x \rangle \in \mathbb{R}$

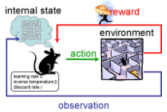
Strategic Linear Contextual Bandits

Arms respond in **Equilibrium:**
arm strategies $\in \mathbf{NE}(\mathrm{Algorithm})$

For $t = 1, \ldots, T$:

1) Every **arm i** $\in [K]$ **privately** observes **true avg reward $\mu^*_{t,i}$** $=$  $\in \mathbb{R}$
2) Every **arm i** $\in [K]$ reports a **gamed value $\mu_{t,i}$** $=$  $\in \mathbb{R}$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \mu^*_{t,i_t} + \eta_t$

↗
known *environment*



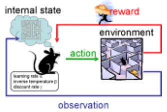
- Let's swap features $x \in \mathbb{R}^d$ for the implied **avg reward μ** $\coloneqq \langle \theta^*, x \rangle \in \mathbb{R}$
- We're happy when the arms "cooperate" and the ***reports $\mu_{t,i}$*** match the ***truth $\mu^*_{t,i}$***

Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

For $t = 1, \dots, T$:

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes **true avg reward** $\boldsymbol{\mu^*_{t,i}} =$ $\in \mathbb{R}$
2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed** **value** $\boldsymbol{\mu_{t,i}} =$ $\in \mathbb{R}$ to the **Algorithm**
3) **Algorithm** plays arm $i_t =$ $\in [K]$ and receives reward $r^*_t(i_t) := \mu^*_{t,i_t} + \eta_t$

known *environment* 

- Let's swap features $x \in \mathbb{R}^d$ for the implied **avg reward** $\boldsymbol{\mu} := \langle \theta^*, x \rangle \in \mathbb{R}$
- We're happy when the arms "cooperate" and the ***reports*** $\boldsymbol{\mu_{t,i}}$ match the ***truth*** $\boldsymbol{\mu^*_{t,i}}$

Strategic Linear Contextual Bandits

Arms respond in **Equilibrium**:
arm strategies $\in \mathbf{NE}(\text{Algorithm})$

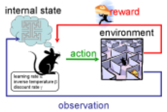
For $t = 1, \ldots, T$:

1) Every **arm** $\boldsymbol{i} \in [K]$ **privately** observes **true avg reward** $\boldsymbol{\mu^*_{t,i}}$ =  $\in \mathbb{R}$

2) Every **arm** $\boldsymbol{i} \in [K]$ reports a **gamed value** $\boldsymbol{\mu_{t,i}}$ =  $\in \mathbb{R}$ to the **Algorithm**

3) **Algorithm** plays arm i_t =  $\in [K]$ and receives reward $r^*_t(i_t) \coloneqq \mu^*_{t,i_t} + \eta_t$

↗ ***known environment***



- Let's swap features $x \in \mathbb{R}^d$ for the implied **avg reward** $\boldsymbol{\mu} \coloneqq \langle \theta^*, x \rangle \in \mathbb{R}$
- We're happy when the arms "cooperate" and the ***reports*** $\boldsymbol{\mu_{t,i}}$  match the ***truth*** $\boldsymbol{\mu^*_{t,i}}$ 

How can we incentivize the arms to "cooperate" with us while minimizing regret?

Grim Trigger Strategy

How can we incentivize the arms to "cooperate" with us?

Repeatedly:

1) each arm privately observes $\mu_{t,i}^*$

2) each arm tells us $\mu_{t,i}$

3) We observe $r_{t,i_t}^* \sim D(\mu_{t,i_t}^*)$

Grim Trigger Strategy

How can we incentivize the arms to "cooperate" with us?

Repeatedly:

1) each arm privately observes $\mu^*_{t,i}$

2) each arm tells us $\mu_{t,i}$

3) We observe $r^*_{t,i_t} \sim D(\mu^*_{t,i_t})$

Grim Trigger Strategy

How can we incentivize the arms to "cooperate" with us?

From Iterated Social Dilemmas: *"If you defect once, I will defect permanently."*

Repeatedly:

1) each arm privately observes $\mu^*_{t,i}$

2) each arm tells us $\mu_{t,i}$

3) We observe $r^*_{t,i_t} \sim D(\mu^*_{t,i_t})$

Grim Trigger Strategy

How can we incentivize the arms to "cooperate" with us?

From Iterated Social Dilemmas: *"If you defect once, I will defect permanently."*

If ***reported avg reward*** $\boldsymbol{\mu_{t,i}}$ **>** ***true avg reward*** $\boldsymbol{\mu^*_{t,i}}$*,*

eliminate arm i *forever.*

Repeatedly:

1) each arm privately observes $\mu^*_{t,i}$

2) each arm tells us $\mu_{t,i}$

3) We observe $r^*_{t,i_t} \sim D(\mu^*_{t,i_t})$

Grim Trigger Strategy

How can we incentivize the arms to "cooperate" with us?

From Iterated Social Dilemmas: *"If you defect once, I will defect permanently."*

If ***reported avg reward*** $\boldsymbol{\mu_{t,i}}$ **>** ***true avg reward*** $\boldsymbol{\mu^*_{t,i}}$*,*

eliminate arm i *forever.*

But we don't observe $\boldsymbol{\mu^*_{t,i}}$

Repeatedly:

1) each arm privately observes $\mu^*_{t,i}$

2) each arm tells us $\mu_{t,i}$

3) We observe $r^*_{t,i_t} \sim D(\mu^*_{t,i_t})$

Grim Trigger Strategy

How can we incentivize the arms to “cooperate” with us?

From Iterated Social Dilemmas: *“If you defect once, I will defect permanently.”*

If ***reported avg reward*** $\boldsymbol{\mu_{t,i}}$ **>** ***observed (true) reward*** $\boldsymbol{r^*_{t,i}}$,

eliminate arm i forever.

But we don‘t observe $\boldsymbol{\mu^*_{t,i}}$

Repeatedly:

1) each arm privately observes $\mu^*_{t,i}$

2) each arm tells us $\mu_{t,i}$

3) We observe $r^*_{t,i_t} \sim D(\mu^*_{t,i_t})$

Grim Trigger Strategy

How can we incentivize the arms to "cooperate" with us?

From Iterated Social Dilemmas: "*If you defect once, I will defect permanently.*"

If ***reported avg reward*** $\boldsymbol{\mu_{t,i}}$ **>** ***observed (true) reward*** $\boldsymbol{r^*_{t,i}}$*,*

eliminate arm i *forever.*

But $\boldsymbol{r^*_{t,i}}$ *is noisy*

Repeatedly:

1) each arm privately observes $\mu^*_{t,i}$

2) each arm tells us $\mu_{t,i}$

3) We observe $r^*_{t,i_t} \sim D(\mu^*_{t,i_t})$

Grim Trigger Strategy

How can we incentivize the arms to "cooperate" with us?

From Iterated Social Dilemmas: *"If you defect once, I will defect permanently."*

If ***reported avg reward*** $\sum_t \mu_{t,i}$ **>** ***optimistic estimate*** $\sum_t \boldsymbol{r}^*_{t,i} + \sqrt{\boldsymbol{n}_\tau(\boldsymbol{i})}$

eliminate arm i forever.

Repeatedly:

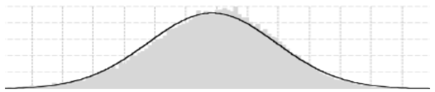
1) each arm privately observes $\mu^*_{t,i}$

2) each arm tells us $\mu_{t,i}$

3) We observe $r^*_{t,i_t} \sim D(\mu^*_{t,i_t})$

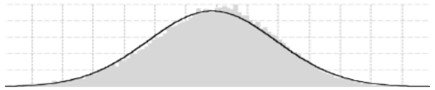
What does the Grim Trigger mean for the arms?

What does the Grim Trigger mean for the arms?



The observations $\sum_t r^*_{t,i}$ *concentrate* around the average $\sum_t \mu^*_{t,i}$

What does the Grim Trigger mean for the arms?



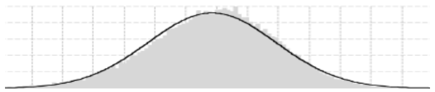
The observations $\sum_t r_{t,i}^*$ *concentrate* around the average $\sum_t \mu_{t,i}^*$

This implies that with high probability:

If $\sum_t \boldsymbol{\mu_{t,i}} - \sum_t \boldsymbol{\mu_{t,i}^*} > \mathbf{2}\sqrt{\boldsymbol{n_\tau(i)}}$, *we eliminate arm i.* 

reports* − *truth

What does the Grim Trigger mean for the arms?



The observations $\sum_t r^*_{t,i}$ *concentrate* around the average $\sum_t \mu^*_{t,i}$

This implies that with high probability:

If $\sum_t \boldsymbol{\mu_{t,i}} - \sum_t \boldsymbol{\mu^*_{t,i}} > \mathbf{2}\sqrt{\boldsymbol{n_\tau(i)}}$*, we eliminate arm* i*.*

reports* − *truth

The arms can misreport to us … but not too often !

Greedy Grim Trigger Mechanism (**GGTM**)

Greedy Grim Trigger Mechanism (**GGTM**)

Greedy Selection:

Play $i_t = \arg\max_{i \in \text{alive}} \mu_{t,i}$

$\oplus$

Grim Trigger: 

If $\sum_t \mu_{t,i} > \sum_t r^*_{t,i} + \sqrt{n_\tau(i)}$, eliminate arm i.

Greedy Grim Trigger Mechanism (GGTM)

Greedy Selection:

Play $i_t = \arg\max_{i \in \text{alive}} \mu_{t,i}$

$\oplus$

Grim Trigger: 

If $\sum_t \mu_{t,i} > \sum_t r^*_{t,i} + \sqrt{n_\tau(i)}$, eliminate arm i.

Theoretical Results:

Greedy Grim Trigger Mechanism (GGTM)

Greedy Selection:

Play $i_t = \arg\max_{i \in \text{alive}} \mu_{t,i}$

$\oplus$

Grim Trigger: 

If $\sum_t \mu_{t,i} > \sum_t r^*_{t,i} + \sqrt{n_\tau(i)}$, eliminate arm i.

Theoretical Results:

1) Under **GGTM**, being **truthful** is a $\sqrt{T}$-Nash Equilibrium.

Greedy Grim Trigger Mechanism (**GGTM**)

Greedy Selection:

Play $i_t = \arg\max_{i \in \text{alive}} \mu_{t,i}$

$\oplus$

Grim Trigger: 

If $\sum_t \mu_{t,i} > \sum_t r^*_{t,i} + \sqrt{n_\tau(i)}$, eliminate arm i.

Theoretical Results:

1) Under **GGTM**, being **truthful** is a $\sqrt{T}$-Nash Equilibrium.
2) If the arms play any NE under **GGTM**, then $R_T(\text{GGTM}) \leq K\sqrt{T} + K^2\sqrt{KT}$

Greedy Grim Trigger Mechanism (GGTM)

Greedy Selection:

Play $i_t = \arg\max_{i \in \text{alive}} \mu_{t,i}$

$\oplus$

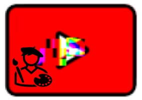
Grim Trigger: 

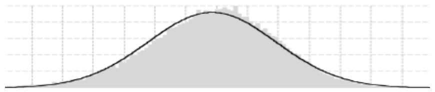
If $\sum_t \mu_{t,i} > \sum_t r^*_{t,i} + \sqrt{n_\tau(i)}$, eliminate arm i.

Theoretical Results:

1) Under **GGTM**, being **truthful** is a $\sqrt{T}$-Nash Equilibrium.
2) If the arms play any NE under **GGTM**, then $R_T(\textbf{GGTM}) \le K\sqrt{T} + K^2\sqrt{KT}$

cost of manipulation
& uncertainty





Greedy Grim Trigger Mechanism (GGTM)

Greedy Selection:

Play $i_t = \arg\max_{i \in \text{alive}} \mu_{t,i}$

$\oplus$

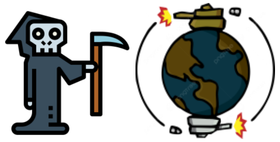
Grim Trigger: 

If $\sum_t \mu_{t,i} > \sum_t r_{t,i}^* + \sqrt{n_\tau(i)}$, eliminate arm i.

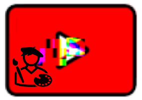
Theoretical Results:

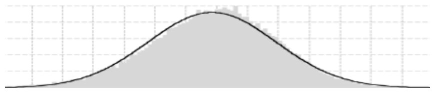
1) Under **GGTM**, being **truthful** is a $\sqrt{T}$-Nash Equilibrium.
2) If the arms play any NE under **GGTM**, then $R_T(\text{GGTM}) \leq K\sqrt{T} + K^2\sqrt{KT}$

cost of mechanism design



cost of manipulation
& uncertainty





Suppose the environment $\boldsymbol{\theta}^*$ is **unknown ...**

Things get complicated ...

We observe: gamed context $\boldsymbol{x}_{t,i}$ and reward $\boldsymbol{r}^*_{t,i} := \langle \boldsymbol{\theta}^*, \boldsymbol{x}^*_{t,i} \rangle + \eta_t$

We don't observe: true context $\boldsymbol{x}^*_{t,i}$ and parameter $\boldsymbol{\theta}^*$

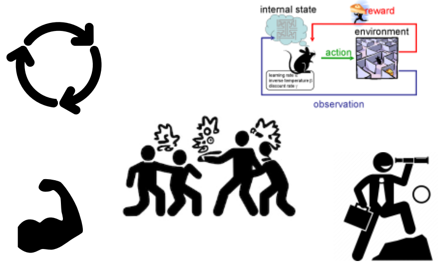
The arms can manipulate our estimate of $\boldsymbol{\theta}^*$ ...

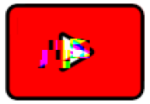
Estimating $\boldsymbol{\theta}^*$ accurately becomes **impossible**?!

Another time ...

Short Recap

- Strategic Interactive Decision-Making
  - **Reinforcement Learning** + **Mechanism Design**
  - Objective: **Strategic Robustness** + **Incentive Alignment**



- Strategic Linear Contextual Bandits
  - Strategic agents  manipulating contexts 
  - Grim Trigger Mechanism  from Iterated Social Dilemmas
  - Mechanism Design becomes *approximate* 

- *There are many more problems like this left to study ...*