Characterizing Truthful Multi-Armed Bandit Mechanisms

Moshe Babaioff
Microsoft Research

Yogi Sharma
Cornell University

Alex Slivkins
Microsoft Research

The 2nd EaGL Theory of Computation Workshop
October, 2009

Characterizing Truthful Multi-Armed Bandit Mechanisms

Moshe Babaioff
Microsoft Research

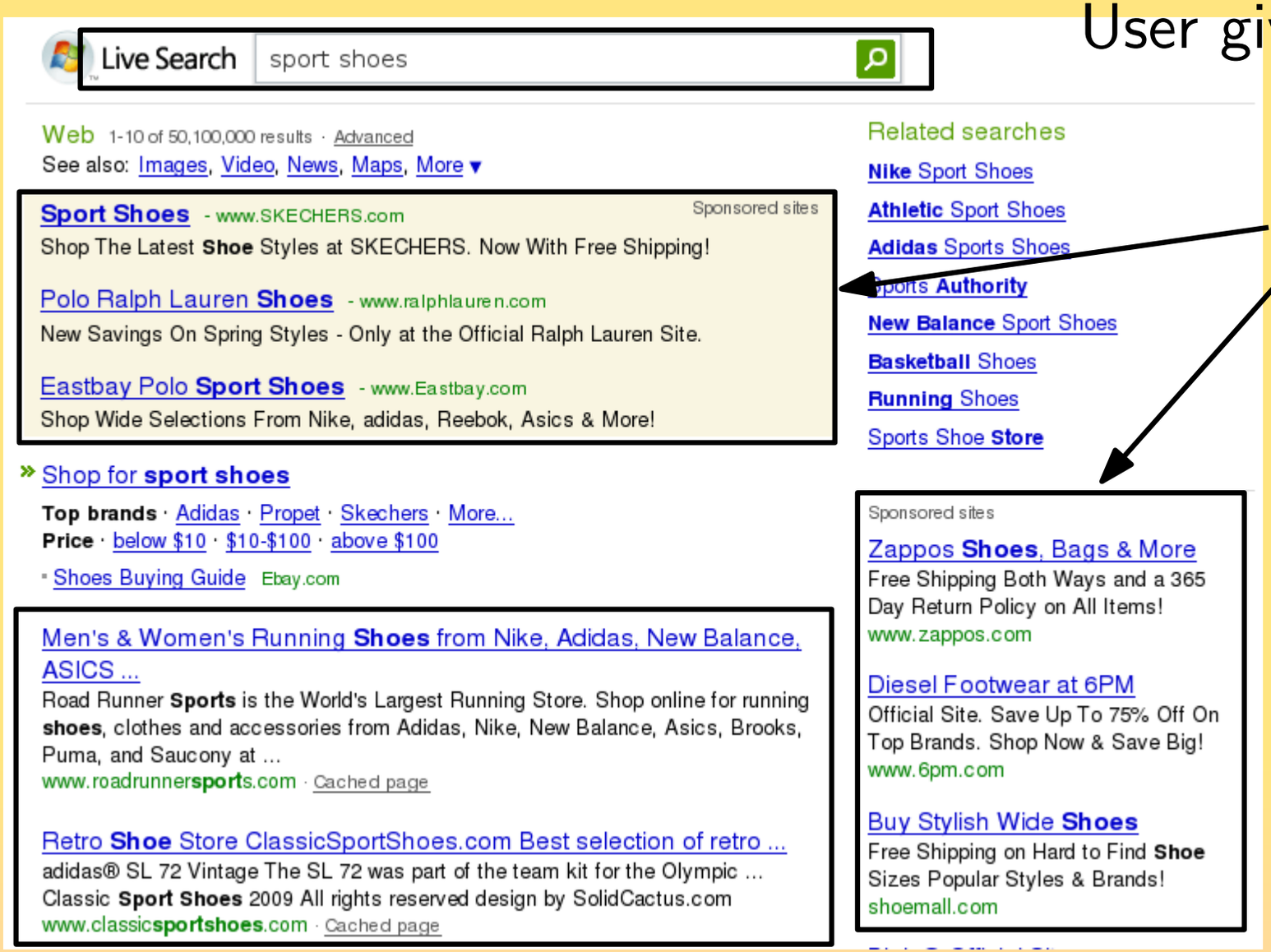
Yogi Sharma
Cornell University

Alex Slivkins
Microsoft Research

The 2nd EaGL Theory of Computation Workshop
October, 2009

Online advertising

User gives search query



Advertisements

Search results

Online advertising

User gives search query

Live Search | sport shoes

Web 1-10 of 50,100,000 results · Advanced
See also: Images, Video, News, Maps, More ▾

Sponsored sites

Sport Shoes - www.SKECHERS.com
Shop The Latest **Shoe** Styles at SKECHERS. Now With Free Shipping!

Polo Ralph Lauren **Shoes** - www.ralphlauren.com
New Savings On Spring Styles - Only at the Official Ralph Lauren Site.

Eastbay Polo **Sport Shoes** - www.Eastbay.com
Shop Wide Selections From Nike, adidas, Reebok, Asics & More!

» Shop for **sport shoes**

Top brands · Adidas · Propet · Skechers · More...
Price · below $10 · $10-$100 · above $100

▪ Shoes Buying Guide Ebay.com

Men's & Women's Running **Shoes** from Nike, Adidas, New Balance, ASICS ...
Road Runner **Sports** is the World's Largest Running Store. Shop online for running **shoes**, clothes and accessories from Adidas, Nike, New Balance, Asics, Brooks, Puma, and Saucony at ...
www.roadrunner**sport**s.com · Cached page

Retro **Shoe** Store ClassicSportShoes.com Best selection of retro ...
adidas® SL 72 Vintage The SL 72 was part of the team kit for the Olympic ... Classic **Sport Shoes** 2009 All rights reserved design by SolidCactus.com
www.classic**sportshoes**.com · Cached page

Related searches

Nike Sport Shoes
Athletic Sport Shoes
Adidas Sports Shoes
Sports Authority
New Balance Sport Shoes
Basketball Shoes
Running Shoes
Sports Shoe **Store**

Sponsored sites

Zappos **Shoes**, Bags & More
Free Shipping Both Ways and a 365 Day Return Policy on All Items!
www.zappos.com

Diesel Footwear at 6PM
Official Site. Save Up To 75% Off On Top Brands. Shop Now & Save Big!
www.6pm.com

Buy Stylish Wide **Shoes**
Free Shipping on Hard to Find **Shoe** Sizes Popular Styles & Brands!
shoemall.com

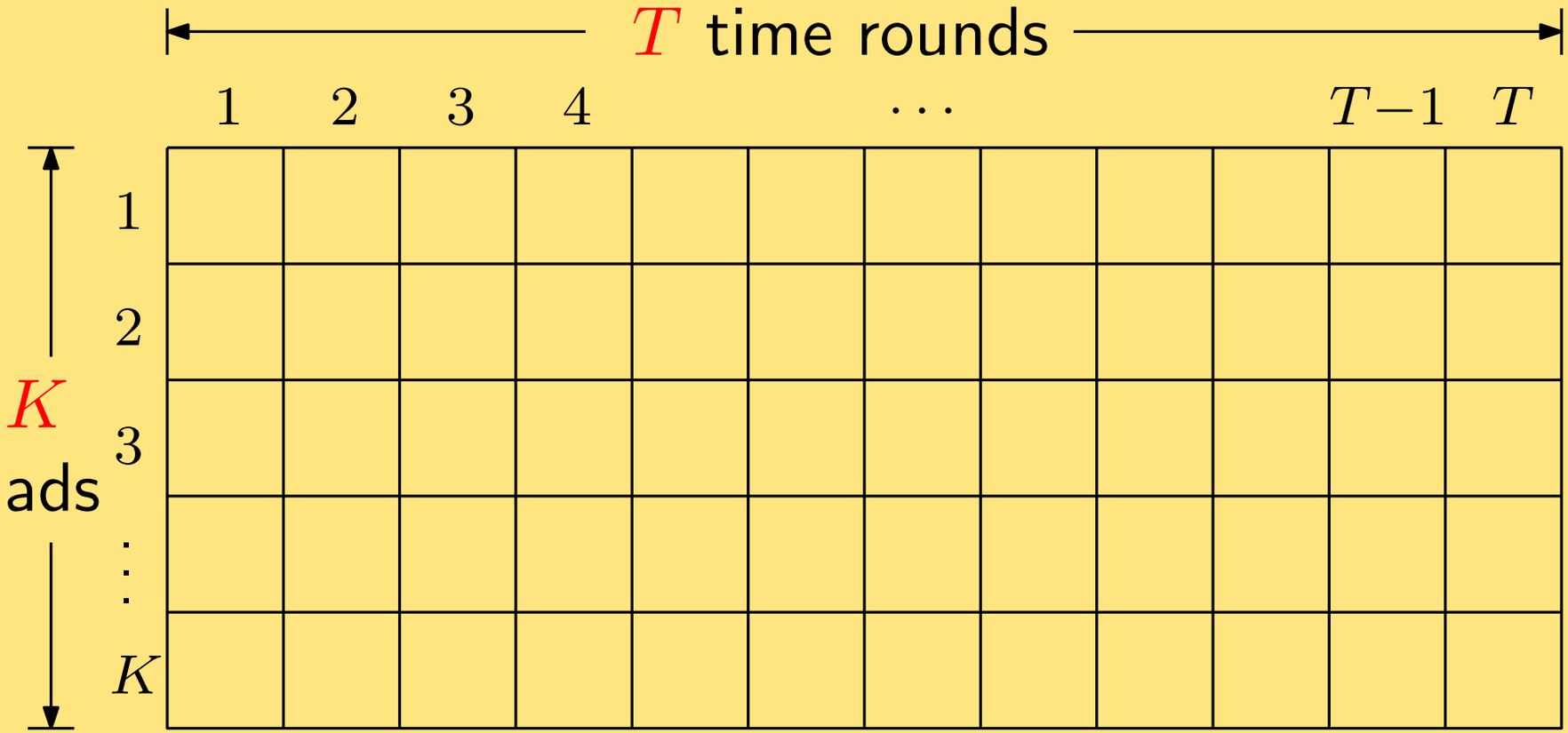
Search results

Advertisements

Clicks

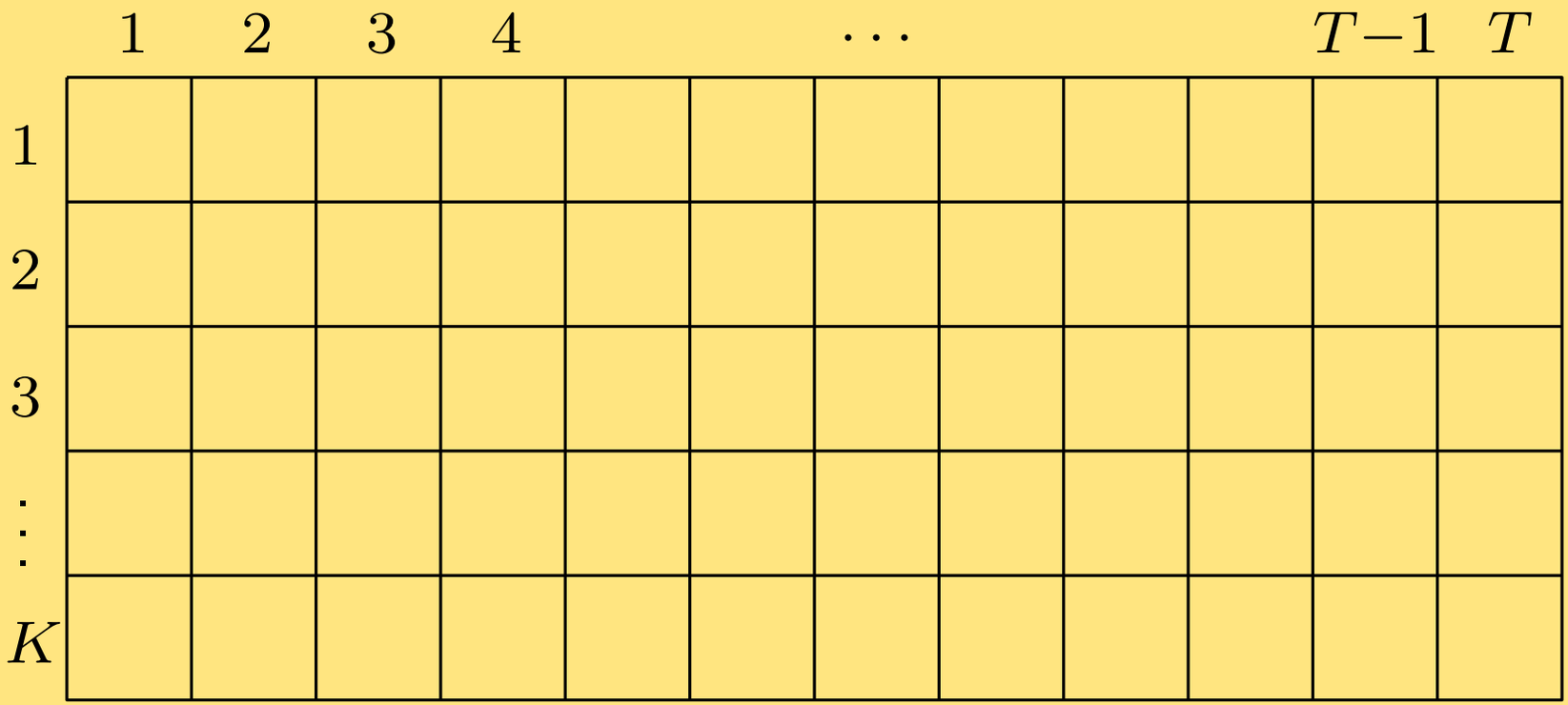
Payments

Model: Dynamics and terminology



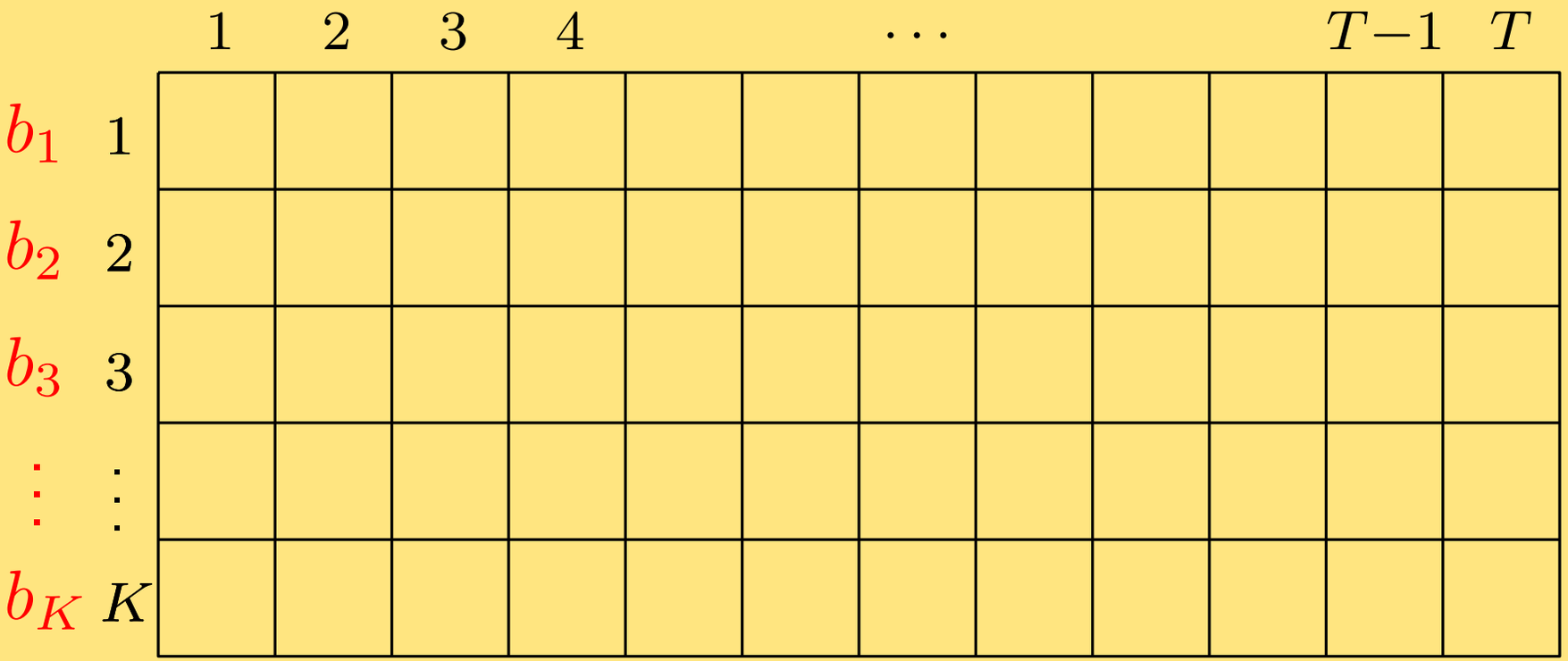
Model: Dynamics and terminology

Single slot



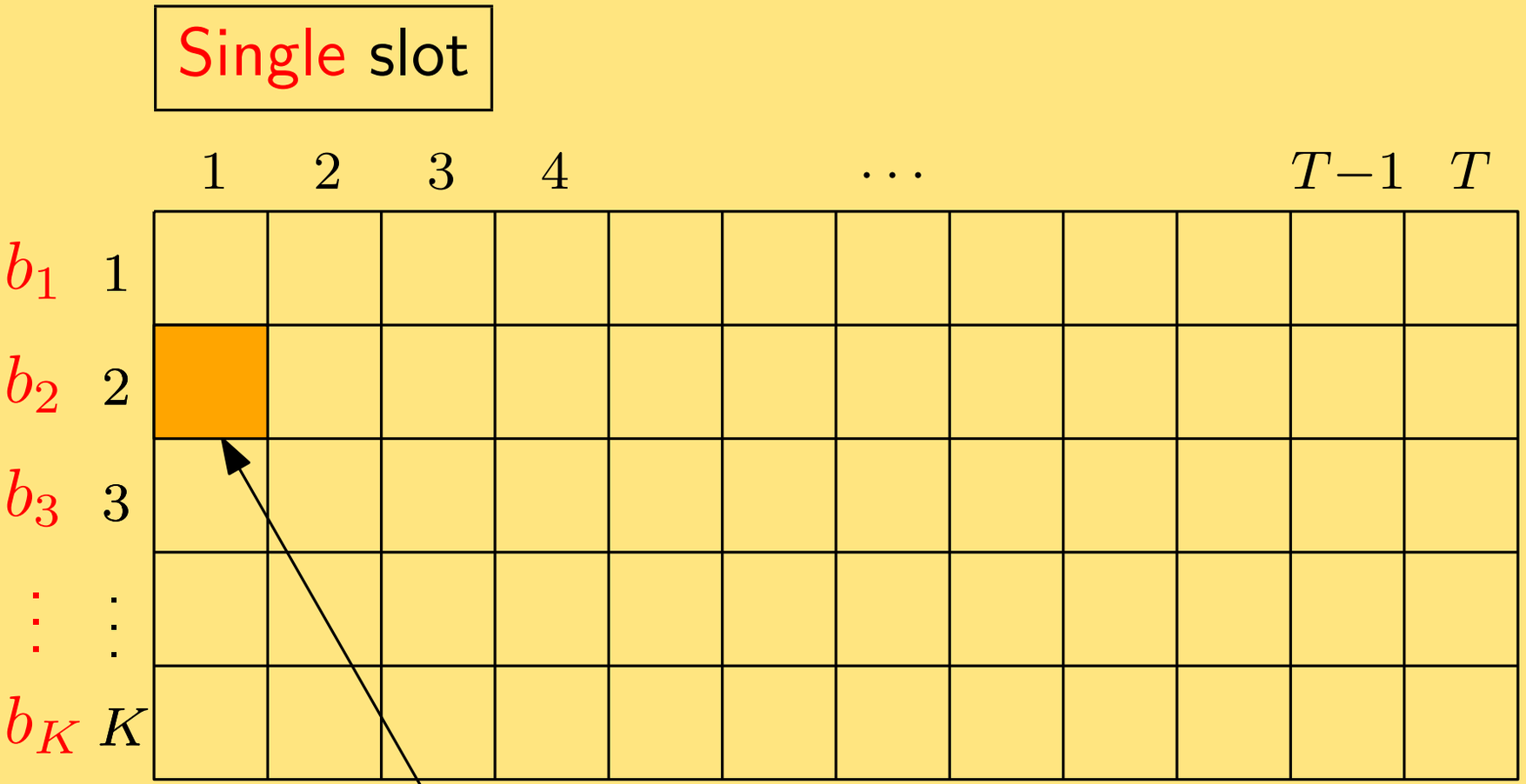
Model: Dynamics and terminology

Single slot



bid: proxy for value per click

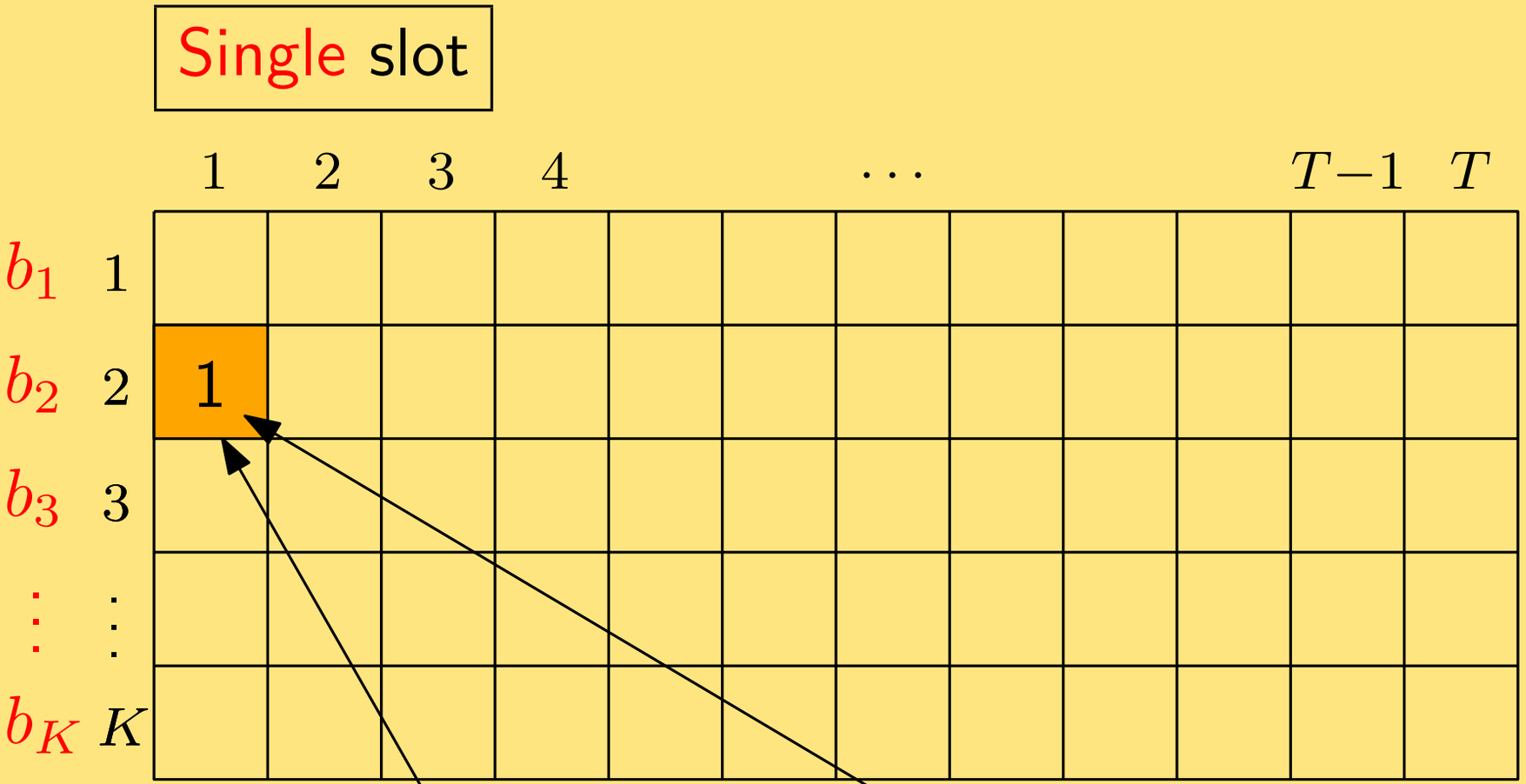
Model: Dynamics and terminology



bid: proxy for value per click

- Ad allocated, gets impression

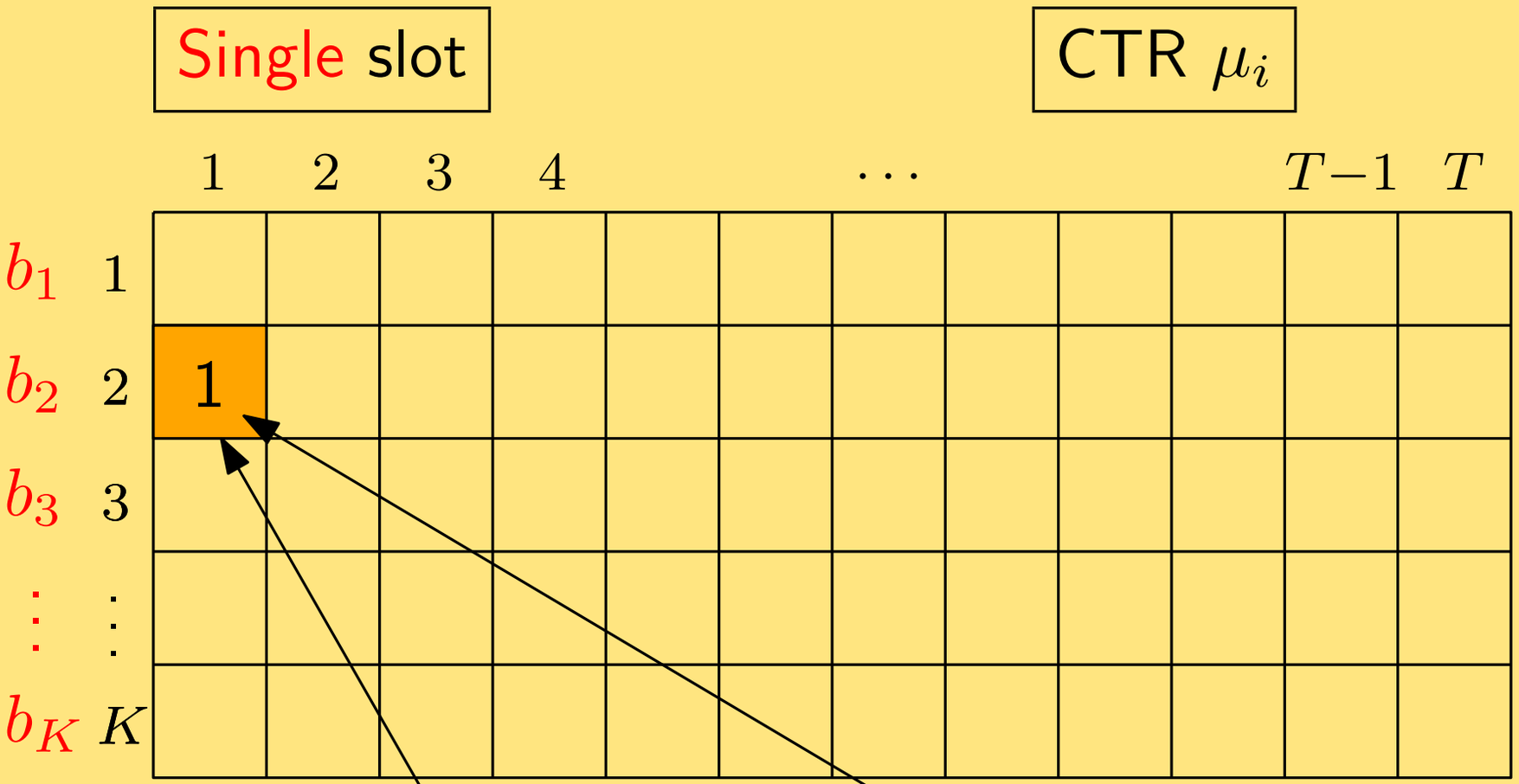
Model: Dynamics and terminology



bid: proxy for value per click

- Ad allocated, gets impression
- Ad clicked (1) or not (0)
  - Derives value on click

Model: Dynamics and terminology

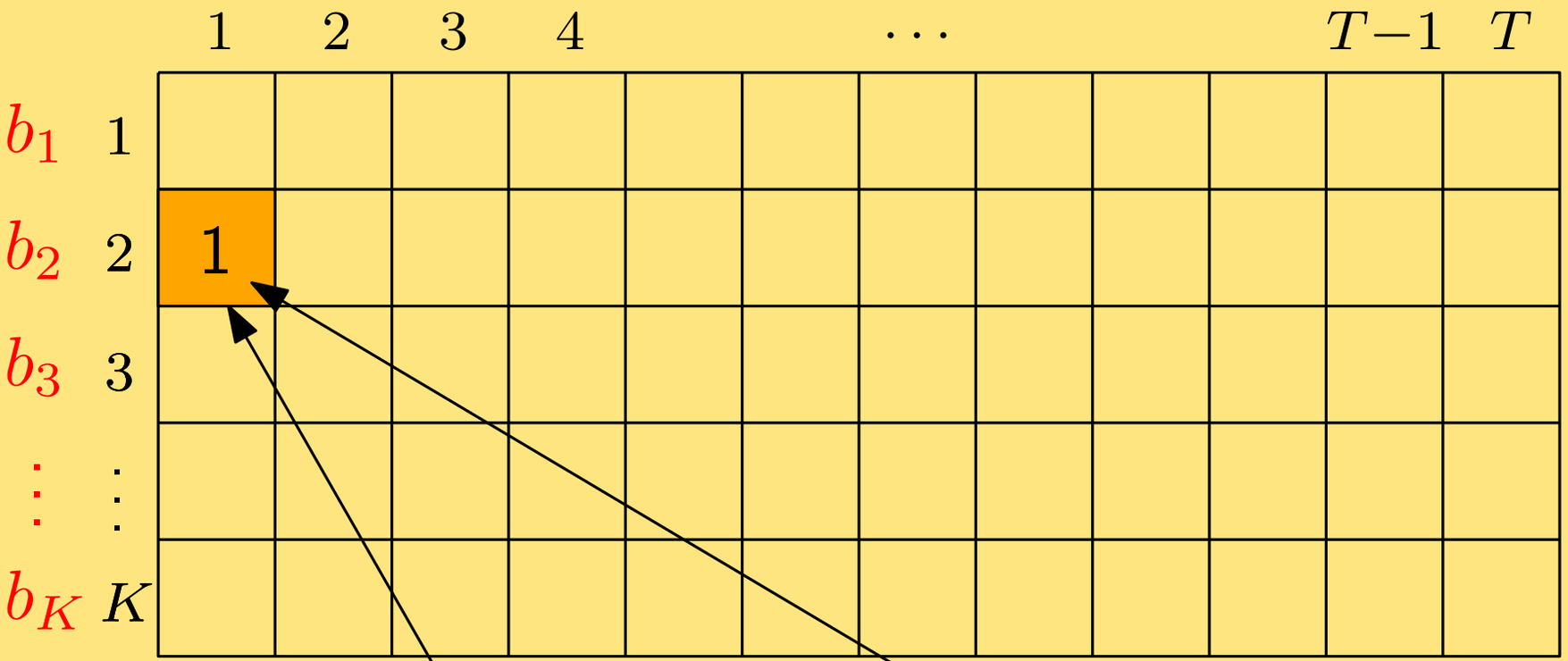


bid: proxy for value per click

- Ad allocated, gets impression
- Ad clicked (1) or not (0)
  - Derives value on click

Model: Dynamics and terminology

Single slot	Value per click v_i	CTR μ_i

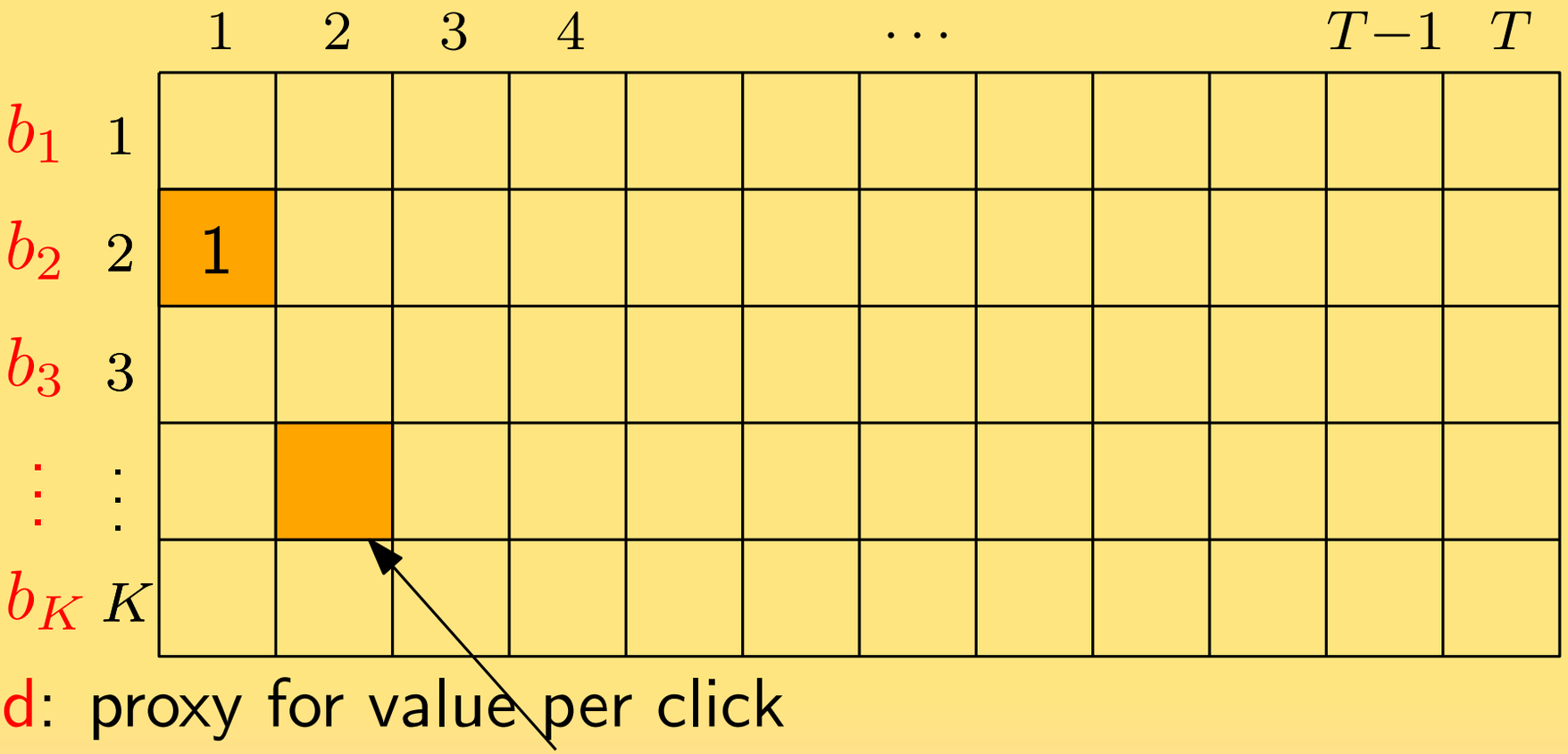


bid: proxy for value per click

- Ad allocated, gets impression
- Ad clicked (1) or not (0)
  - Derives value on click

Model: Dynamics and terminology

Single slot	Value per click v_i	CTR μ_i

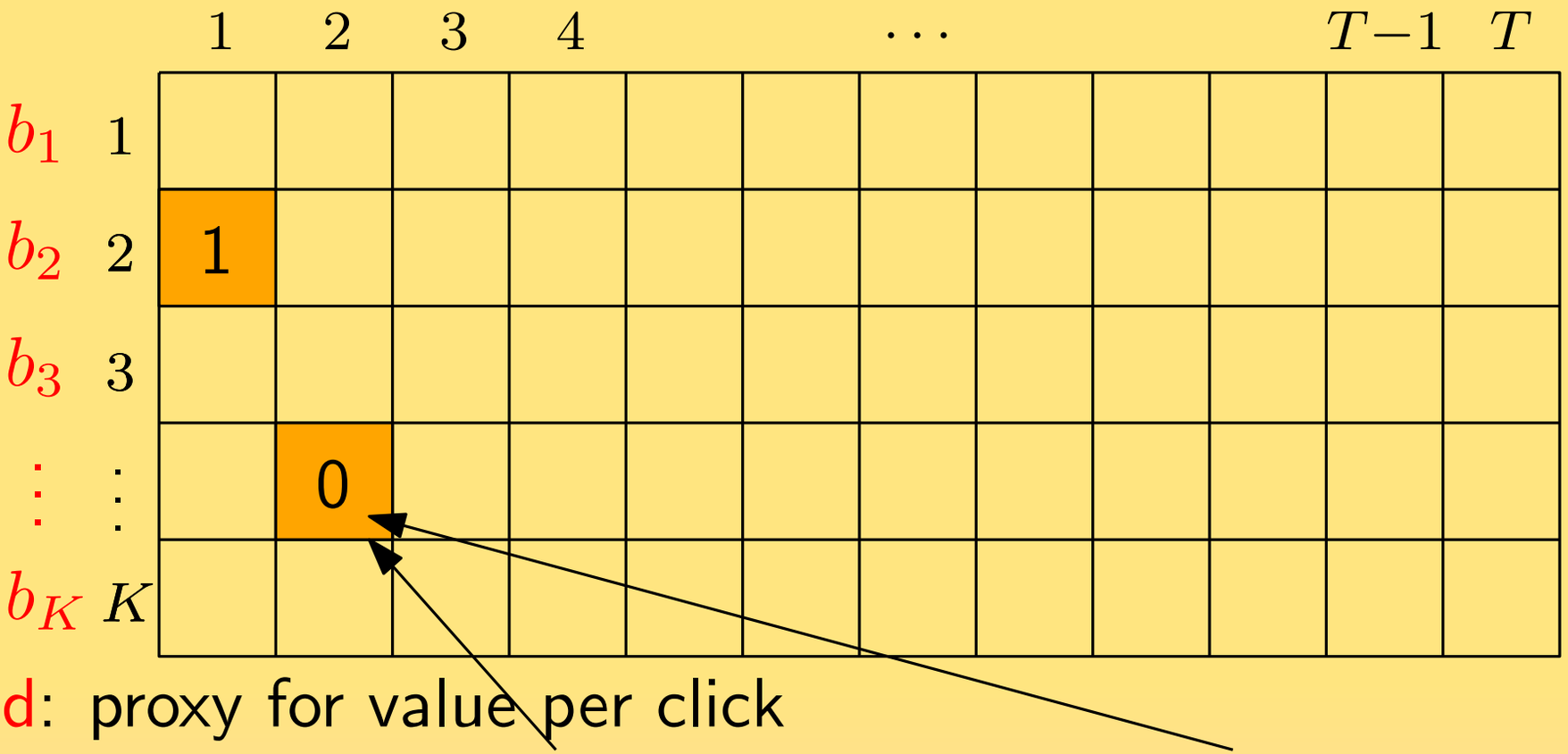


bid: proxy for value per click

- Ad allocated, gets impression
- Ad clicked (1) or not (0)
  - Derives value on click

Model: Dynamics and terminology

Single slot	Value per click v_i	CTR μ_i

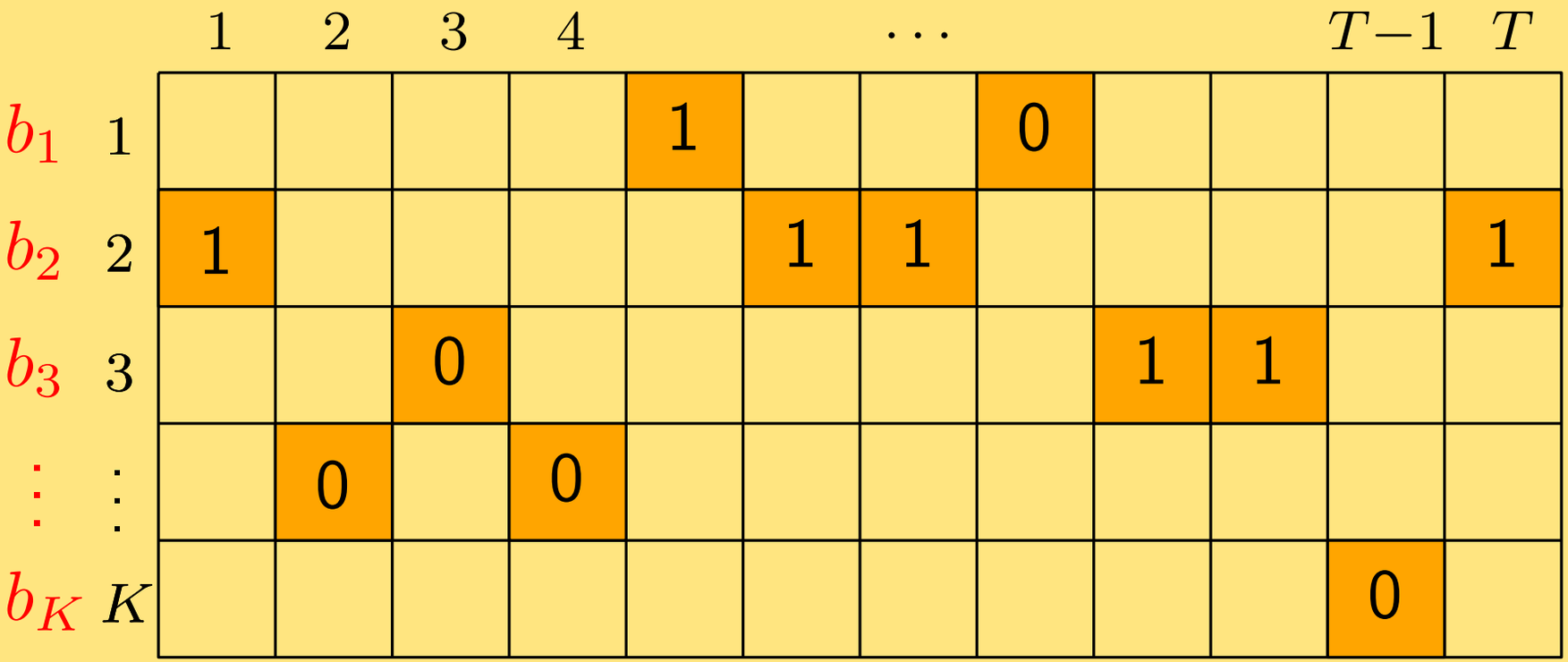


bid: proxy for value per click

- Ad allocated, gets impression
- Ad clicked (1) or not (0)
  - Derives value on click

Model: Dynamics and terminology

Single slot	Value per click v_i	CTR μ_i

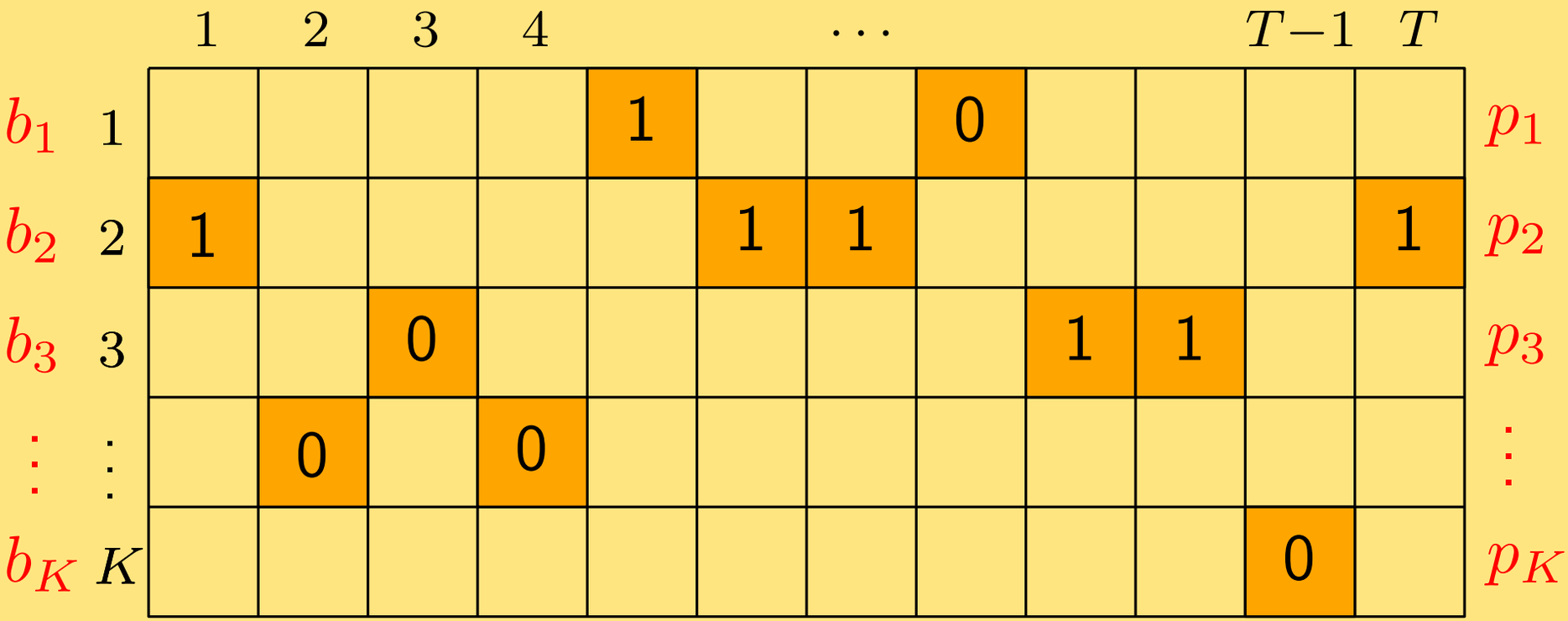


bid: proxy for value per click

- Ad allocated, gets impression
- Ad clicked (1) or not (0)
  - Derives value on click

Model: Dynamics and terminology

Single slot	Value per click v_i	CTR μ_i



bid: proxy for value per click

Prices for clicks

- Ad allocated, gets impression
- Ad clicked (1) or not (0)
  - Derives value on click

Goal for allocation: Maximize social welfare

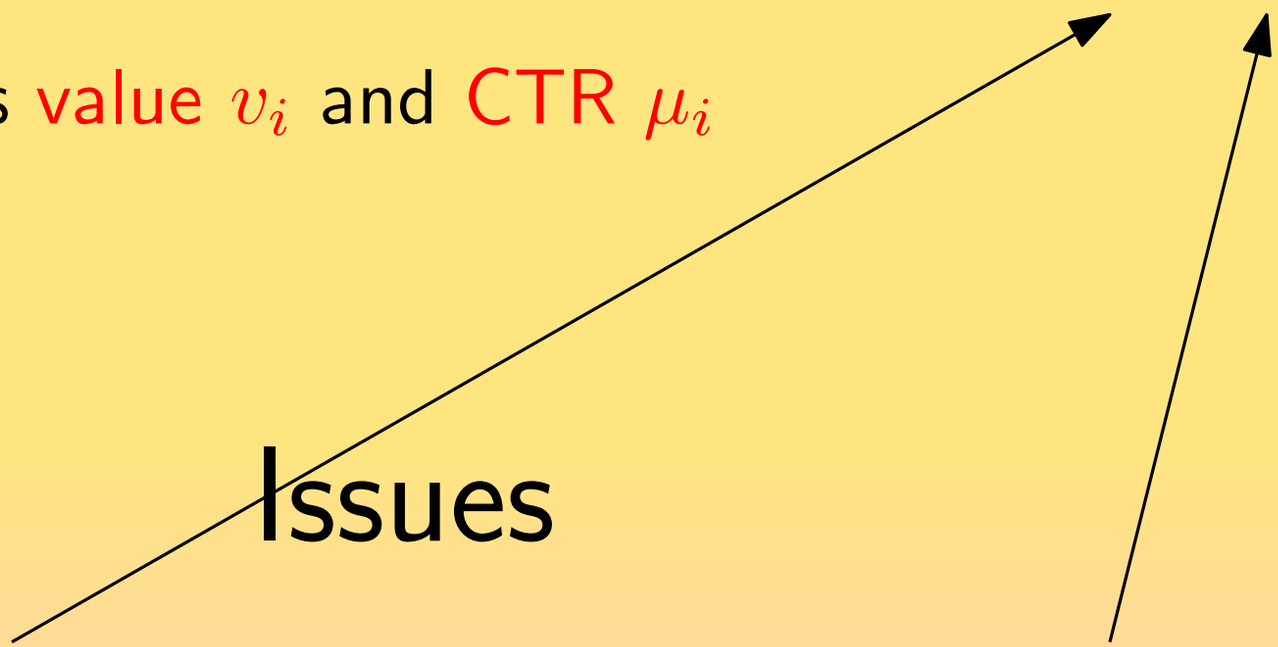
- Maximize social welfare (auctioneer + advertisers)
  - Payments cancel out

- **Easy**: Show ad with highest expected value (max $v_i \cdot \mu_i$)
  - Recall: ad has value v_i and CTR μ_i

Goal for allocation: Maximize social welfare

- Maximize social welfare (auctioneer + advertisers)
  - Payments cancel out

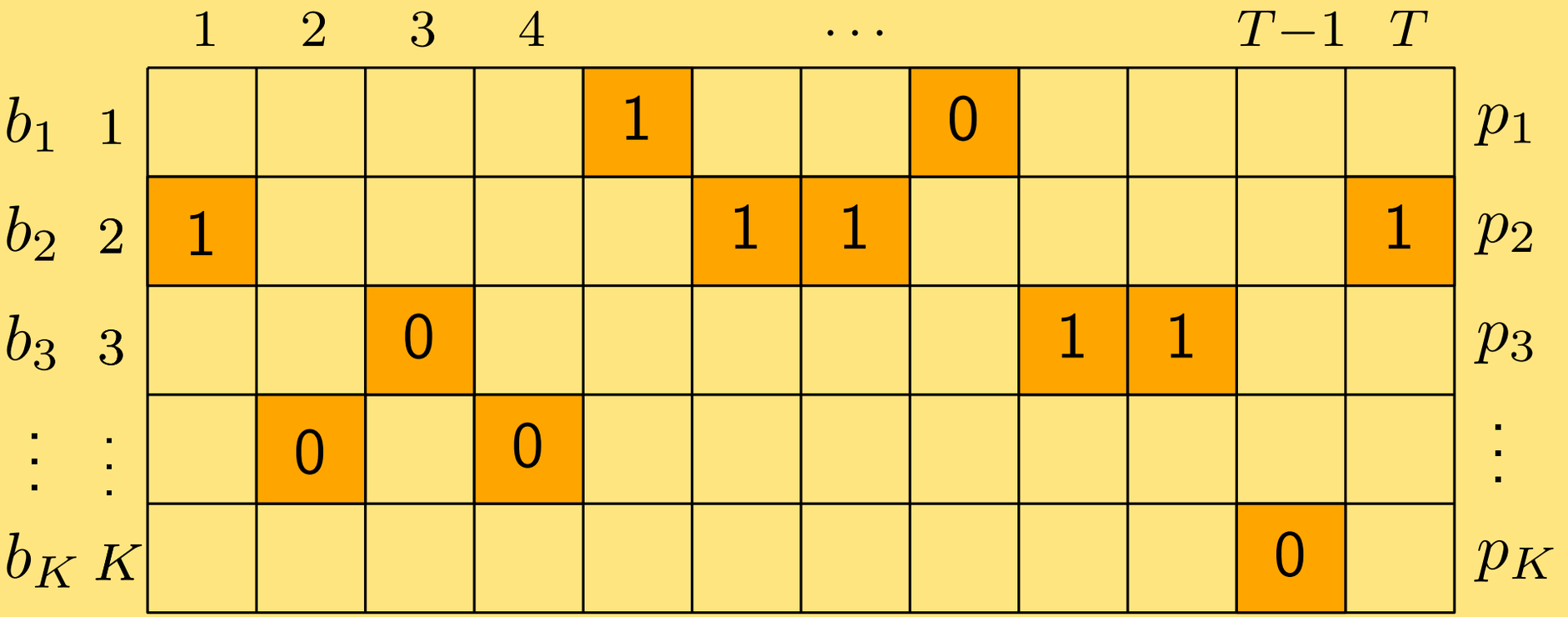
- **Easy**: Show ad with highest expected value (max $v_i \cdot \mu_i$)
  - Recall: ad has value v_i and CTR μ_i



- **Value are private to advertisers**
- **CTRs are unknown to anyone**

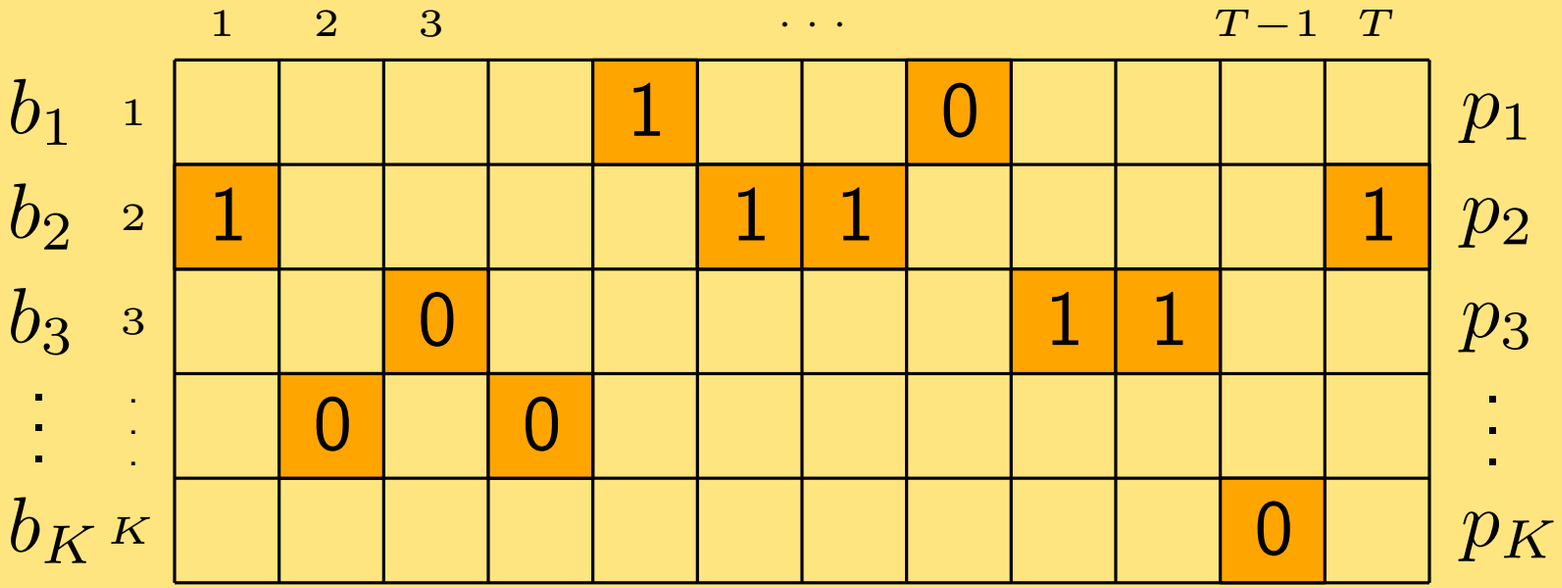
Model: Private values and unknown CTRs

Single slot	Private value per click v_i	Unknown CTR μ_i



Special case: Private values, known CTRs

(Single slot)	(Private value per click v_i)	(~~Unknown~~ CTR μ_i)



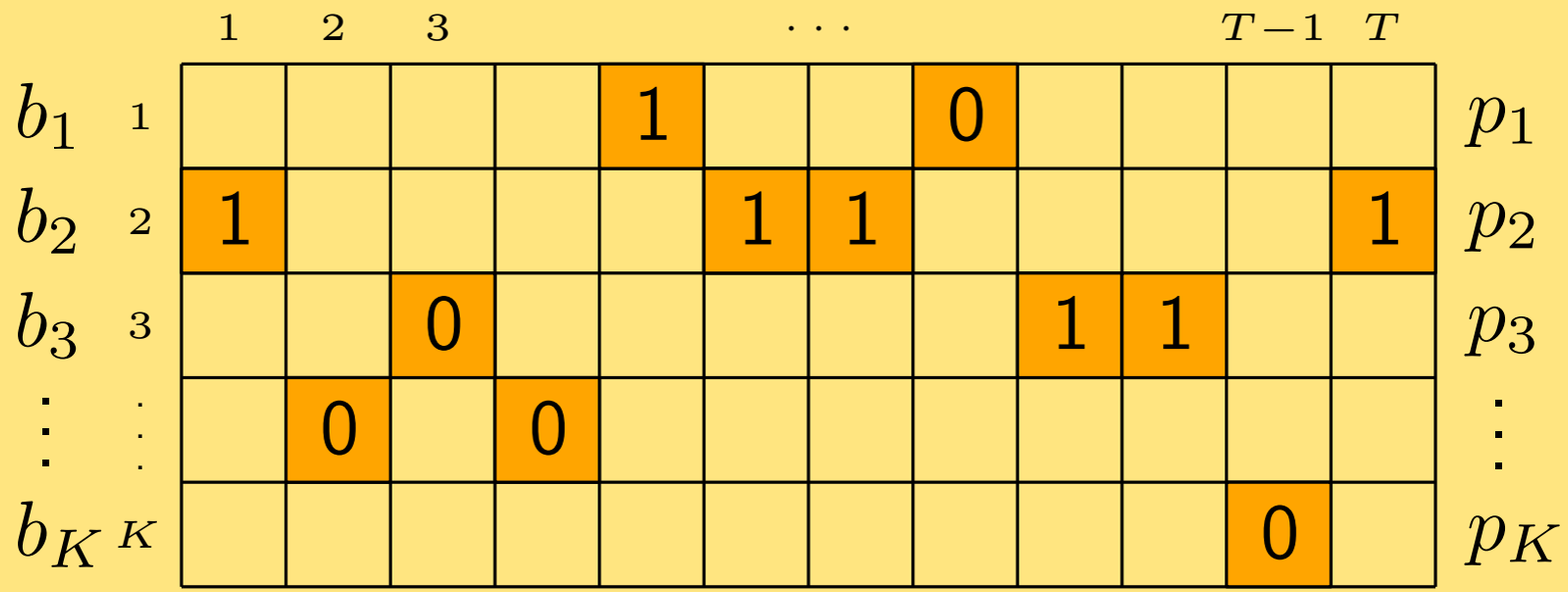
"Like" a second price auction:

- Elicit bids b_i, and order agents: $b_1\mu_1 \geq b_2\mu_2 \geq \cdots \geq b_K\mu_K$
- Allocation: agent 1 (this maximizes expected value)
- Price: $\frac{b_2\mu_2}{\mu_1}$ for agent 1

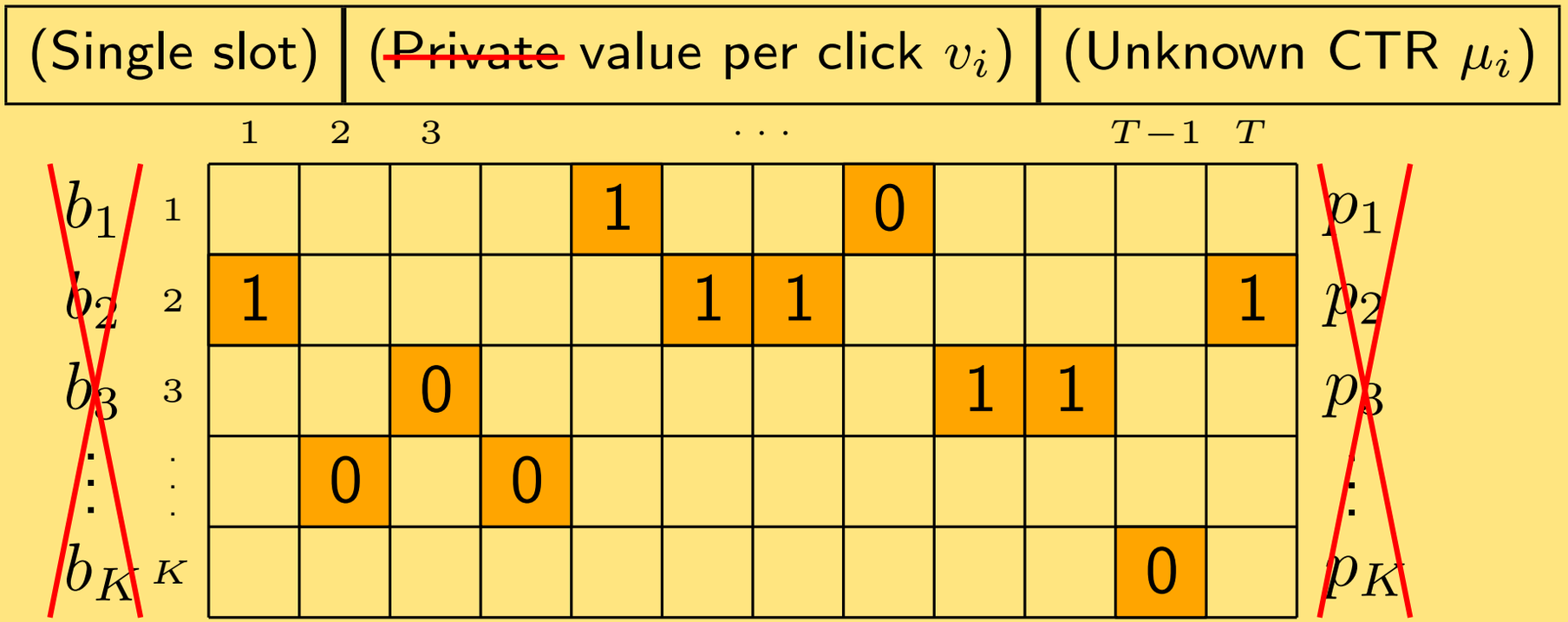
Truthtelling dominant strategy, maximizes social welfare

Special case: Unit values, unknown CTRs

(Single slot)	(~~Private~~ value per click v_i)	(Unknown CTR μ_i)



Special case: Unit values, unknown CTRs

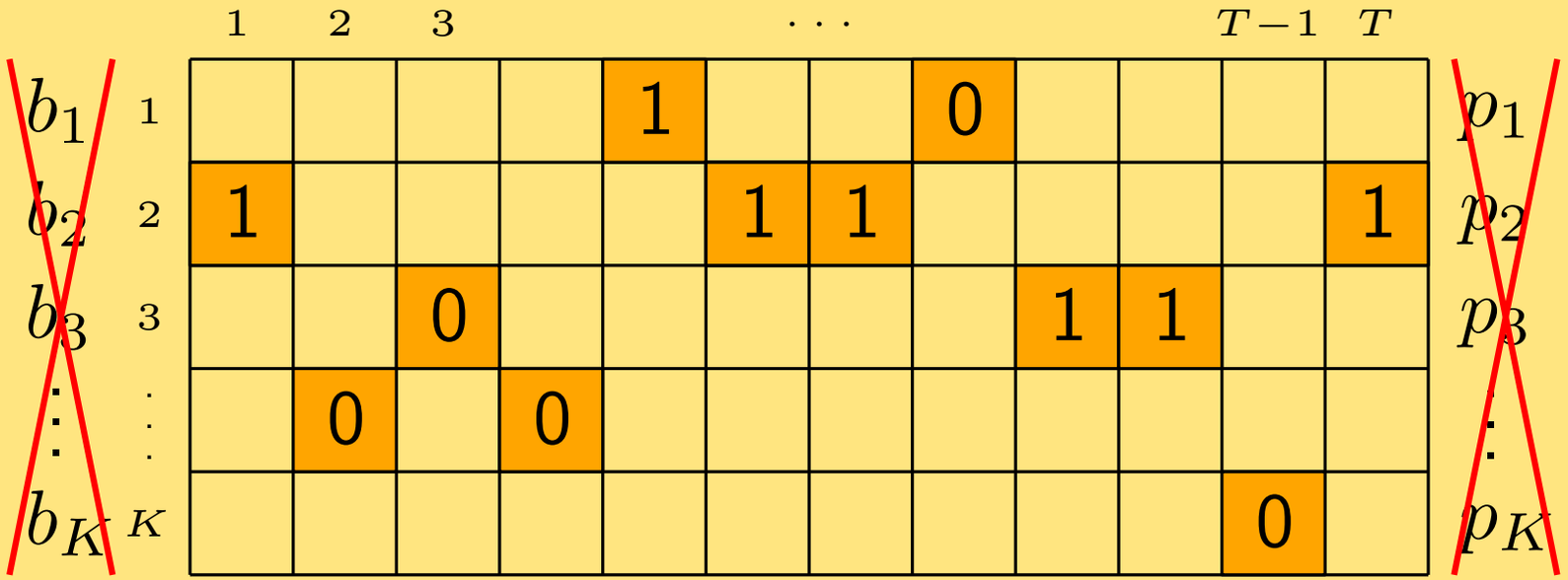


(Stocastic) Multi-Armed Bandit Problem in Learning Theory

- Tradeoff: Exploration versus Exploitation

Special case: Unit values, unknown CTRs

(Single slot)	(~~Private~~ value per click v_i)	(Unknown CTR μ_i)



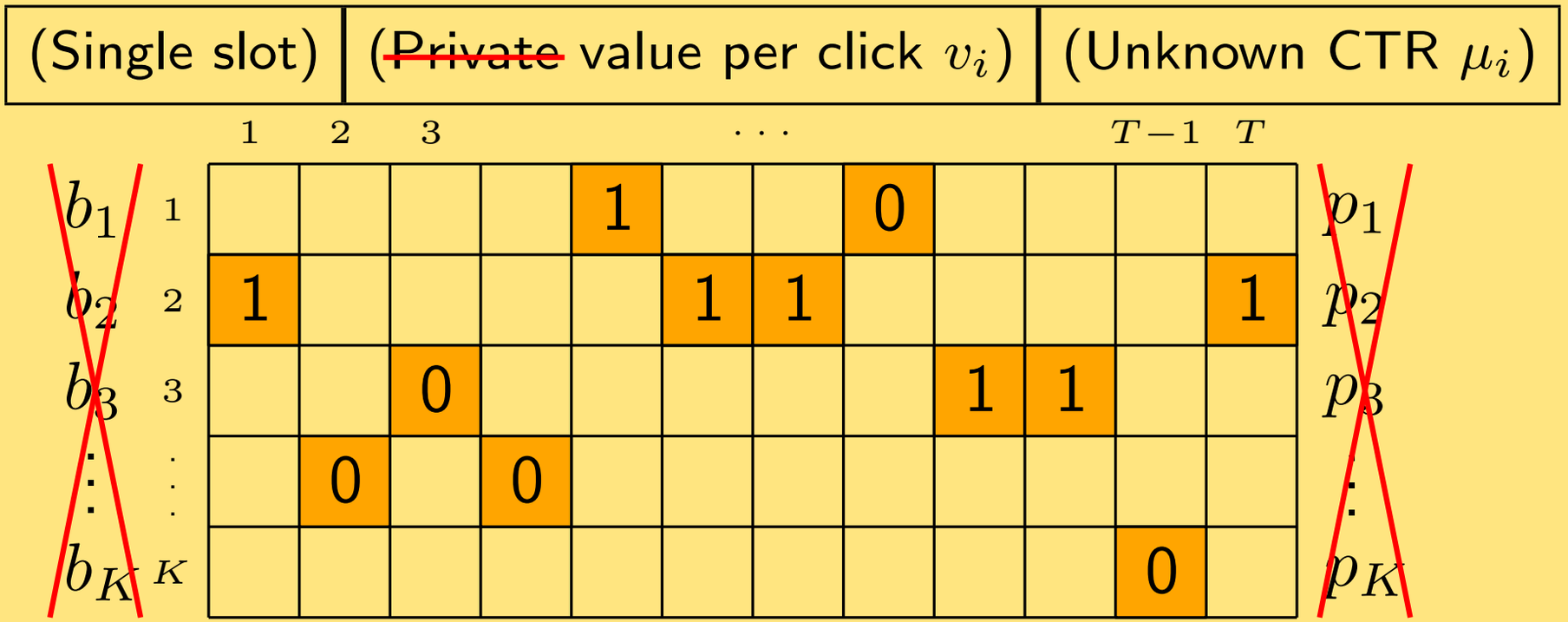
(Stocastic) Multi-Armed Bandit Problem in Learning Theory

- Tradeoff: Exploration versus Exploitation

Regret $= T \cdot \max_i \mu_i -$ Algorithms's Value

Performance measure

Special case: Unit values, unknown CTRs



(Stocastic) Multi-Armed Bandit Problem in Learning Theory

- Tradeoff: Exploration versus Exploitation

$$\text{Regret} = T \cdot \max_i \mu_i - \text{Algorithms's Value}$$

- Achievable regret $= \tilde{O}(\sqrt{KT})$

General case: Private values, unknown CTRs

Mechanism design meets Learning

General case: Private values, unknown CTRs

Mechanism design meets Learning

"Strategic issue"

Values are private information

(Truthfulness)

"Learning issue"

CTRs are unknown

(Efficiency)

General case: Private values, unknown CTRs

Mechanism design meets Learning

"Strategic issue"
Values are private information
(Truthfulness)

"Learning issue"
CTRs are unknown
(Efficiency)

Truthful Multi-Armed Bandit Problem

General case: Private values, unknown CTRs

Mechanism design meets Learning

"Strategic issue"
Values are private information
(Truthfulness)

"Learning issue"
CTRs are unknown
(Efficiency)

Truthful Multi-Armed Bandit Problem

Main questions and (informal) results

General case: Private values, unknown CTRs

Mechanism design meets Learning

"Strategic issue"
Values are private information
(Truthfulness)

"Learning issue"
CTRs are unknown
(Efficiency)

Truthful Multi-Armed Bandit Problem

Main questions and (informal) results

- What is the structure of truthful mechanisms?

General case: Private values, unknown CTRs

Mechanism design meets Learning

"Strategic issue"
Values are private information
(Truthfulness)

"Learning issue"
CTRs are unknown
(Efficiency)

Truthful Multi-Armed Bandit Problem

Main questions and (informal) results

- What is the structure of truthful mechanisms?

- Truthful mechanisms regret versus MAB algorithms regret?

General case: Private values, unknown CTRs

Mechanism design meets Learning

"Strategic issue"
Values are private information
(Truthfulness)

"Learning issue"
CTRs are unknown
(Efficiency)

Truthful Multi-Armed Bandit Problem

Main questions and (informal) results

- What is the structure of truthful mechanisms?
  - Must separate exploration and exploitation.
- Truthful mechanisms regret versus MAB algorithms regret?

General case: Private values, unknown CTRs

Mechanism design meets Learning

"Strategic issue"
Values are private information
(Truthfulness)

"Learning issue"
CTRs are unknown
(Efficiency)

Truthful Multi-Armed Bandit Problem

Main questions and (informal) results

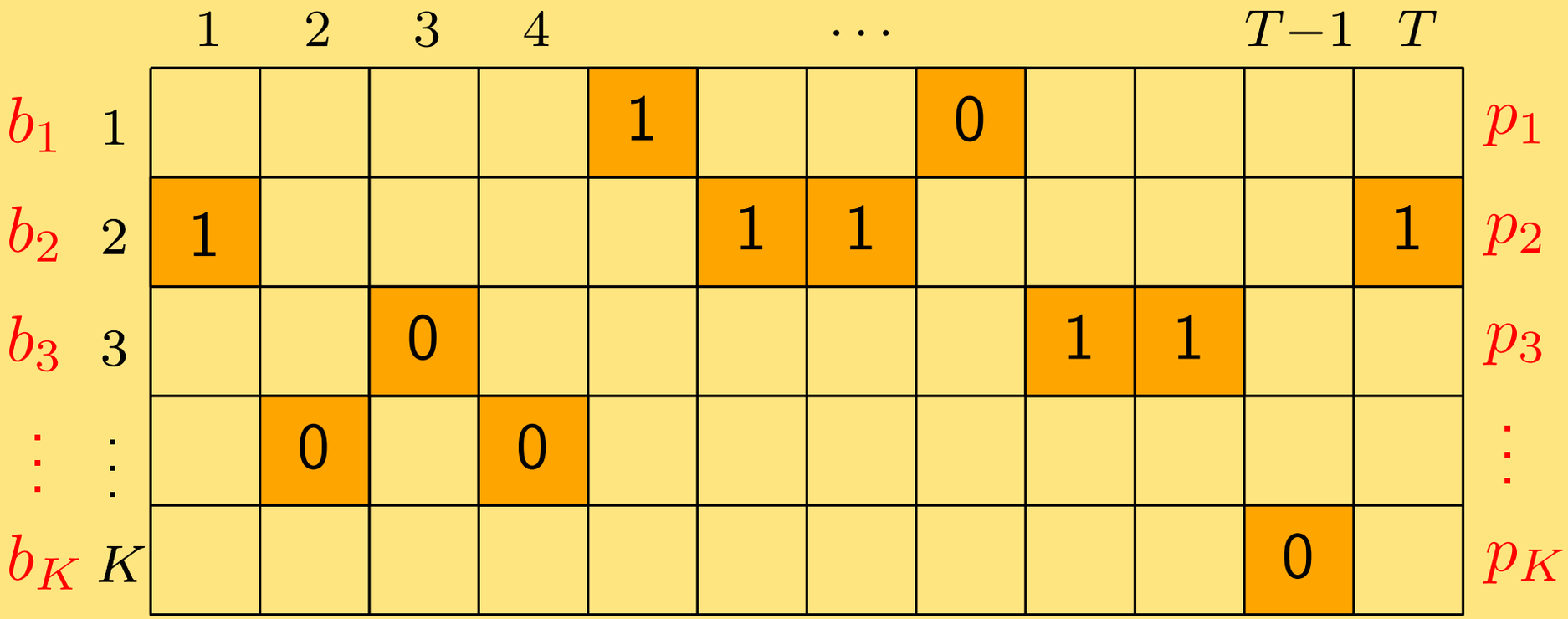
- What is the structure of truthful mechanisms?
  - Must separate exploration and exploitation.
- Truthful mechanisms regret versus MAB algorithms regret?
  - Truthful mechanisms have much higher regret.

Overview

- ✓ Overview and informal results
- MAB mechanism design and need for a characterization
- Characterization and lower bounds
- Conclusions and open problems

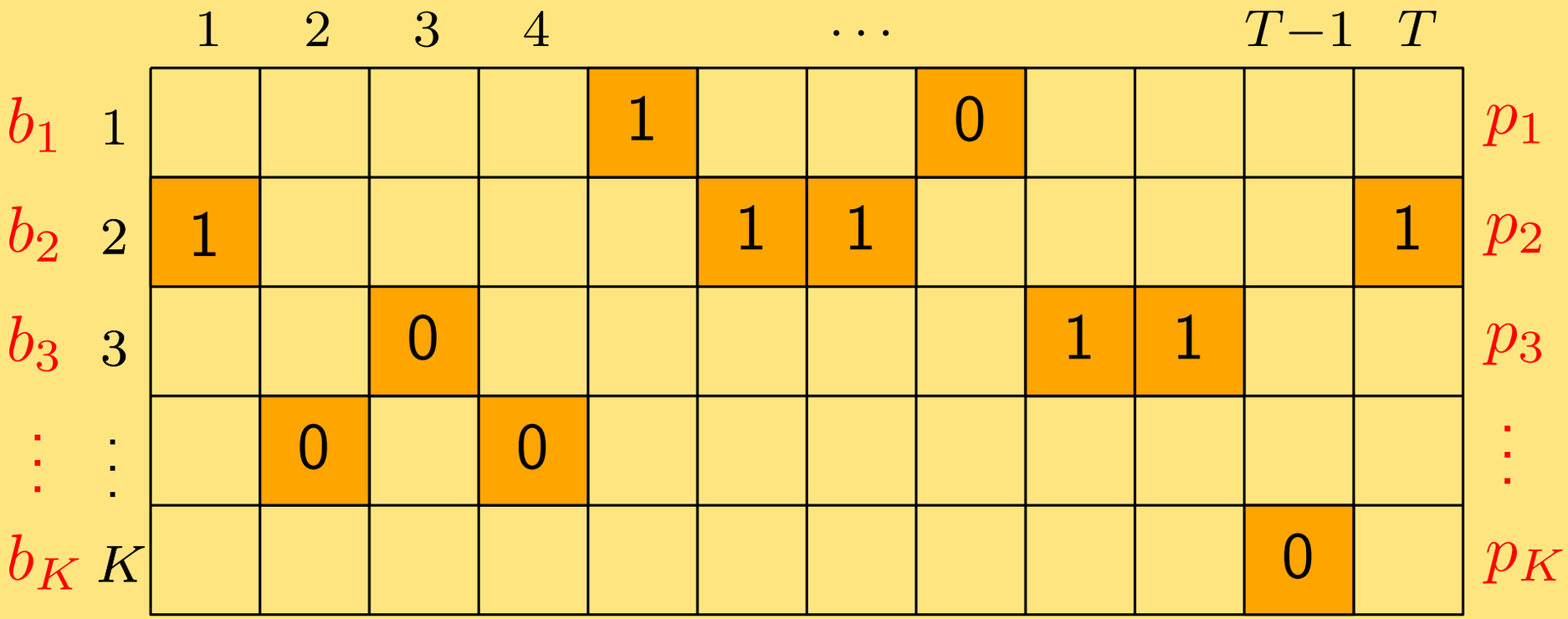
MAB mechanism design

Single slot	Unknown CTR μ_i	Private value per click v_i



MAB mechanism design

Single slot	Unknown CTR μ_i	Private value per click v_i



$$\text{utility}_i = v_i \cdot (\#\text{clicks}_i) - \text{price}_i$$

Requirements on mechanism

- Deterministic mech (for now)

Requirements on mechanism

- Deterministic mech (for now)
- Realization ρ: click info for all agents, all rounds

	1	2	3						T
1	1	0	0	1	1	0	0	1	0
2	0	0	0	1	1	0	0	1	1
.	0	1	1	0	1	1	1	0	0
.	1	1	0	0	1	1	1	0	1
K	0	0	1	0	0	1	0	0	1

Requirements on mechanism

- Deterministic mech (for now)
- Realization ρ: click info for all agents, all rounds

	1	2	3						T
1	1	0	0	1	1	0	0	1	0
2	0	0	0	1	1	0	0	1	1
⋮	0	1	1	0	1	1	1	0	0
	1	1	0	0	1	1	1	0	1
K	0	0	1	0	0	1	0	0	1

Truthful

For any realization and for any bid profile, truthtelling maximizes utility.

Normalized

For any realization and any bid profile, average price per click is in $[0, b_i]$.

- Mechanism can only use bids and previously observed clicks.

Myerson's characterization

Myerson's characterization

Theorem [Myerson '81, Archer Tardos '01]: Let (A, p) be a normalized mechanism. Then,

(A, p) is truthful for each realization

$\Updownarrow$

#$\mathrm{clicks}_i(b_i)$ monotone non-decreasing in b_i for each realization.

Myerson's characterization

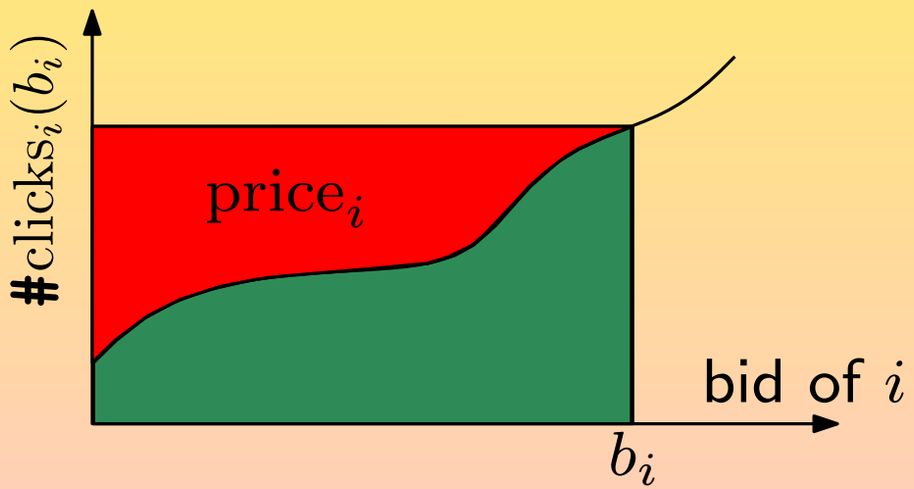
Theorem [Myerson '81, Archer Tardos '01]: Let (A, p) be a normalized mechanism. Then,

(A, p) is truthful for each realization

$\Updownarrow$

#$\text{clicks}_i(b_i)$ monotone non-decreasing in b_i for each realization.

Payments are uniquely determined



Myerson's characterization

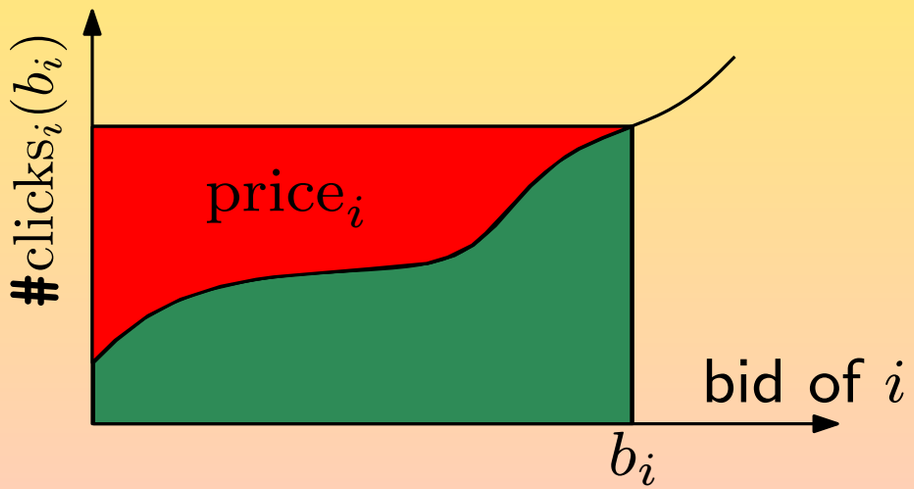
Theorem [Myerson '81, Archer Tardos '01]: Let (A, p) be a normalized mechanism. Then,

(A, p) is truthful for each realization

$\Updownarrow$

#$\mathrm{clicks}_i(b_i)$ monotone non-decreasing in b_i for each realization.

Payments are uniquely determined



- Mechanism design reduces to monotone algorithm design!

Myerson's characterization

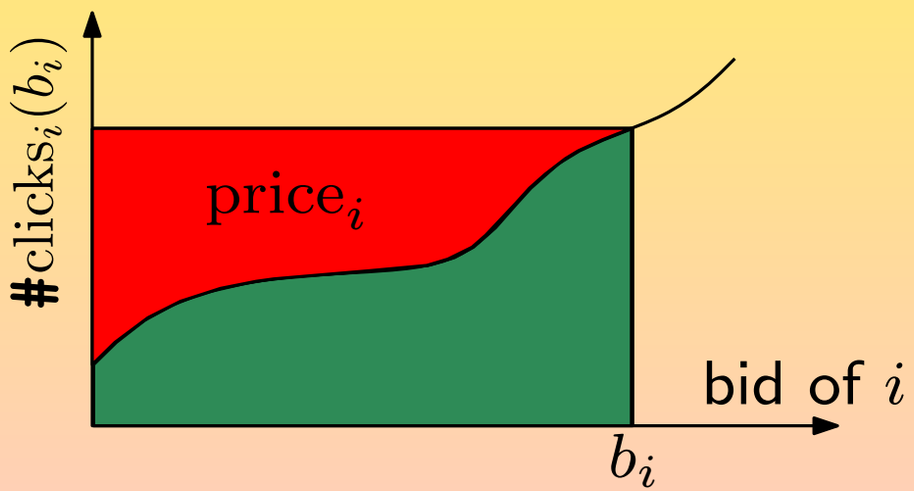
Theorem [Myerson '81, Archer Tardos '01]: Let (A, p) be a normalized mechanism. Then,

(A, p) is truthful for each realization

$\Updownarrow$

$\#\text{clicks}_i(b_i)$ monotone non-decreasing in b_i for each realization.

Payments are uniquely determined



- Mechanism design reduces to monotone algorithm design!
- **Not in our setting!**

Does Myerson's characterization suffice?

$\text{bid}_1 < \text{bid}_2$

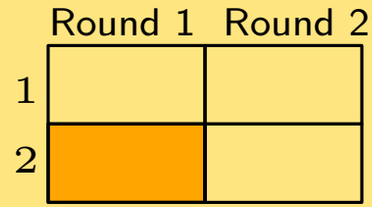
	Round 1	Round 2
1		
2		

$\text{bid}_1 \geq \text{bid}_2$

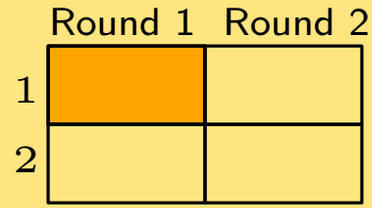
	Round 1	Round 2
1		
2		

Does Myerson's characterization suffice?

$\text{bid}_1 < \text{bid}_2$

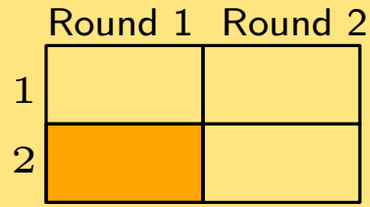


$\text{bid}_1 \geq \text{bid}_2$

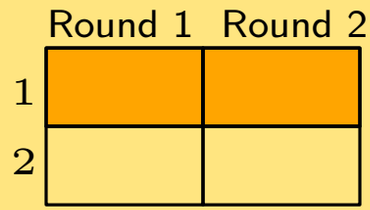


Does Myerson's characterization suffice?

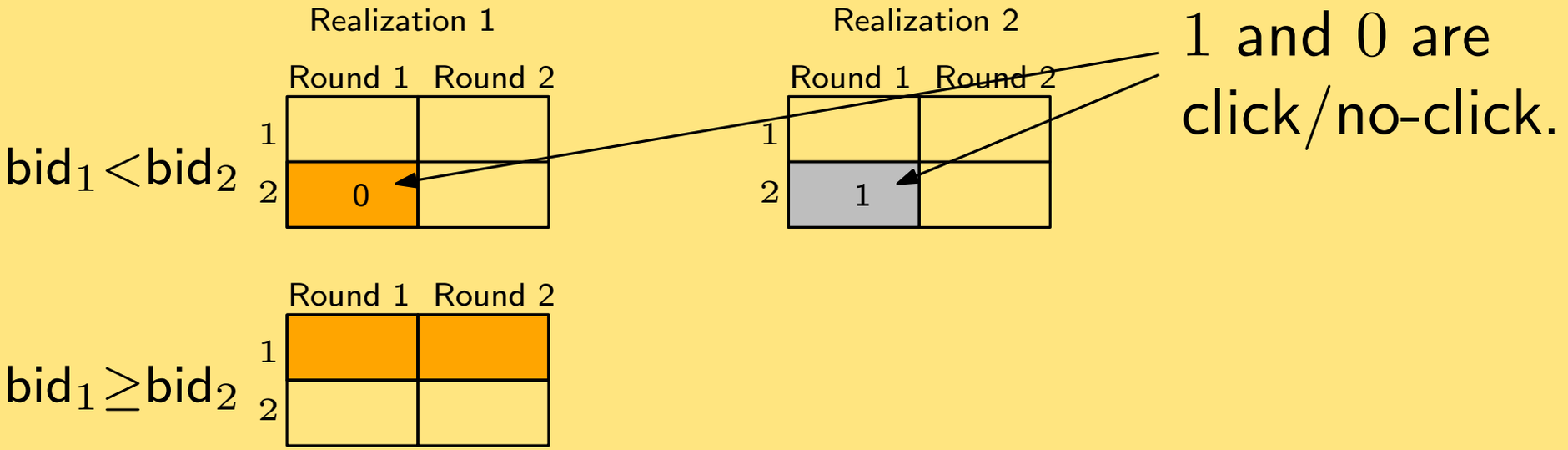
$\text{bid}_1 < \text{bid}_2$



$\text{bid}_1 \geq \text{bid}_2$



Does Myerson's characterization suffice?



Mechanism "prefers" agent 1

Does Myerson's characterization suffice?

Realization 1

$\text{bid}_1 < \text{bid}_2$

	Round 1	Round 2
1		
2	0	

Realization 2

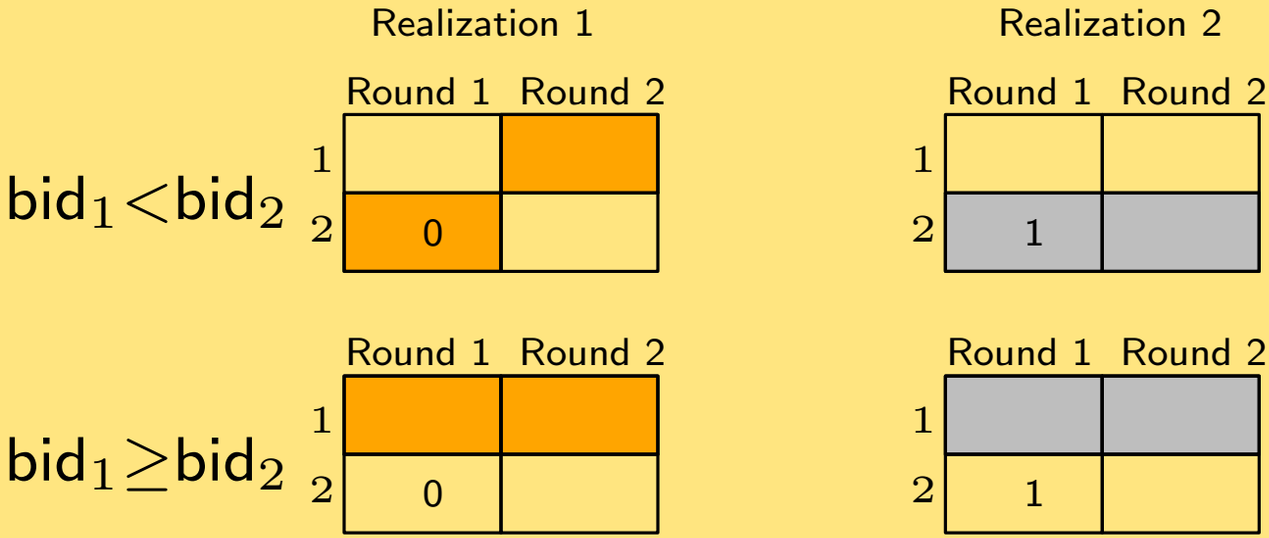
	Round 1	Round 2
1		
2	1	

$\text{bid}_1 \geq \text{bid}_2$

	Round 1	Round 2
1		
2		

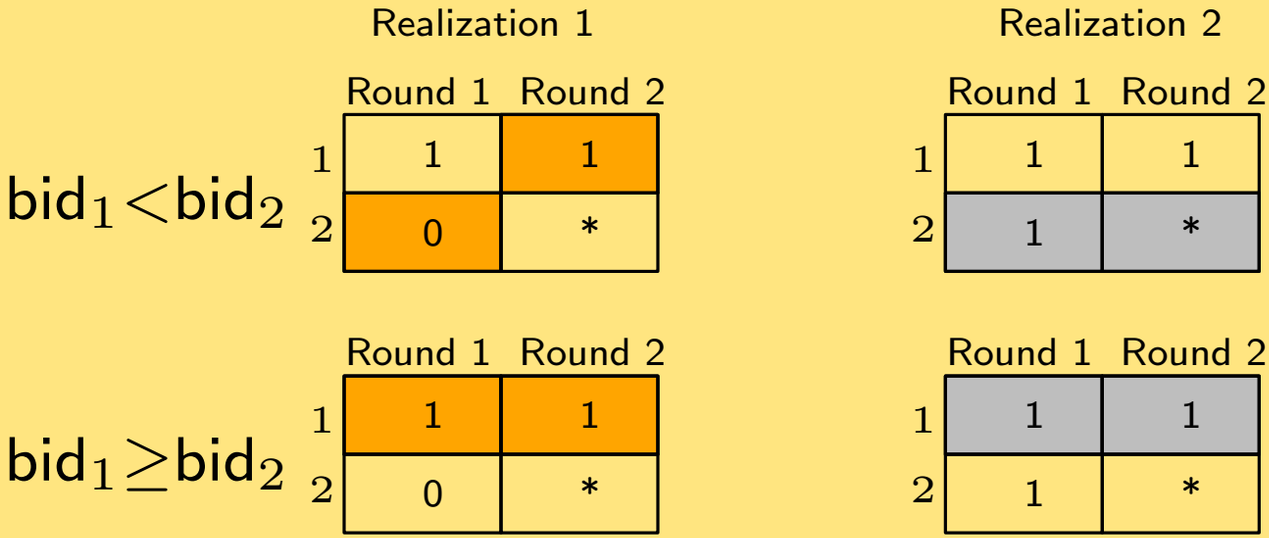
Mechanism "prefers" agent 1

Does Myerson's characterization suffice?



Mechanism "prefers" agent 1

Does Myerson's characterization suffice?



Mechanism "prefers" agent 1

Does Myerson's characterization suffice?

Realization 1

$\text{bid}_1 < \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	*

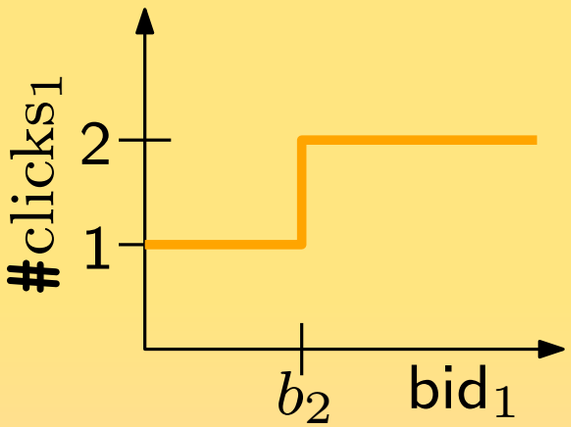
Realization 2

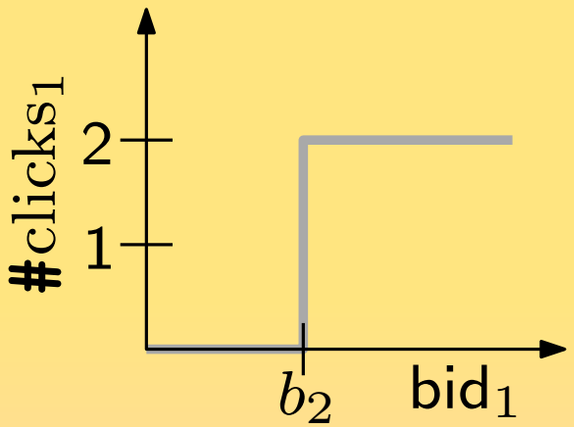
	Round 1	Round 2
1	1	1
2	1	*

$\text{bid}_1 \geq \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	*

	Round 1	Round 2
1	1	1
2	1	*





Does Myerson's characterization suffice?

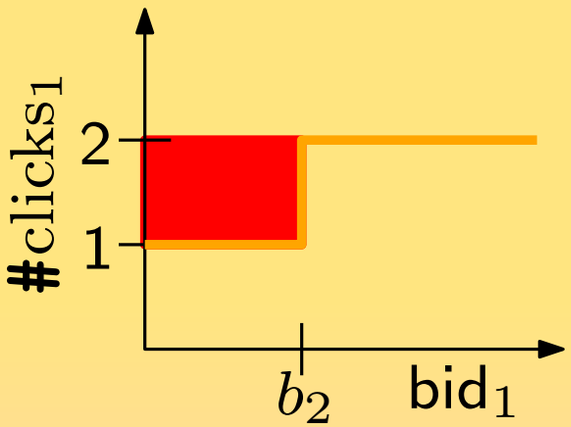
Realization 1

$\text{bid}_1 < \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	*

$\text{bid}_1 \geq \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	*

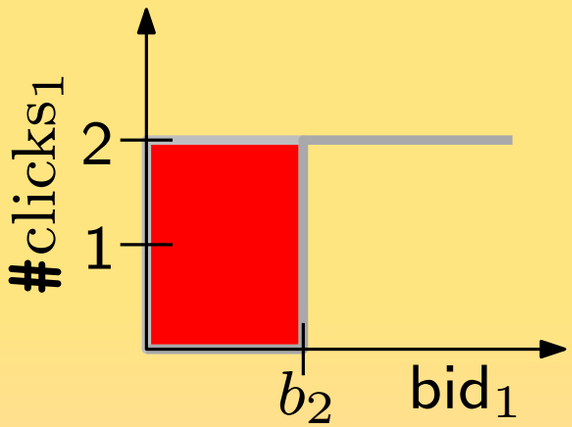


$p_1(\infty, b_2) = b_2$

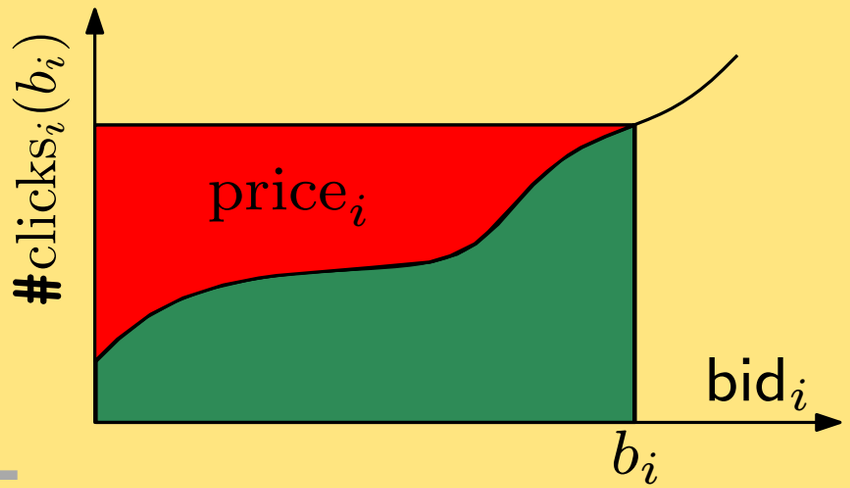
Realization 2

	Round 1	Round 2
1	1	1
2	1	*

	Round 1	Round 2
1	1	1
2	1	*



$p_1(\infty, b_2) = 2b_2$



Does Myerson's characterization suffice?

Realization 1

$\text{bid}_1 < \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	*

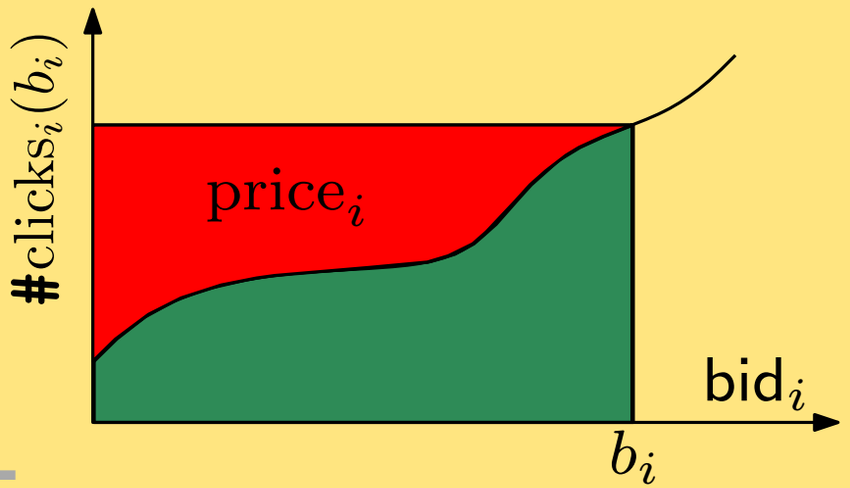
Realization 2

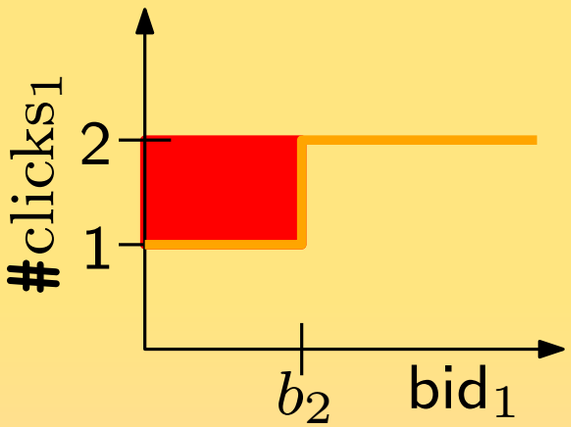
	Round 1	Round 2
1	1	1
2	1	*

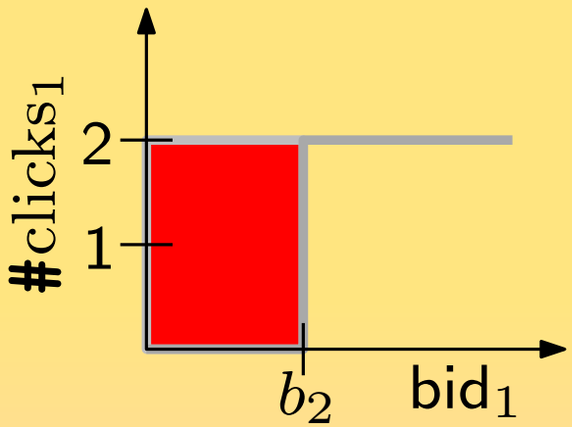
$\text{bid}_1 \geq \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	*

	Round 1	Round 2
1	1	1
2	1	*







$p_1(\infty, b_2) = b_2$

$p_1(\infty, b_2) = 2b_2$

Price cannot be calculated for agent 1.

What is the problem here?

Realization 1

$\text{bid}_1 < \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	

Realization 2

	Round 1	Round 2
1	1	1
2	1	*

$\text{bid}_1 \geq \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	

	Round 1	Round 2
1	1	1
2	1	*

What is the problem here?

Realization 1

$\text{bid}_1 < \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	

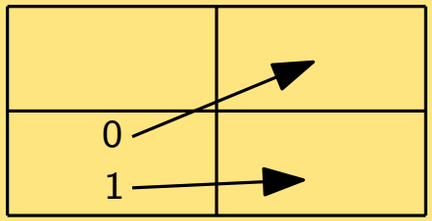
$\text{bid}_1 \geq \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	

Realization 2

	Round 1	Round 2
1	1	1
2	1	*

	Round 1	Round 2
1	1	1
2	1	*



Round 1 is influential for agent 1

What is the problem here?

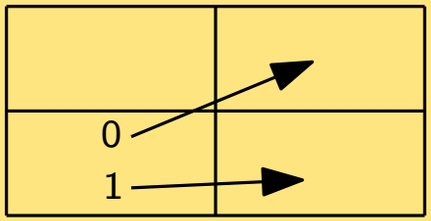
Realization 1

$\text{bid}_1 < \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	

Realization 2

	Round 1	Round 2
1	1	1
2	1	*

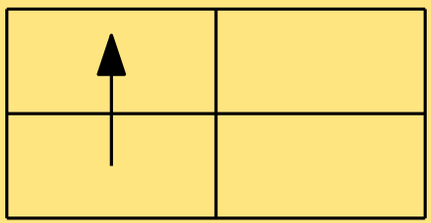


Round 1 is influential for agent 1

$\text{bid}_1 \geq \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	

	Round 1	Round 2
1	1	1
2	1	*



Agent 1 can steal impression in round 1

What is the problem here?

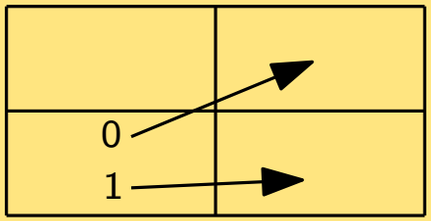
Realization 1

$\text{bid}_1 < \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	

Realization 2

	Round 1	Round 2
1	1	1
2	1	*

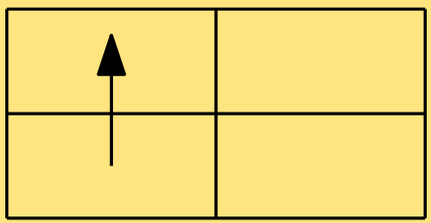


Round 1 is influential for agent 1

$\text{bid}_1 \geq \text{bid}_2$

	Round 1	Round 2
1	1	1
2	0	

	Round 1	Round 2
1	1	1
2	1	*



Agent 1 can steal impression in round 1

That is precisely what is wrong!

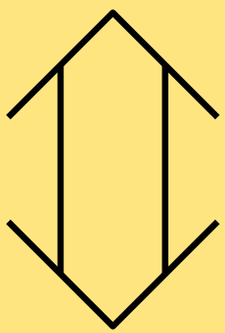
Overview

- ✓ Overview and informal results
- ✓ MAB mechanism design and need for a characterization
- Characterization and lower bounds
- Conclusions and open problems

Characterization for two agents

Let A be a non-degenerate deterministic allocation rule.
Let A be scalefree.

- (A, p) is normalized truthful for some p

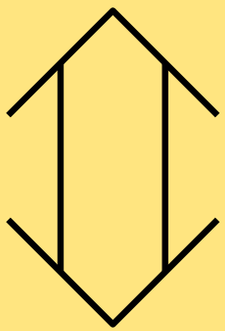


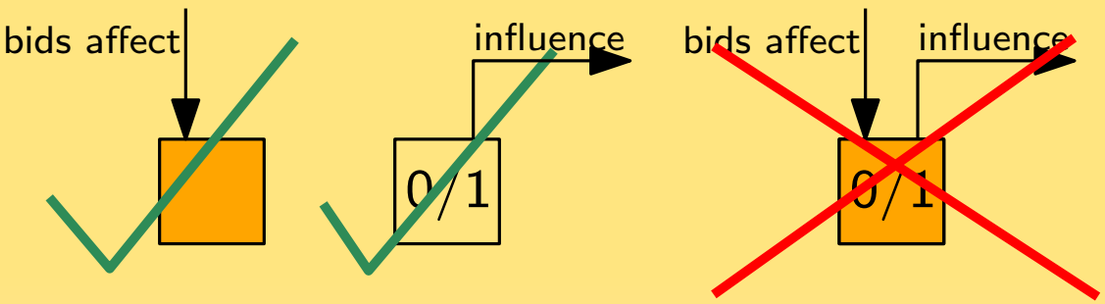
- A is pointwise-monotone
- A is exploration-separated

Characterization for two agents

Let A be a non-degenerate deterministic allocation rule.
Let A be scalefree.

- (A, p) is normalized truthful for some p





- A is pointwise-monotone
- A is exploration-separated

If round t's click info used later, then t's allocation independent of bids.

Consequences of characterization (2 agents)

$R(T) = T \cdot \max_i v_i \mu_i -$ Expected value of mechanism

- Regret: loss in social welfare due to not knowing CTRs

	Truthful mechanisms	Bandit algorithms
Worst case regret	$\Omega(T^{2/3})$	$O(T^{1/2})$
δ-gap regret	$\Omega(\delta \cdot T^{\epsilon})$ for $\epsilon > 0$	$O(\delta^{-1} \cdot \log T)$

$|v_1\mu_1 - v_2\mu_2| \geq \delta \max\{v_1, v_2\}$

Consequences of characterization (2 agents)

$R(T) = T \cdot \max_i v_i \mu_i -$ Expected value of mechanism

- Regret: loss in social welfare due to not knowing CTRs

	Truthful mechanisms	Bandit algorithms
Worst case regret	$\Omega(T^{2/3})$	$O(T^{1/2})$
δ-gap regret	$\Omega(\delta \cdot T^{\epsilon})$ for $\epsilon > 0$	$O(\delta^{-1} \cdot \log T)$

$|v_1\mu_1 - v_2\mu_2| \geq \delta \max\{v_1, v_2\}$

Consequences of characterization (2 agents)

$R(T) = T \cdot \max_i v_i \mu_i -$ Expected value of mechanism

- Regret: loss in social welfare due to not knowing CTRs

	Truthful mechanisms	Bandit algorithms
Worst case regret	$\Omega(T^{2/3})$	$O(T^{1/2})$
δ-gap regret	$\Omega(\delta \cdot T^{\epsilon})$ for $\epsilon > 0$	$O(\delta^{-1} \cdot \log T)$

$|v_1\mu_1 - v_2\mu_2| \geq \delta \max\{v_1, v_2\}$

Consequences of characterization (2 agents)

$R(T) = T \cdot \max_i v_i \mu_i -$ Expected value of mechanism

- Regret: loss in social welfare due to not knowing CTRs

	Truthful mechanisms	Bandit algorithms
Worst case regret	$\Omega(T^{2/3})$	$O(T^{1/2})$
δ-gap regret	$\Omega(\delta \cdot T^{\epsilon})$ for $\epsilon > 0$	$O(\delta^{-1} \cdot \log T)$

$|v_1\mu_1 - v_2\mu_2| \geq \delta \max\{v_1, v_2\}$

- A simple mechanism achieves these bounds.

Related work

Tension between "truthfulness" and "solution quality"

- Scheduling problems [Nisan, Ronen 2001], [Archer, Tardos 2001]
- Combinatorial auctions [Lavi, Mu'alem, Nisan 2003], [Dobzinski, Sundararajan 2008], [Mossel, Papadimitriou, Schapira, Singer 2009]
- CPPP [Papadimitriou, Schapira, Singer 2008]

Dynamic auctions: Bayesian settings

- [Athey and Segal 2007], [Bergemann, Valimaki 2007]

MAB with approximate truthfulness

- Pay per actions auctions [Nazerzadeh, Saberi, Vohra '08]

Truthful MAB auctions

- $\Theta(T^{2/3})$ regret w.r.t. second-price revenue [Devanur, Kakade, Later in this session]

Conclusions and open problems

Truthful deterministic multi-armed bandit mechanisms

- Simple settings that combine strategic and learning issues
- Structure: must separate exploration and exploitation
- Lower bound on regret: algs. better than mechanisms
- A simple mechanism with matching regret bounds

Conclusions and open problems

Truthful deterministic multi-armed bandit mechanisms

- Simple settings that combine strategic and learning issues
- Structure: must separate exploration and exploitation
- Lower bound on regret: algs. better than mechanisms
- A simple mechanism with matching regret bounds

Randomized mechanisms: Partial results and open problems

			Characterization, Lower bounds?
Universally truthful	$\forall_{\text{clicks}}$	$\forall_{\text{seeds}}$ [utility]	In paper
Weakly truthful	$\forall_{\text{clicks}}$	$\mathbb{E}_{\text{seeds}}$ [utility]	Open
Truthful in expectation	$\mathbb{E}_{\text{clicks}}$	$\mathbb{E}_{\text{seeds}}$ [utility]	Open