Introduction and Recent Advances: Multi-Armed Bandit Mechanisms

Y. Narahari[1] **and Dinesh Garg**[2]

[1]Dept. of CSA, IISc, Bangalore
[2]IBM Research, New Delhi

Joint work with
Satyanath Bhat, Arpita Biswas, Shweta Jain, and Debmalya Mandal

January 18, 2014

Outline

[Part 1] (Forward) MAB Mechanisms

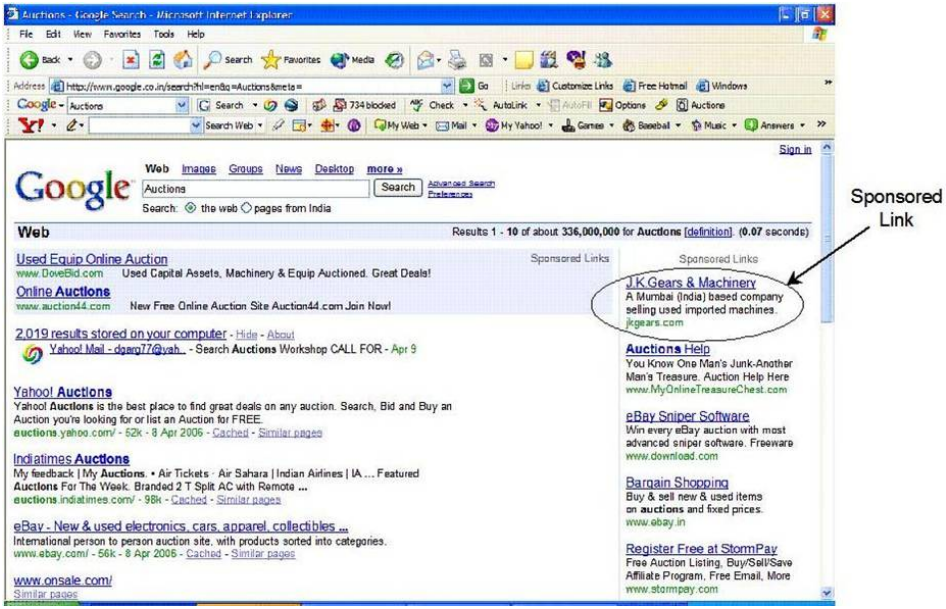
- Motivating example
- Taxonomy of (Forward) MAB Mechanisms
- Characterization for truthfulness
- Welfare maximizing (Forward) MAB mechanisms
- Revenue maximizing (Forward) MAB mechanisms

[Part 2] (Reverse) MAB mechanisms

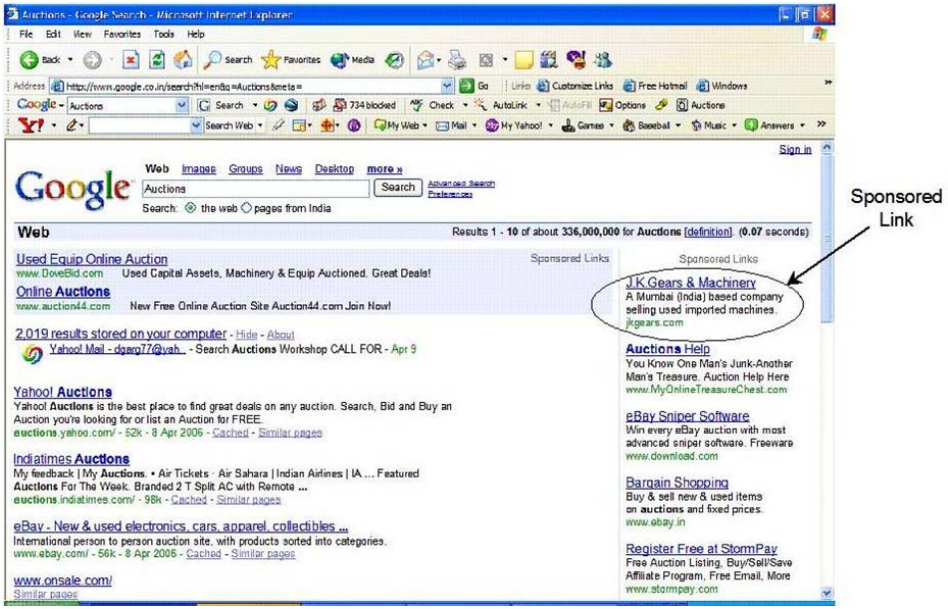
- Motivating example
- Taxonomy of (Reverse) MAB Mechanisms
- Recent results on (Reverse) MAB mechanisms

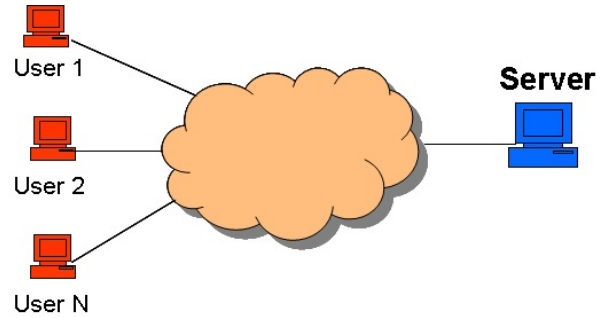
Example 1: Online Advertising

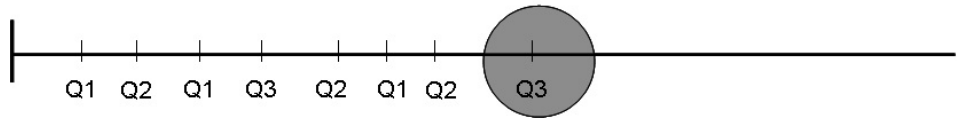
Example 1: Online Advertising



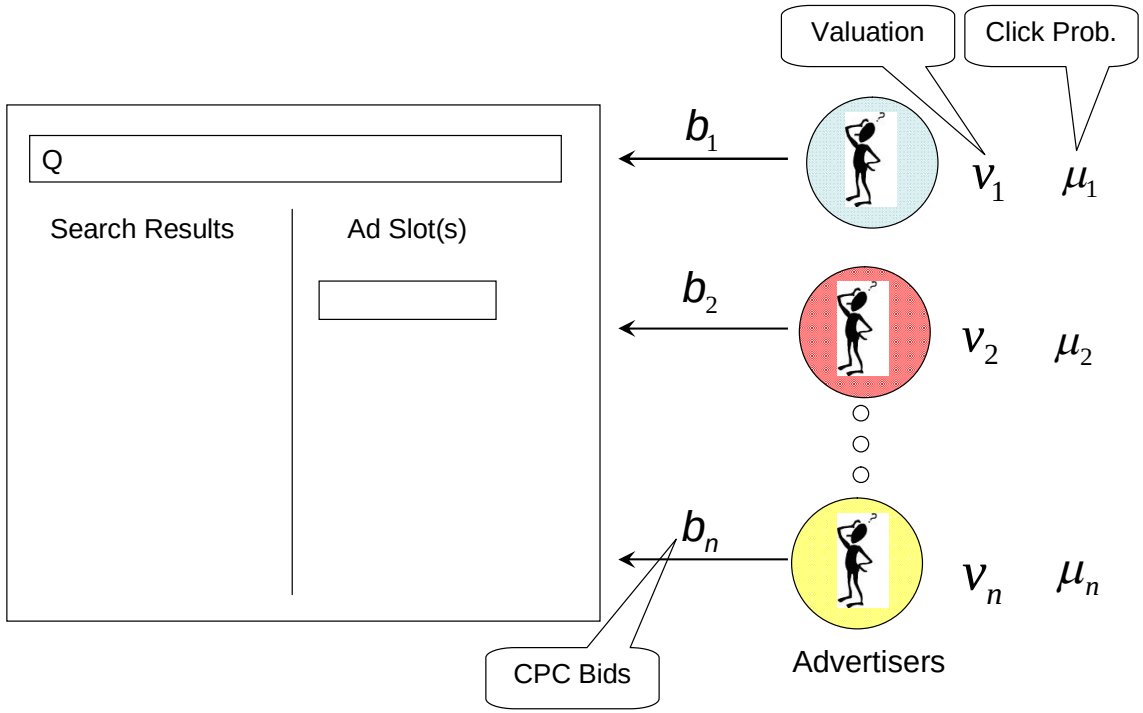
Example 1: Online Advertising







Pay-Per-Click Auctions



- Click probabilities (μ_i) are unknown to the search engine (aka auctioneer)
- Search engine needs to learn click probabilities
- Search engine employs an ad selection rule and a payment rule
- Ad selection rule and payment rule depend on bids b_i and learned $\hat{\mu}_i$
- Advertisers are strategic in reporting their bids b_i
- Search engine's goal: revenue maximization, welfare maximization

Modeling Ad-Auction as a (Forward) MAB Mechanism

- Advertisers $\Leftrightarrow$ Arms

Modeling Ad-Auction as a (Forward) MAB Mechanism

- Advertisers $\Leftrightarrow$ Arms
- $\mu_i \times v_i \Leftrightarrow$ expected value for arm i whenever it is pulled

Modeling Ad-Auction as a (Forward) MAB Mechanism

- Advertisers $\Leftrightarrow$ Arms
- $\mu_i \times v_i \Leftrightarrow$ expected value for arm i whenever it is pulled
- Payment made by arm i if its ad is clicked

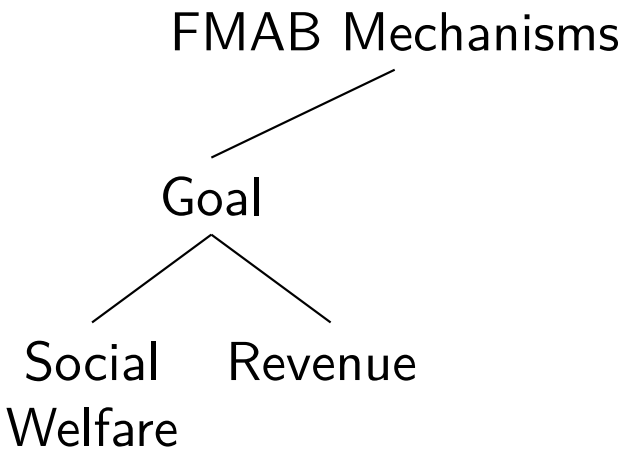
Modeling Ad-Auction as a (Forward) MAB Mechanism

- Advertisers $\Leftrightarrow$ Arms
- $\mu_i \times v_i \Leftrightarrow$ expected value for arm i whenever it is pulled
- Payment made by arm i if its ad is clicked

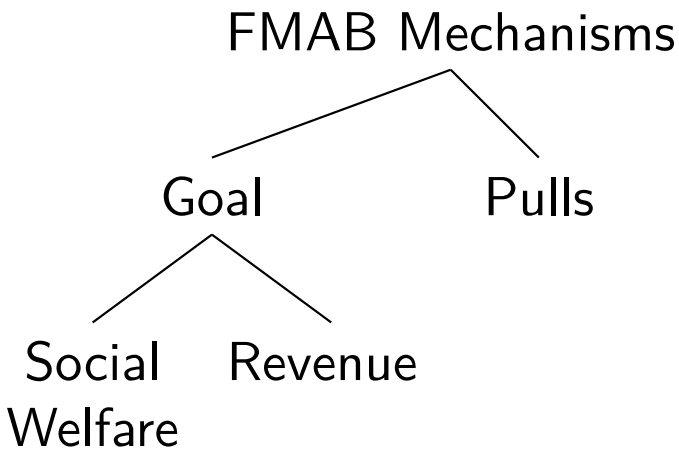
Given: Bids $b_1, b_2, \ldots, b_n$
(Forward) MAB Mechanism Design Problem:
Design an allocation rule (aka pulling strategy) and a payment rule such that

1. Arms bids **truthfully**
2. One of the following quantities gets maximized - *social welfare, revenue*
3. Individual Rationality

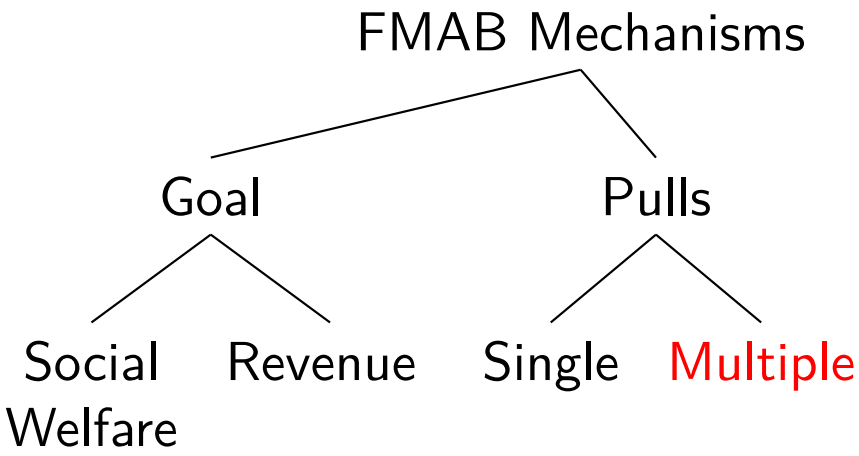
Taxonomy of (Forward) MAB Mechanism



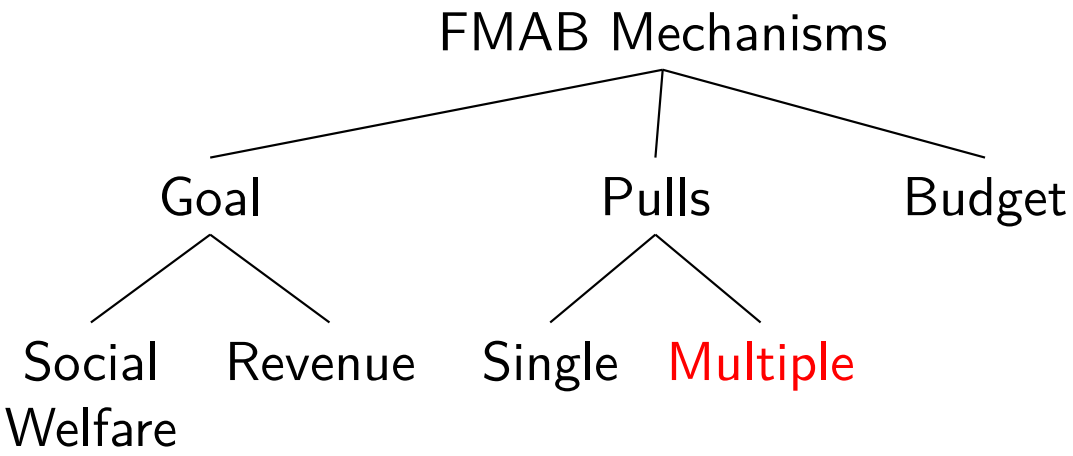
Taxonomy of (Forward) MAB Mechanism



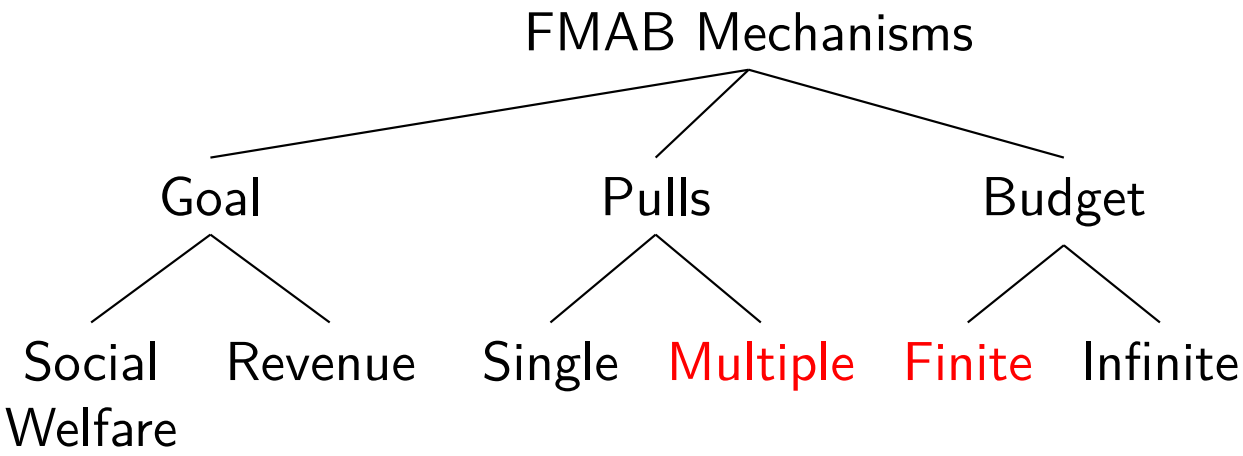
Taxonomy of (Forward) MAB Mechanism



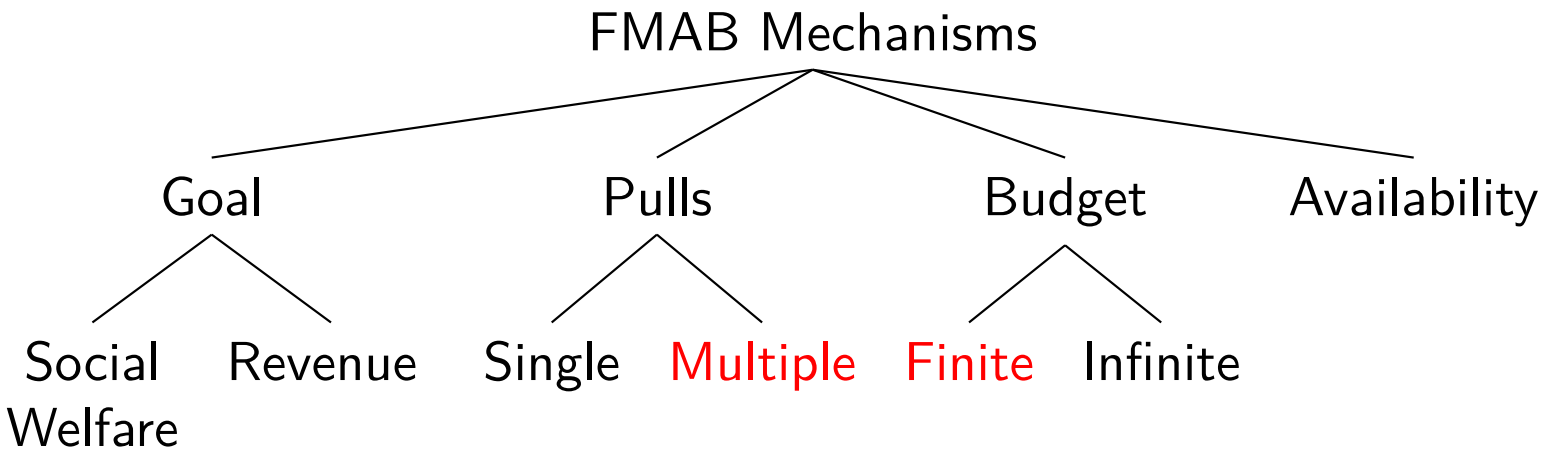
Taxonomy of (Forward) MAB Mechanism



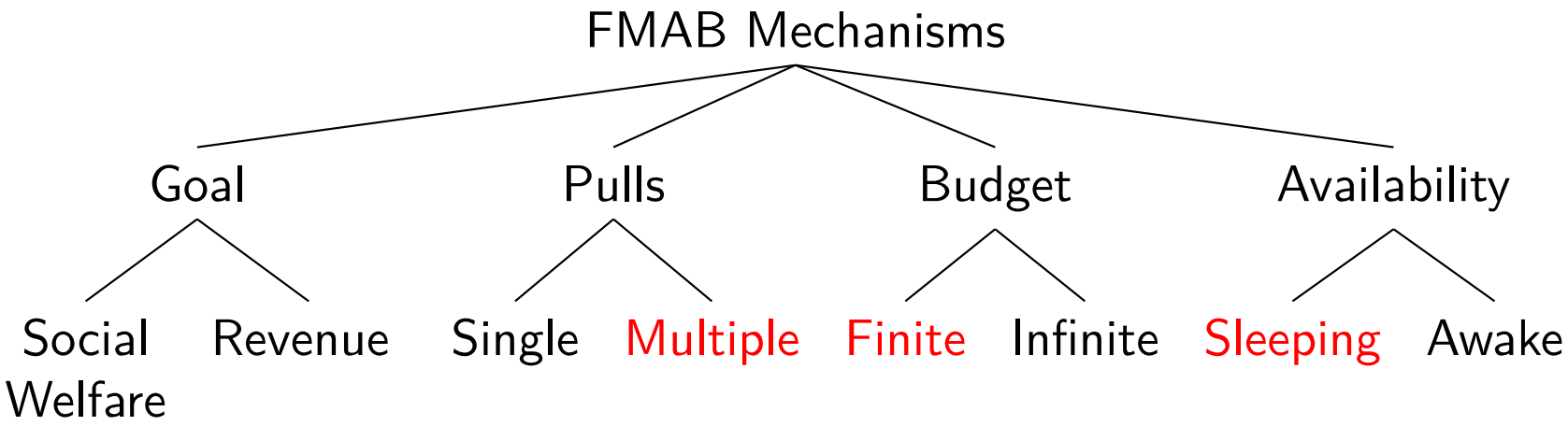
Taxonomy of (Forward) MAB Mechanism



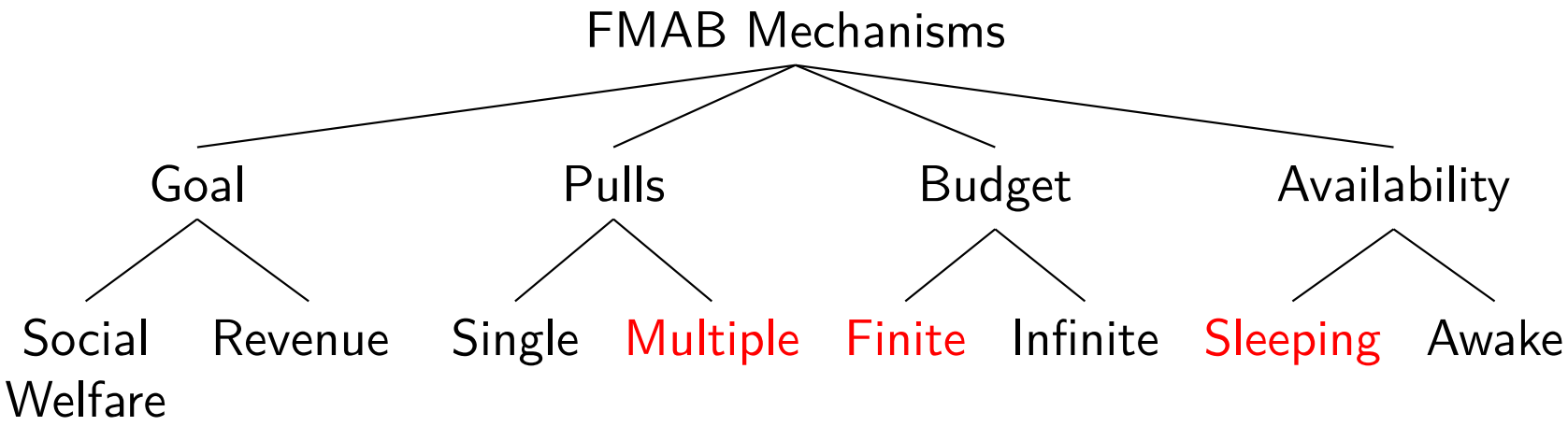
Taxonomy of (Forward) MAB Mechanism



Taxonomy of (Forward) MAB Mechanism



Taxonomy of (Forward) MAB Mechanism



	Single Pull	Multiple Pull
Social Welfare Maximization	Babaioff, Slivkins and Sharma, EC-2009 Babaioff, Kleinberg and Slivkins, EC-2010	Sharma, Gujar and Narahari, Current Science 2012 Mandal and Narahari, AAMAS 2014
Revenue Maximization	Devanur and Kakade, EC-2009	Gatti, Lazaric, and Trovo, EC2012

Characterization of Truthful (Forward) MAB Mechanisms

Definitions:

- *Deterministic MAB Allocation Rule:*
  $\mathcal{A}(b, \text{click realization}, t) = \text{Id of an arm to be pulled}$

Characterization of Truthful (Forward) MAB Mechanisms

Definitions:

- *Deterministic MAB Allocation Rule:*
  $\mathcal{A}(b, \text{click realization}, t) = \text{Id of an arm to be pulled}$
- *Ex-post Monotone MAB Allocation Rule:*
  For each $(b, \text{click realization}, t)$ tuple, if an agent is selected at this round, then she is also selected after increasing her bid (fixing everything else)

Characterization of Truthful (Forward) MAB Mechanisms

Definitions:

- *Deterministic MAB Allocation Rule:*
  $\mathcal{A}(b, \text{click realization}, t) = \text{Id of an arm to be pulled}$
- *Ex-post Monotone MAB Allocation Rule:*
  For each $(b, \text{click realization}, t)$ tuple, if an agent is selected at this round, then she is also selected after increasing her bid (fixing everything else)
- *Exploration Separated MAB Allocation Rule:*
  For any click realization, the allocation in any influential round does not depend on the bids.

Characterization of Truthful (Forward) MAB Mechanisms

Definitions:

- *Deterministic MAB Allocation Rule:*
  $\mathcal{A}(b, \text{click realization}, t) = \text{Id of an arm to be pulled}$
- *Ex-post Monotone MAB Allocation Rule:*
  For each $(b, \text{click realization}, t)$ tuple, if an agent is selected at this round, then she is also selected after increasing her bid (fixing everything else)
- *Exploration Separated MAB Allocation Rule:*
  For any click realization, the allocation in any influential round does not depend on the bids.
- *Influential Round:*
  A round t is influential round for a given click realization, with influenced agent j, if for some bid profile changing the click realization for this round can affect the allocation of agent j in some future round.

[BSS09] M. Babaioff, Y. Sharma, and A. Slivkins, "Characterizing Truthful Multi-Armed Bandit Mechanisms", EC 2009.

Characterization of Truthful (Forward) MAB Mechanisms

Theorem (BSS09)

Consider a (forward) MAB mechanism design problem. Let the allocation rule $\mathcal{A}(\cdot)$ be a deterministic allocation rule. Then

$\mathcal{A}(\cdot), P(\cdot)$ *is truthful* $\Leftrightarrow$ $\mathcal{A}(\cdot)$ *is ex-post monotone and exploration separable*

Characterization of Truthful (Forward) MAB Mechanisms

Theorem (BSS09)

Consider a (forward) MAB mechanism design problem. Let the allocation rule $\mathcal{A}(\cdot)$ be a deterministic allocation rule. Then

$\mathcal{A}(\cdot), P(\cdot)$ *is truthful* $\Leftrightarrow$ $\mathcal{A}(\cdot)$ *is ex-post monotone and exploration separable*

What about regret of such an allocation rule ?

Characterization of Truthful (Forward) MAB Mechanisms

Theorem (BSS09)

Consider a (forward) MAB mechanism design problem. Let the allocation rule $\mathcal{A}(\cdot)$ be a deterministic allocation rule. Then

$\mathcal{A}(\cdot), P(\cdot)$ *is truthful* $\Leftrightarrow$ $\mathcal{A}(\cdot)$ *is ex-post monotone and exploration separable*

What about regret of such an allocation rule ?

Regret for Welfare Maximization

Characterization of Truthful (Forward) MAB Mechanisms

Theorem (BSS09)

Consider a (forward) MAB mechanism design problem. Let the allocation rule $\mathcal{A}(\cdot)$ be a deterministic allocation rule. Then

$\mathcal{A}(\cdot), P(\cdot)$*is truthful* $\Leftrightarrow$ $\mathcal{A}(\cdot)$ *is ex-post monotone and exploration separable*

What about regret of such an allocation rule ?

Regret for Welfare Maximization

$$R(T) = T\mu_{i^*}v_{i^*} - \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n}\mu_i v_i \mathbb{1}_{\mathcal{A}(b,\text{click realization},t)=i}\right]$$

Theorem (BSS09)

Let $\mathcal{A}(\cdot)$ be an exploration-separated deterministic allocation rule. Then

$$R(T, v_{max}) = \Omega(v_{max} n^{1/3} T^{2/3})$$

A (Forward) MAB Mechanisms Achieving Lower Bound

Naive MAB Mechanism

1. For the first τ rounds, explore all the arms in a round robin fashion and charge nothing, where
$$\tau = n^{1/3} T^{2/3} (\log(T))^{1/3}$$
2. $\forall i$, compute the sample mean $\hat{\mu}_i$ for the click probability μ_i
3. $i^* \leftarrow \text{argmax}_i \ \hat{\mu}_i \times b_i$
4. For the remaining rounds, allocate i^* and charge him $\frac{smax(\hat{\mu}_i \times b_i)}{\hat{\mu_{i^*}}}$ if click happens

A (Forward) MAB Mechanisms Achieving Lower Bound

Naive MAB Mechanism

1. For the first τ rounds, explore all the arms in a round robin fashion and charge nothing, where
$$\tau = n^{1/3} T^{2/3} (\log(T))^{1/3}$$
2. $\forall i$, compute the sample mean $\hat{\mu}_i$ for the click probability μ_i
3. $i^* \leftarrow \text{argmax}_i \ \hat{\mu}_i \times b_i$
4. For the remaining rounds, allocate i^* and charge him $\frac{smax(\hat{\mu}_i \times b_i)}{\hat{\mu_{i^*}}}$ if click happens

Theorem (BSS09)

Naive MAB mechanism achieves a (welfare maximizing) regret of $O(v_{max} n^{1/3} T^{2/3} \log^{2/3} T)$

Revenue Maximization in (Forward) MAB Mechanisms

Regret for Revenue Maximization

Revenue Maximization in (Forward) MAB Mechanisms

Regret for Revenue Maximization

$$R(T) = \underbrace{T \times smax\{\mu_i b_i\}}_{\text{Revenue of Omniscient Vickrey Auction}} - \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} \mu_i P_i \mathbb{1}_{\mathcal{A}(b,\text{click realization},t)=i}\right]}_{\text{Revenue of any MAB Mechanism}}$$

Revenue Maximization in (Forward) MAB Mechanisms

Regret for Revenue Maximization

$$R(T) = \underbrace{T \times smax\{\mu_i b_i\}}_{\text{Revenue of Omniscient Vickrey Auction}} - \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} \mu_i P_i \mathbb{1}_{\mathcal{A}(b,\text{click realization},t)=i}\right]}_{\text{Revenue of any MAB Mechanism}}$$

Theorem (DK09)

The worst-case (revenue maximizing) regret of any truthful (forward) MAB is $\Omega(T^{2/3})$

Revenue Maximization in (Forward) MAB Mechanisms

Regret for Revenue Maximization

$$R(T) = \underbrace{T \times smax\{\mu_i b_i\}}_{\text{Revenue of Omniscient Vickrey Auction}} - \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} \mu_i P_i \mathbb{1}_{\mathcal{A}(b,\text{click realization},t)=i}\right]}_{\text{Revenue of any MAB Mechanism}}$$

Theorem (DK09)

The worst-case (revenue maximizing) regret of any truthful (forward) MAB is $\Omega(T^{2/3})$

Theorem (DK09)

There exists a truthful (forward) MAB mechanism for whom the worst-case (revenue maximizing) regret is $\Omega(n^{1/3}T^{2/3})$

[BSS09] N. Devanur, and S. Kakade, "The Price of Truthfulness for Pay-Per-Click Auctions", EC 2009.

Improving the Regret Bounds Further

Recall To ensure truthfulness,

- A deterministic MAB mechanisms must be *exploration-separated*
- It incurs high regret $O(n^{1/3}T^{2/3})$

Improving the Regret Bounds Further

Recall To ensure truthfulness,

- A deterministic MAB mechanisms must be *exploration-separated*
- It incurs high regret $O(n^{1/3}T^{2/3})$

Can randomization Help?

Improving the Regret Bounds Further

Recall To ensure truthfulness,

- A deterministic MAB mechanisms must be *exploration-separated*
- It incurs high regret $O(n^{1/3}T^{2/3})$

Can randomization Help?

- *Stochastically Monotone MAB Allocation Rule:*
  Allocation rule is monotone in expectation over nature randomness (i.e. click realizations).

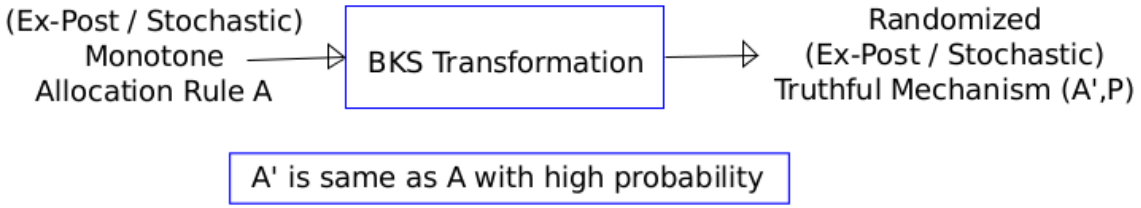
Improving the Regret Bounds Further

Recall To ensure truthfulness,

- A deterministic MAB mechanisms must be *exploration-separated*
- It incurs high regret $O(n^{1/3}T^{2/3})$

Can randomization Help?

- *Stochastically Monotone MAB Allocation Rule:*
  Allocation rule is monotone in expectation over nature randomness (i.e. click realizations).



Theorem (BKS10)

- *UCB algorithm is stochastically-monotone with regret $O(\sqrt{nT \log T})$. Hence, BKS transformation gives a **stochastically** truthful with same regret.*
- *A modified version of UCB is ex-post monotone with regret $O(\sqrt{nT \log T})$. Hence, BKS transformation gives an **ex-post** truthful with same regret.*

Extension to Multi Slot (aka Batch Pull) Setting

Externalities in Click Probabilities

For multi-slot sponsored search auction, the click probability of an ad is determined by two types of externalities :

- Position-Dependent Externality
- Ad-Dependent Externality

GLT12

A generalization of exploration-separated mechanism presented in DK09 with the following worst-case regret guarantees

- Position-Dependent Externality : $O(n^{\frac{1}{3}} k^{\frac{2}{3}} T^{\frac{2}{3}})$
- Ad-Dependent Externality : $O(nk^{\frac{2}{3}} T^{\frac{2}{3}})$

MN14

- Study the regret of ex-post monotone allocation rule for multi-slot sponsored search
- For more details visit the poster !!

[GLT12] N. Gatti, A. Lazaric, F. Trovo, "A truthful learning mechanism for contextual multi-slot sponsored search auctions with externalities", EC 2012

[MN14] D. Mandal, Y. Narahari, "A Novel Ex-post Truthful Mechanism for Multi-Slot Sponsored Search Auctions", AAMAS 2014

Outline

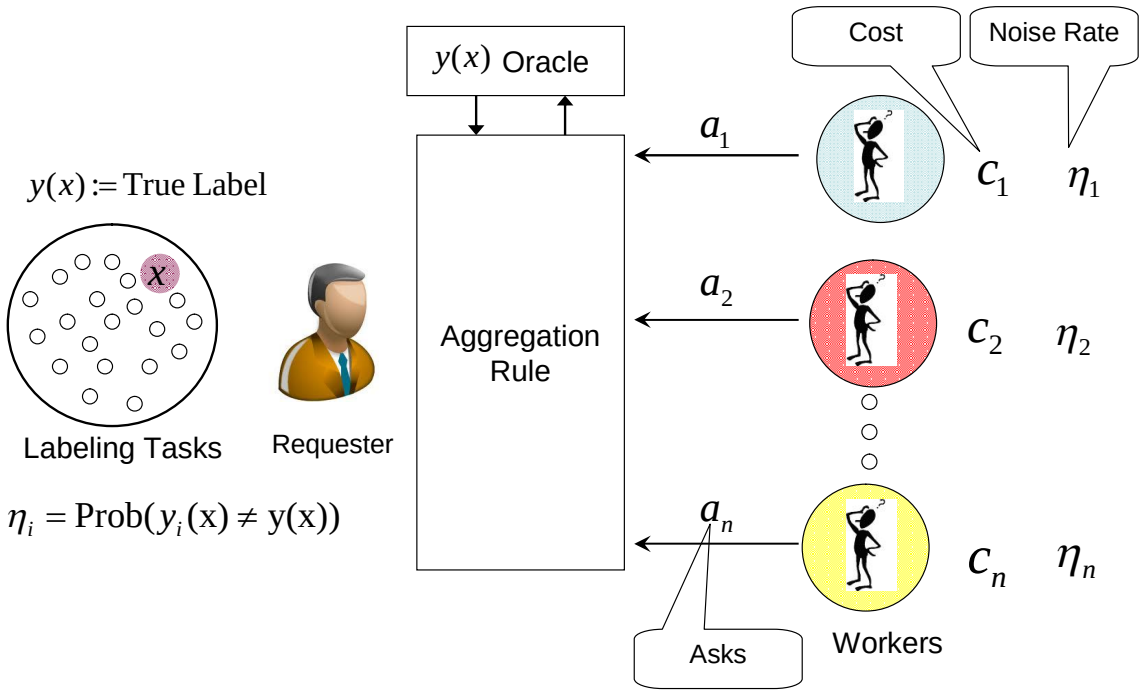
[Part 1] (Forward) MAB Mechanisms

- Motivating example
- Taxonomy of (Forward) MAB Mechanisms
- Characterization for truthfulness
- Welfare maximizing (Forward) MAB mechanisms
- Revenue maximizing (Forward) MAB mechanisms

[Part 2] (Reverse) MAB mechanisms

- Motivating example
- Taxonomy of (Reverse) MAB Mechanisms
- Recent results on (Reverse) MAB mechanisms

Example 2: Crowdsourcing the Labels



- Noise rates (ρ_i) are unknown to the requester (aka auctioneer)
- Requester needs to learn noise rates
- Requester employs a worker(s) selection rule and a payment rule
- Worker(s) selection rule and payment rule are functions of a_i and learned $\hat{\rho}_i$
- Workers are strategic in reporting their asks a_i
- Requester's goal: error minimization, cost minimization

Modelling Crowdsourcing as MAB mechanism

- Workers $\Leftrightarrow$ Arms
- $\eta_i \times l + (1 - \eta_i) \times r \Leftrightarrow$ expected value received by the requester whenever arm i is pulled
- $-c_i \Leftrightarrow$ value received by the arm whenever it is pulled

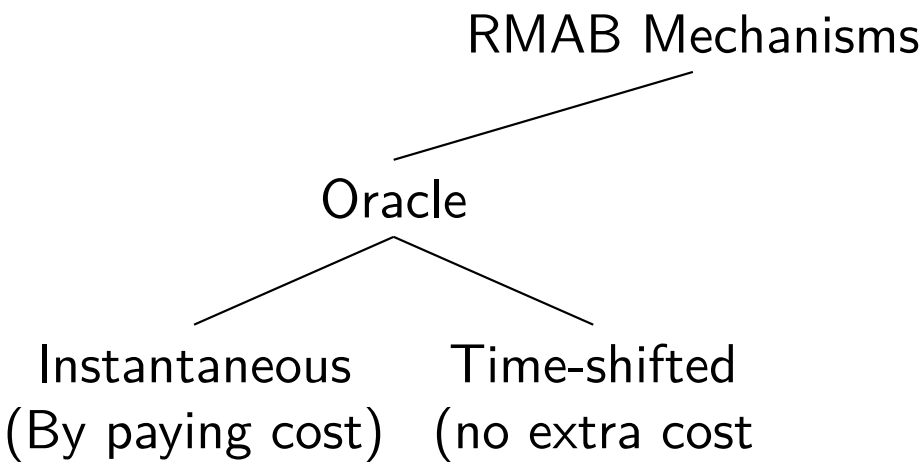
Remark 1: Often, payment has to be made to the crowdworkers irrespective of correctness of their reported labels.

Given: Asks $a_1, a_2, \ldots, a_n$
(Reverse) MAB Mechanism Design Problem:
Design an allocation rule (aka pulling strategy) and a payment rule such that

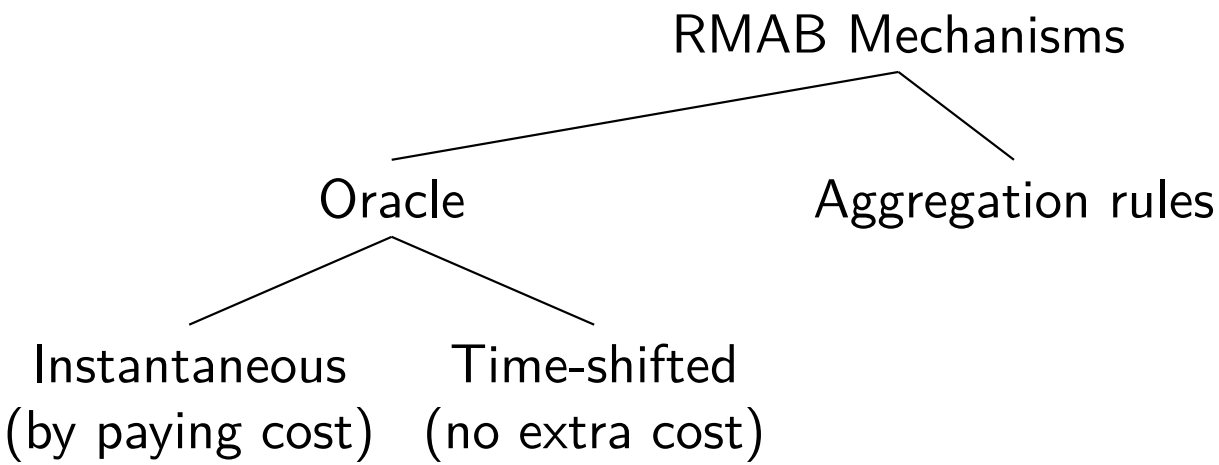
1. Arms asks **truthfully**
2. One of the following quantities gets minimized - *overall labeling error, overall cost of labeling*
3. Individual Rationality

Remark 2: If requester has an infinite procurement budget then the problem reduces to learning from experts.

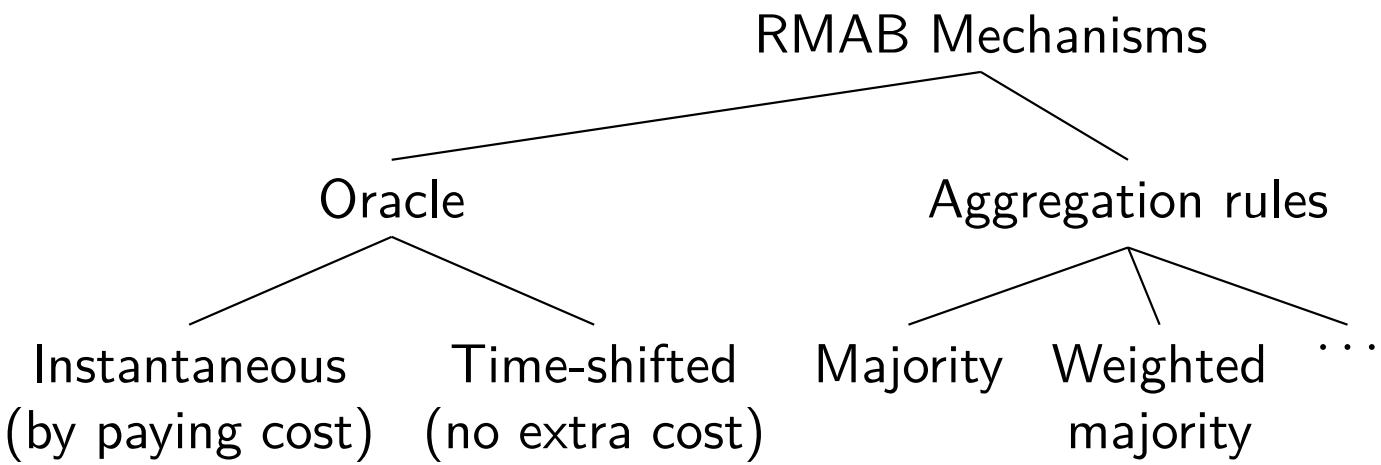
Taxonomy of (Reverse) MAB mechanism



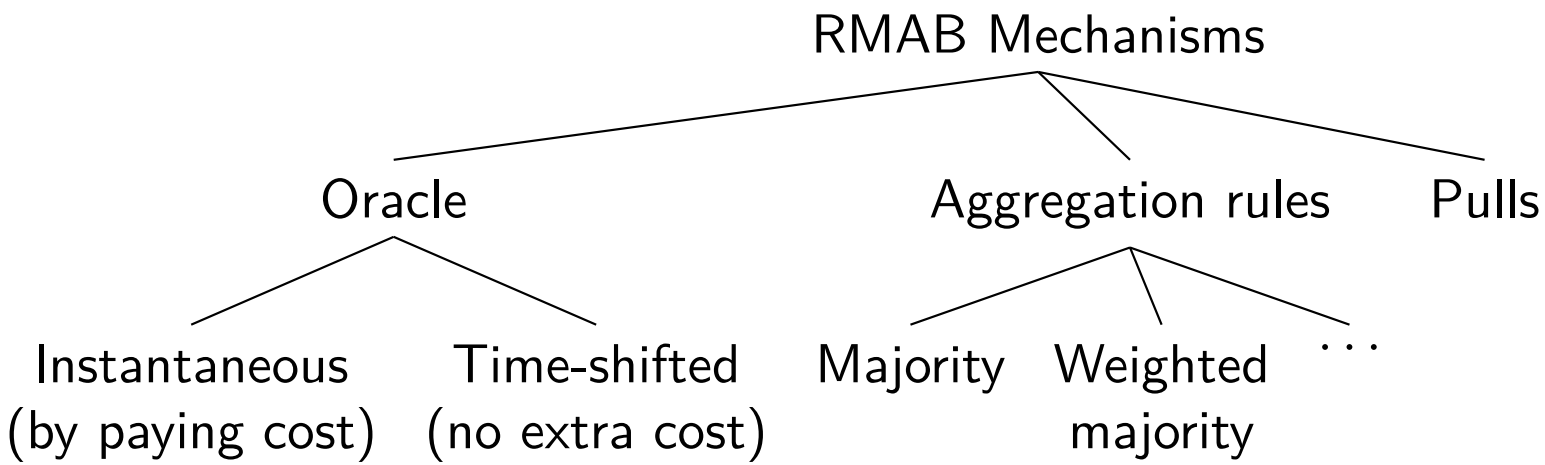
Taxonomy of (Reverse) MAB mechanism



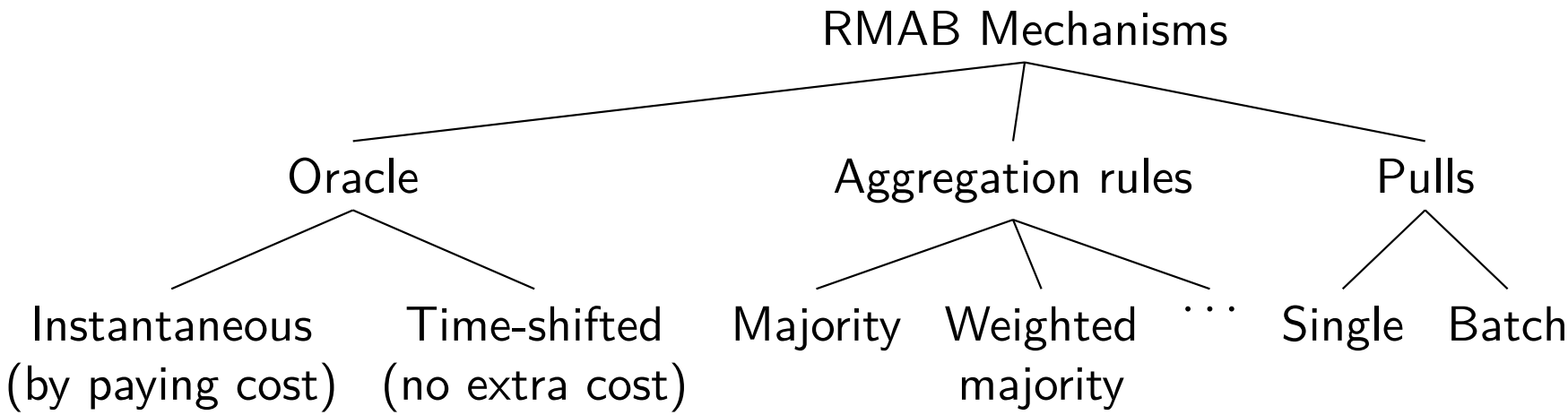
Taxonomy of (Reverse) MAB mechanism



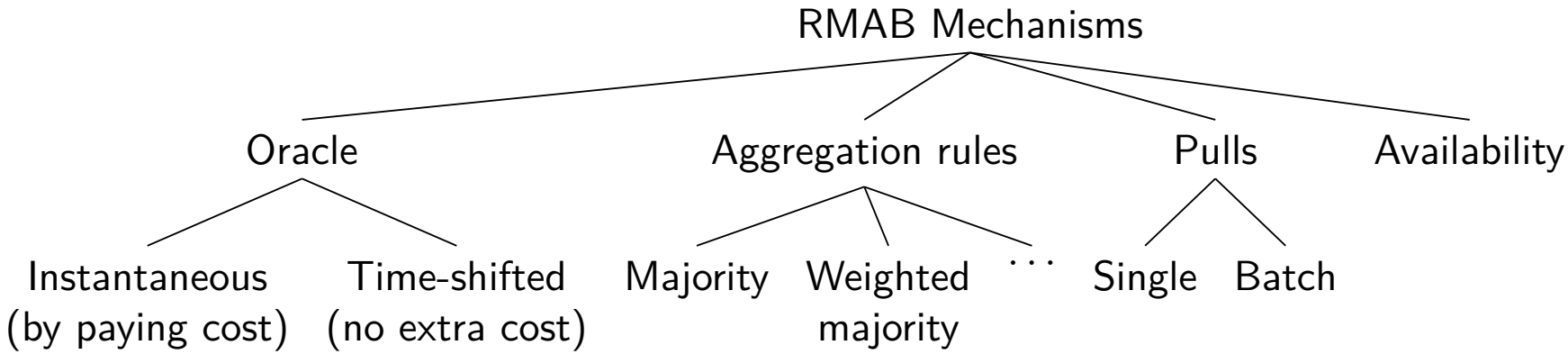
Taxonomy of (Reverse) MAB mechanism



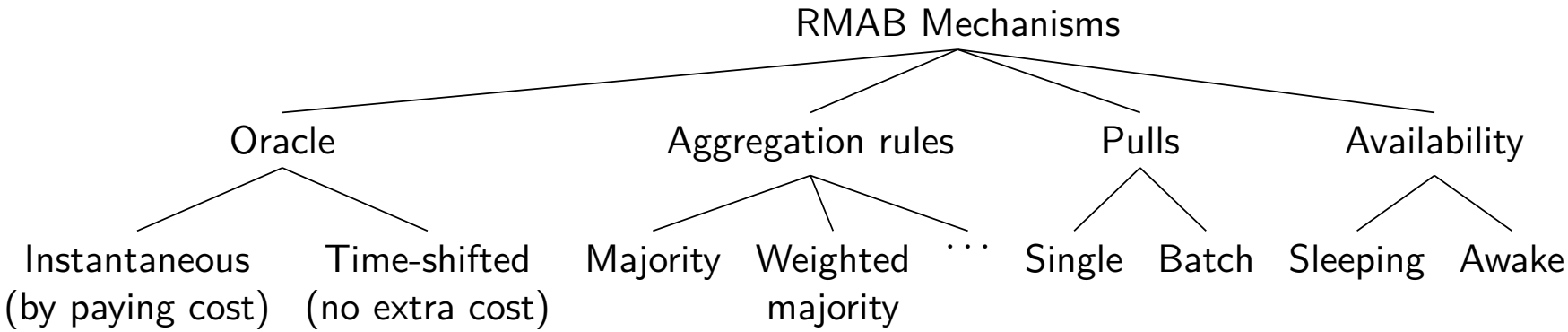
Taxonomy of (Reverse) MAB mechanism



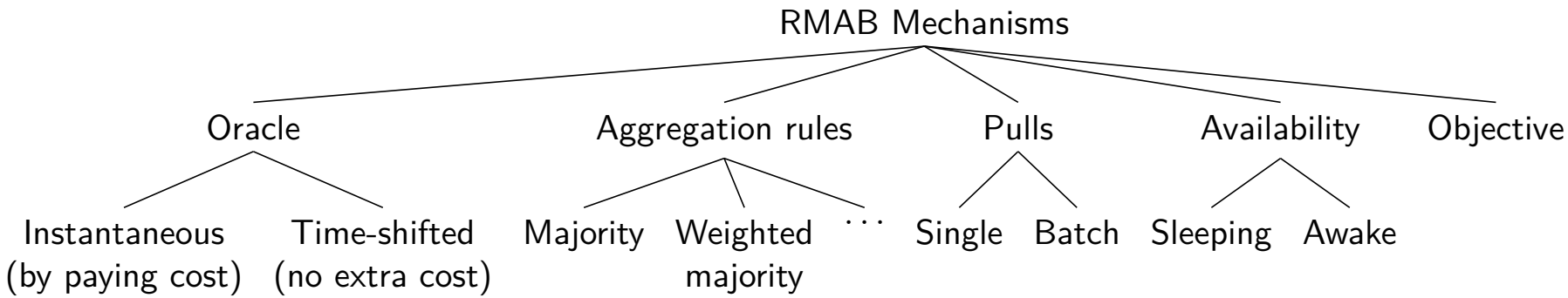
Taxonomy of (Reverse) MAB mechanism



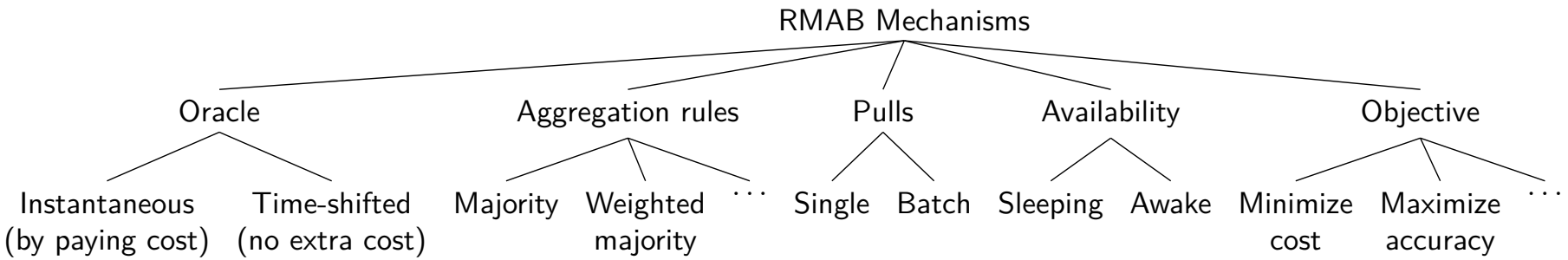
Taxonomy of (Reverse) MAB mechanism



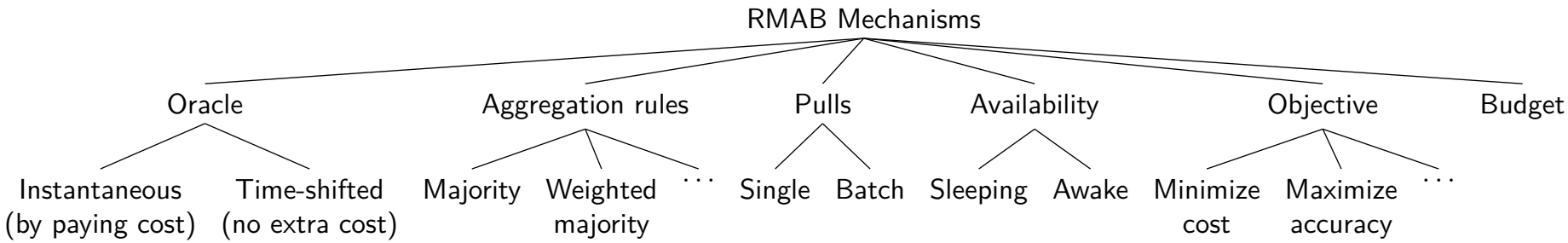
Taxonomy of (Reverse) MAB mechanism



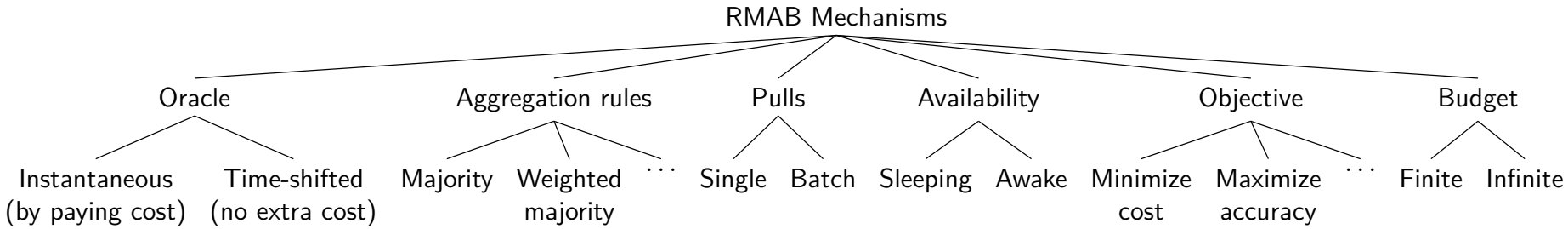
Taxonomy of (Reverse) MAB mechanism



Taxonomy of (Reverse) MAB mechanism



Taxonomy of (Reverse) MAB mechanism



Cost Minimizing Truthful (Reverse) MAB Mechanism

Given:

- Asks $a_1, a_2, \ldots, a_n$
- Instantaneous verification of labels is possible (with zero payment to oracle)
- Aggregation rule is *majority voting rule*

(Reverse) MAB Mechanism Design Problem:
Design an allocation rule (aka pulling strategy) and a payment rule such that

1. Arms report their costs **truthfully**
2. Overall cost of labeling is minimized and for every label an accuracy level is maintained
3. Individual Rationality

Solution: please visit the poster!!

[JGZN14] S. Jain, S. Gujar, O. Zoeter, and Y. Narahari, "A quality assuring MAB mechanism with Incentive Compatible Learning", AAMAS 2014

Summary

MAB mechanisms

- An upcoming area
- Generic framework to model a rich class of contemporary problems
- Forward MAB mechanisms have seen some progress in recent times but still there are many open issues
- Reverse MAB mechanisms are much less explored

Thank You!

Questions?