Spotify ML Day, July 9th, 2018

Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits

James McInerney, Ben Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, Rishabh Mehrotra

Spotify

email: jamesm@spotify.com

Talk Outline

1. Pareto principle for content producers
2. a causal diagnosis of filter bubbles in recommendation
3. contextual bandits for recommendation
4. explained recommendations
5. introducing Bart (bandits for recommendations as treatments)
6. offline and online experiments on homepage data
7. conclusions & future work

A small number of producers dominate consumption in culture

producer popularity

leads to leads to

exposure to new consumers

A small number of producers dominate consumption in culture

e.g. musicians, authors, actors

producer popularity

leads to

leads to

exposure to new consumers

A small number of producers dominate consumption in culture

e.g. musicians, authors, actors

producer popularity

leads to

leads to

exposure to new consumers

e.g. word of mouth, media coverage

A small number of producers dominate consumption in culture

e.g. musicians, authors, actors

producer popularity

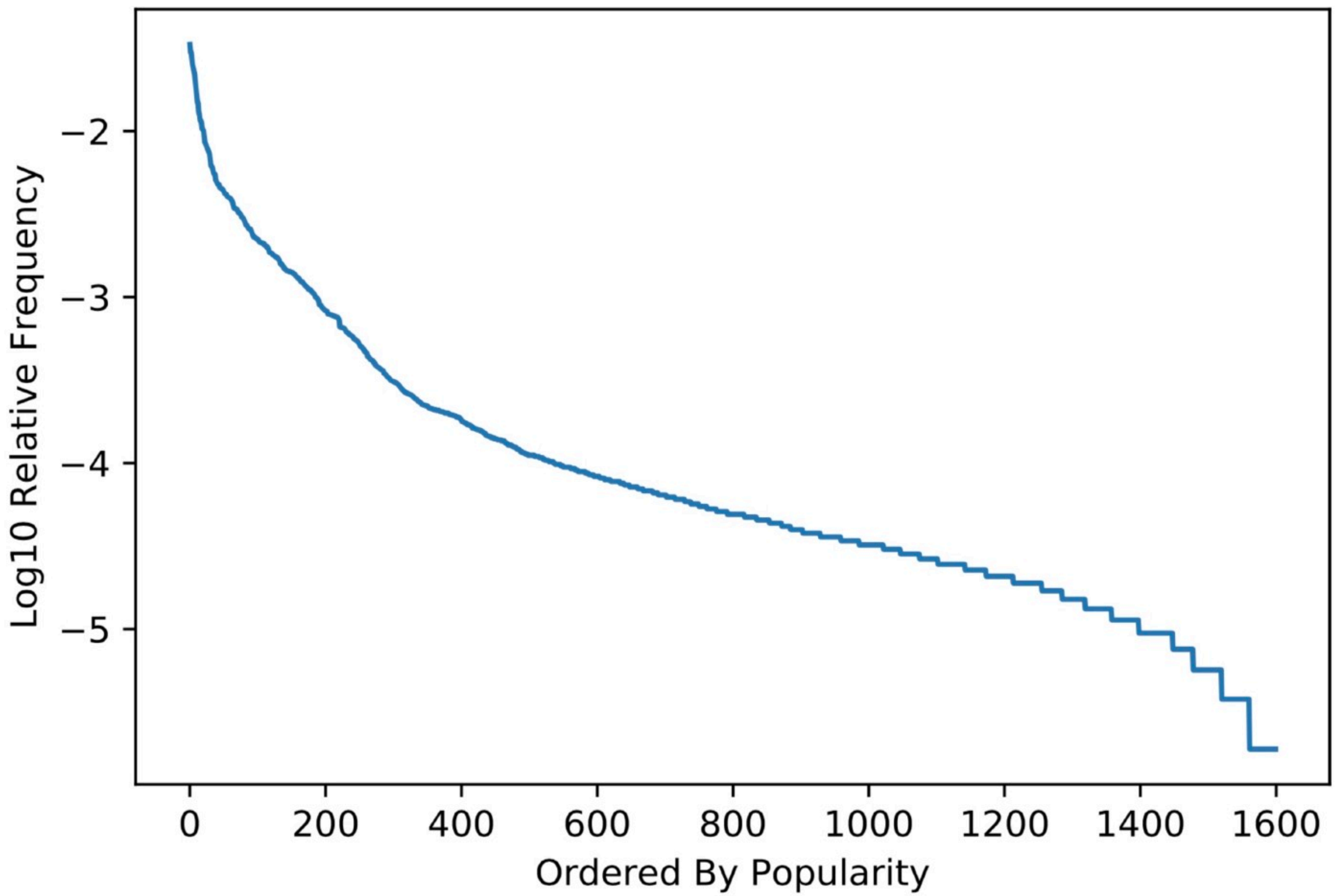
leads to leads to

exposure to new consumers

e.g. word of mouth, media coverage

- known as the Matthew effect or Pareto principle

A small number of producers dominate consumption in culture



Collaborative filtering perpetuates the Pareto principle

e.g. matrix factorization

$$\mathbf{Y} = \mathbf{U}\mathbf{V}^\top$$

#users $\times$ #items

#users $\times$ K

#items $\times$ K

Collaborative filtering perpetuates the Pareto principle

e.g. matrix factorization

$$\mathbf{Y} = \mathbf{U}\mathbf{V}^\top$$

$\underbrace{\#\text{users} \times \#\text{items}}_{\text{logged feedback data}}$ $\#\text{users} \times K$ $\#\text{items} \times K$

Collaborative filtering perpetuates the Pareto principle

e.g. matrix factorization

$$\mathbf{Y} = \mathbf{U}\mathbf{V}^{\top}$$

$$\underbrace{\#\text{users} \times \#\text{items}}_{\text{logged feedback data}} \qquad \underbrace{\#\text{users} \times K \quad \#\text{items} \times K}_{\substack{\text{latent variables fitted} \\ \text{to logged feedback data}}}$$

Collaborative filtering perpetuates the Pareto principle

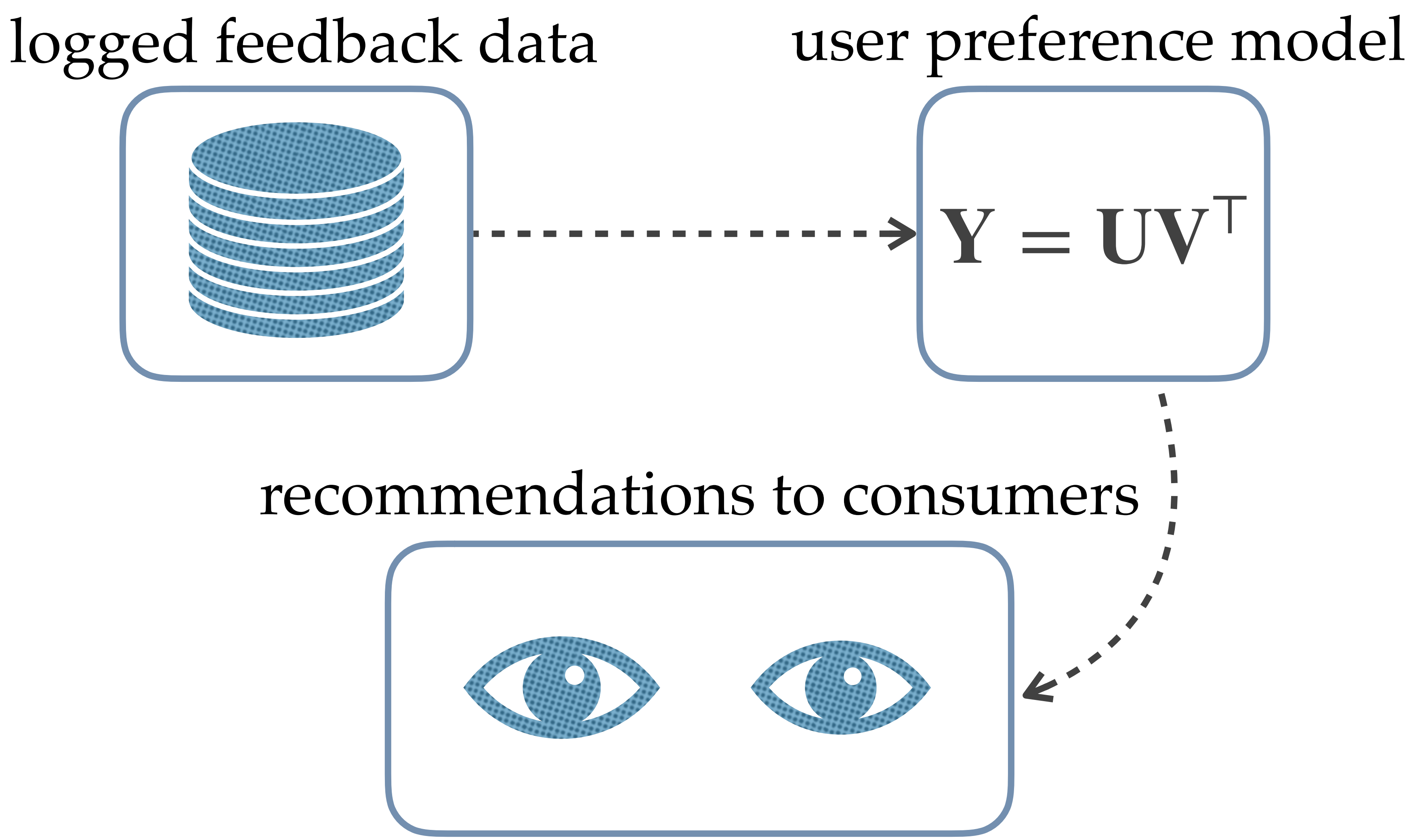
e.g. matrix factorization

$$\mathbf{Y} = \mathbf{U}\mathbf{V}^\top$$

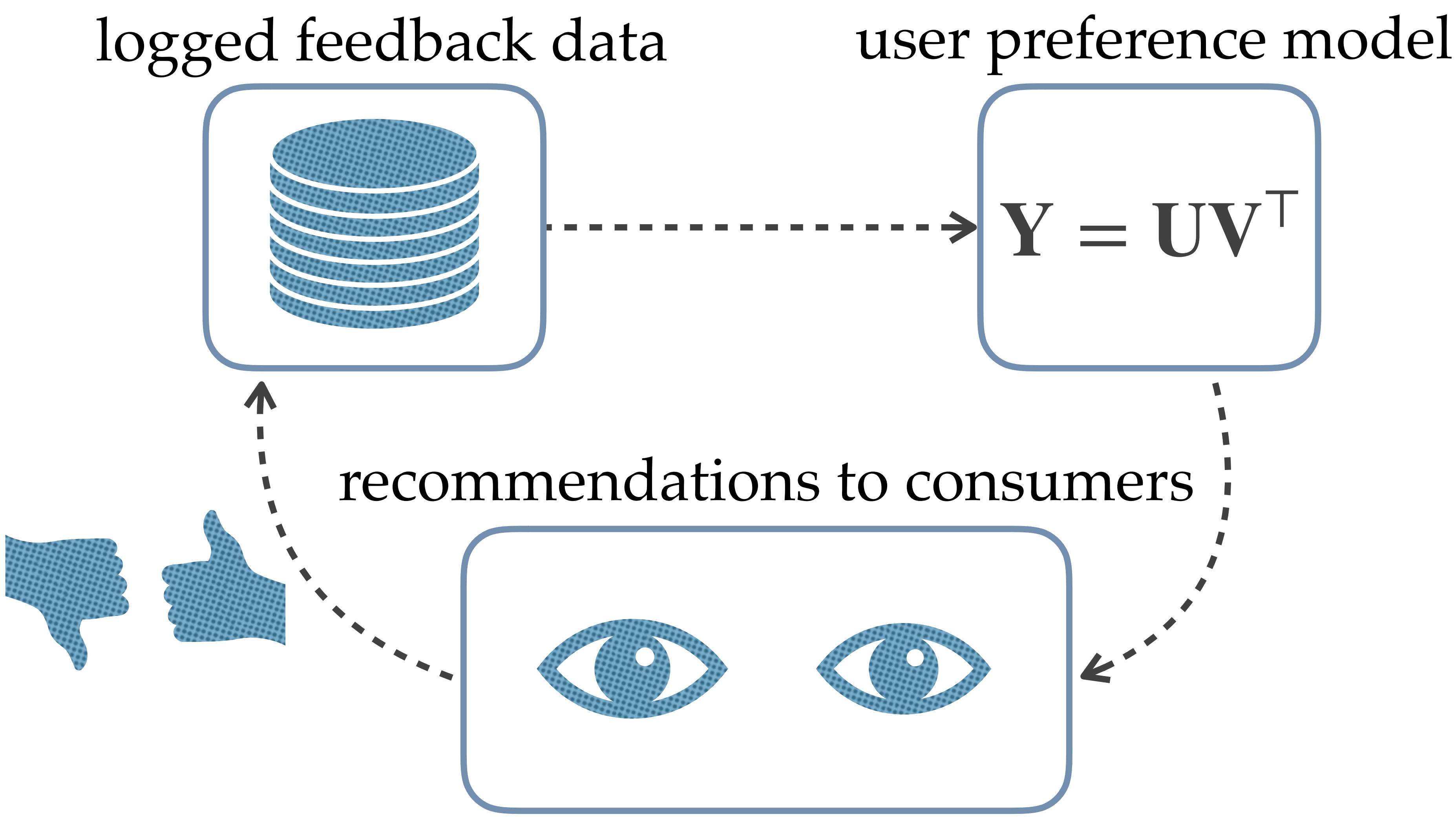
$$\underbrace{\#\text{users} \times \#\text{items}}_{\text{logged feedback data}} \qquad \underbrace{\#\text{users} \times K \quad \#\text{items} \times K}_{\text{latent variables fitted to logged feedback data}}$$

- in general: collaborative filtering engines use implicit feedback data from users to learn a model of user preferences

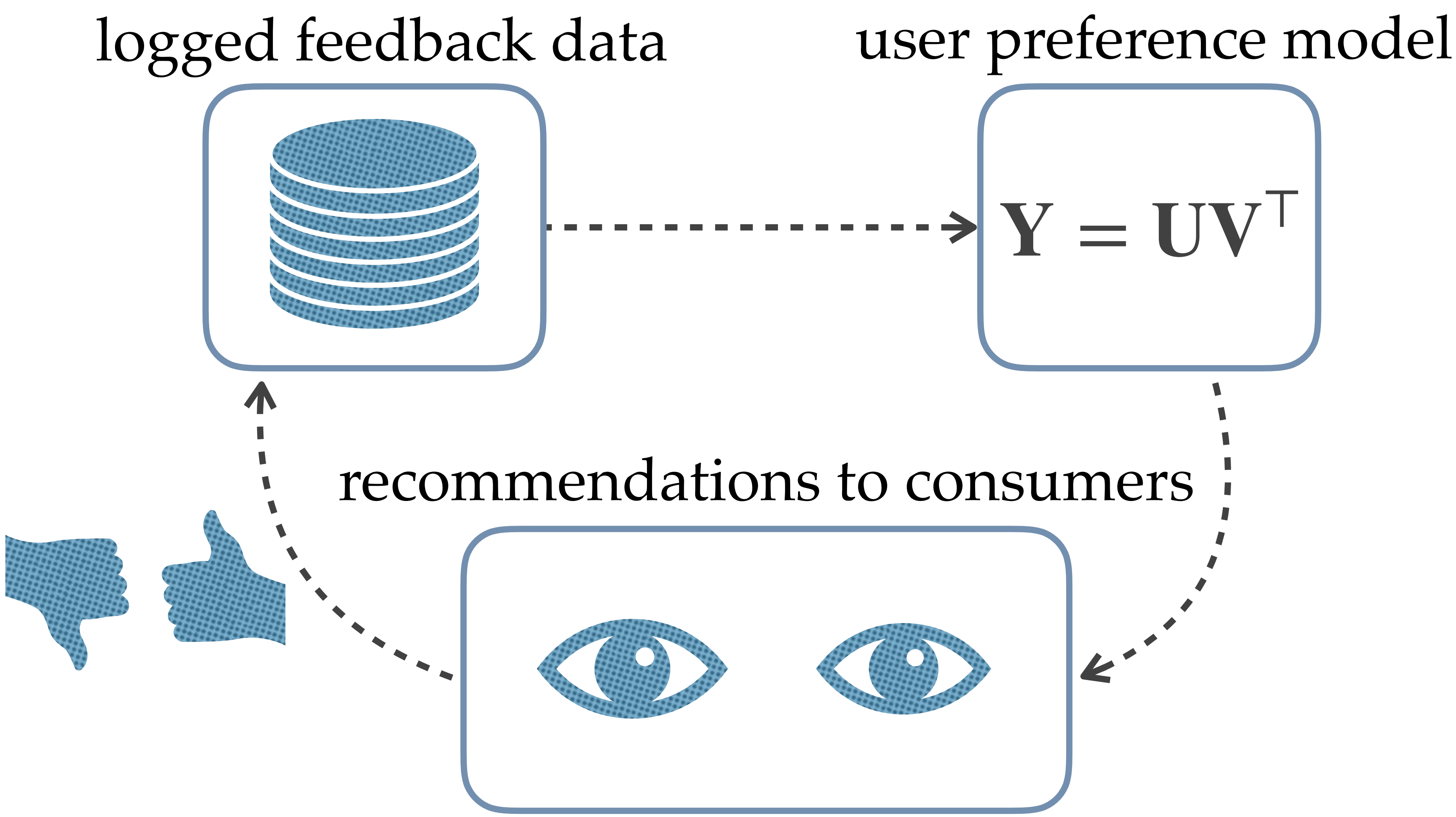
Collaborative filtering perpetuates the Pareto principle



Collaborative filtering perpetuates the Pareto principle



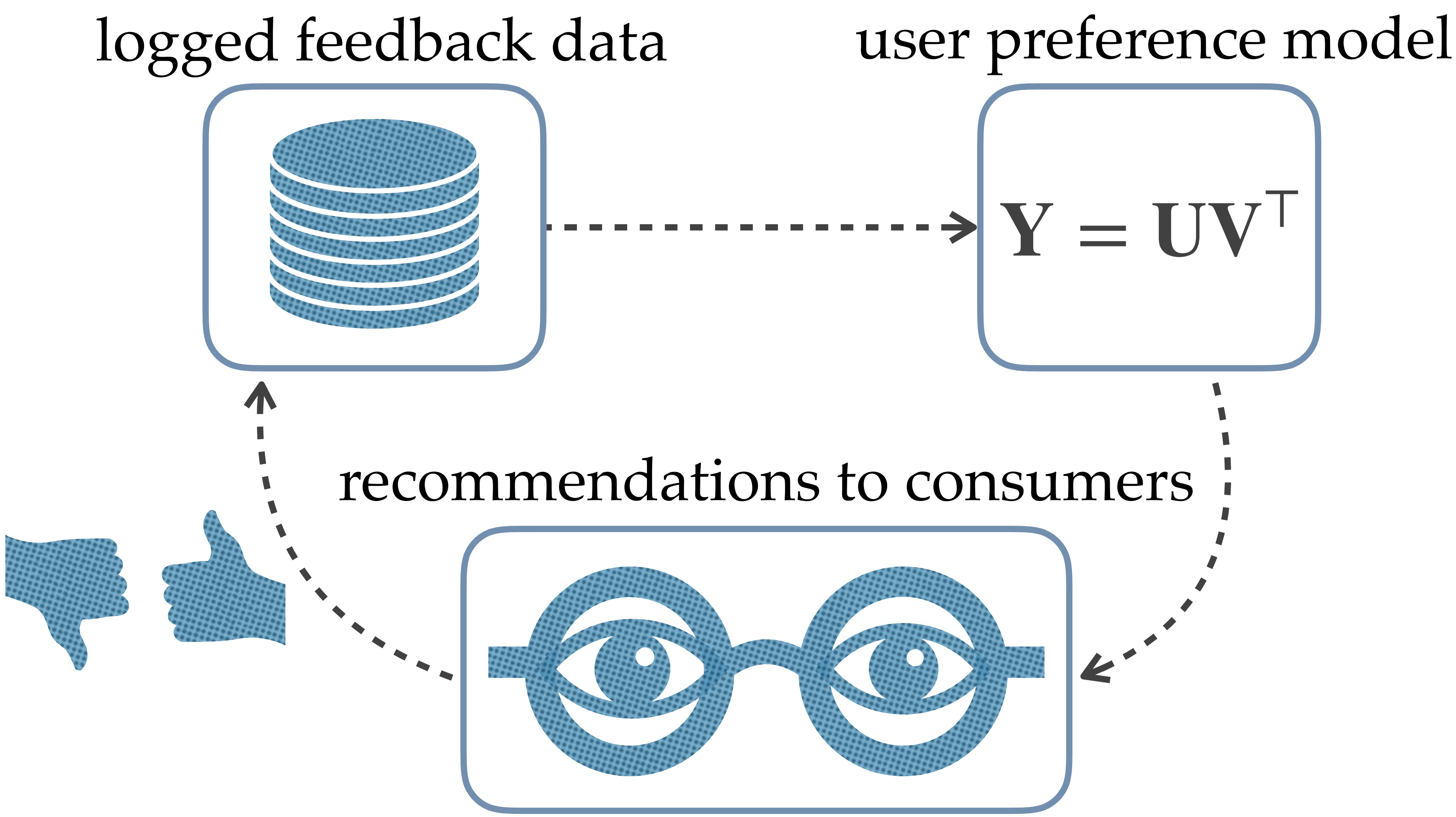
Collaborative filtering perpetuates the Pareto principle



"How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility" (Chaney et al. 2017)

"Modeling User Exposure in Recommendation" (Liang et al. 2016)

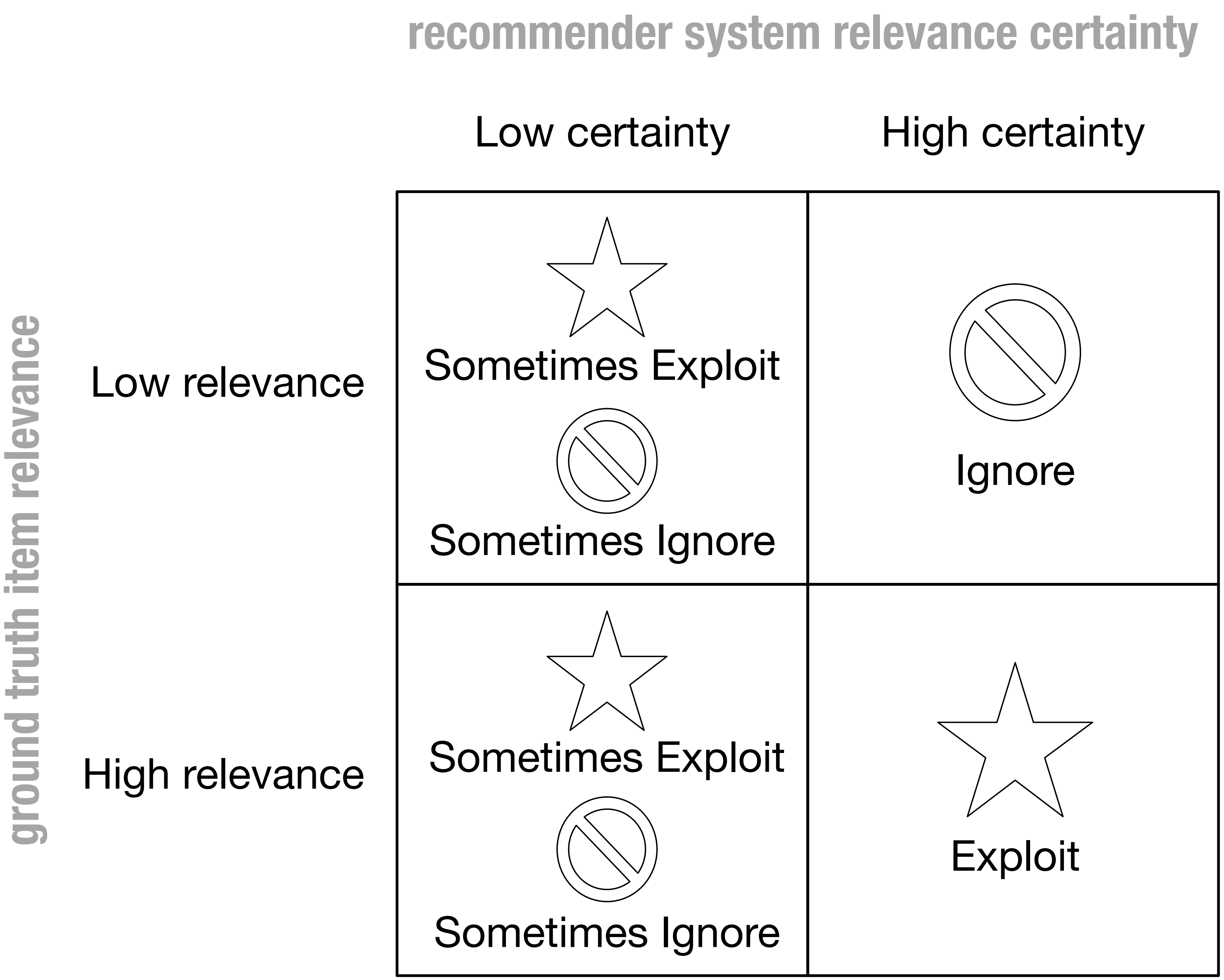
Collaborative filtering perpetuates the Pareto principle



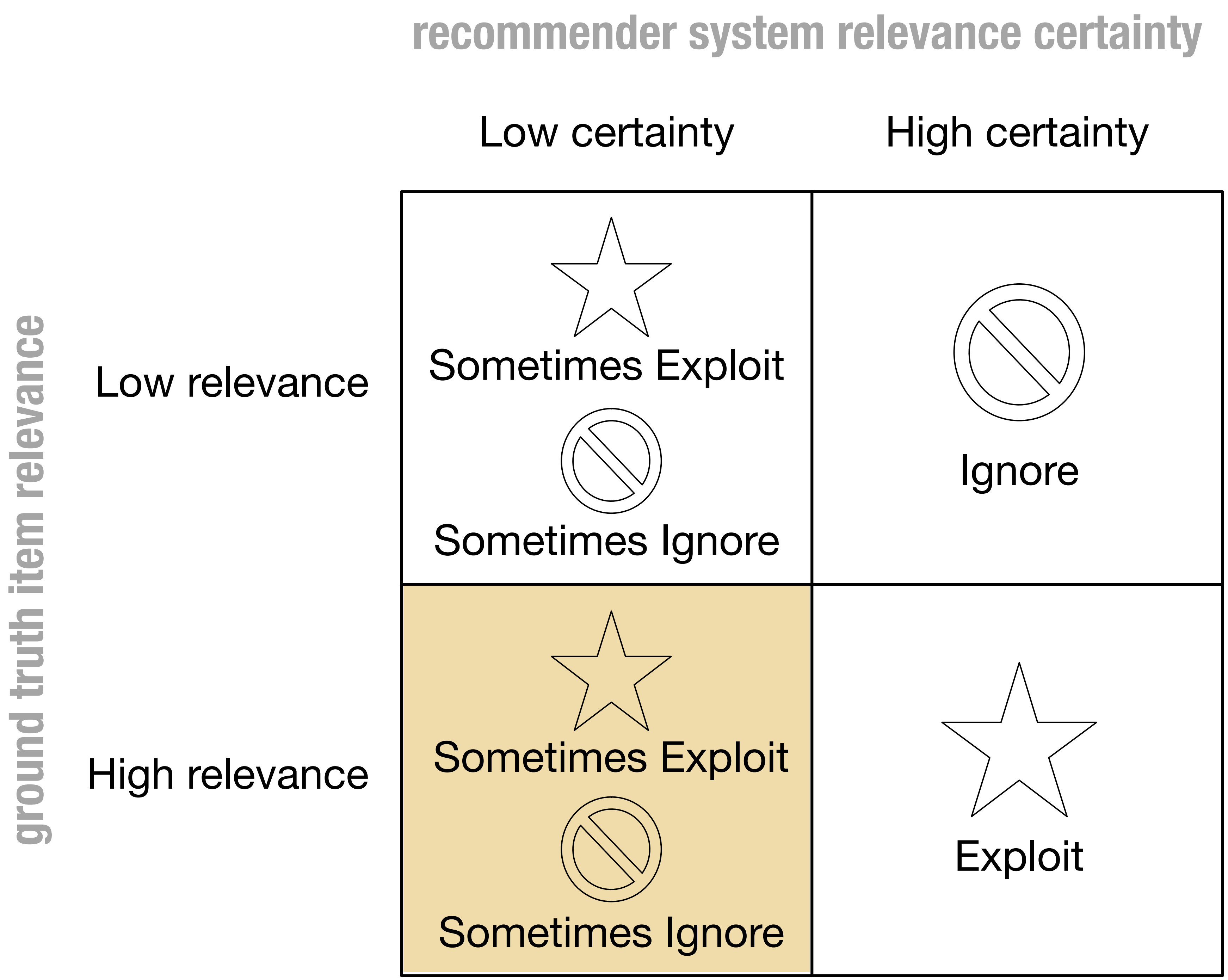
"How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility" (Chaney et al. 2017)

"Modeling User Exposure in Recommendation" (Liang et al. 2016)

Standard collaborative filtering methods are limited because they can only exploit or ignore

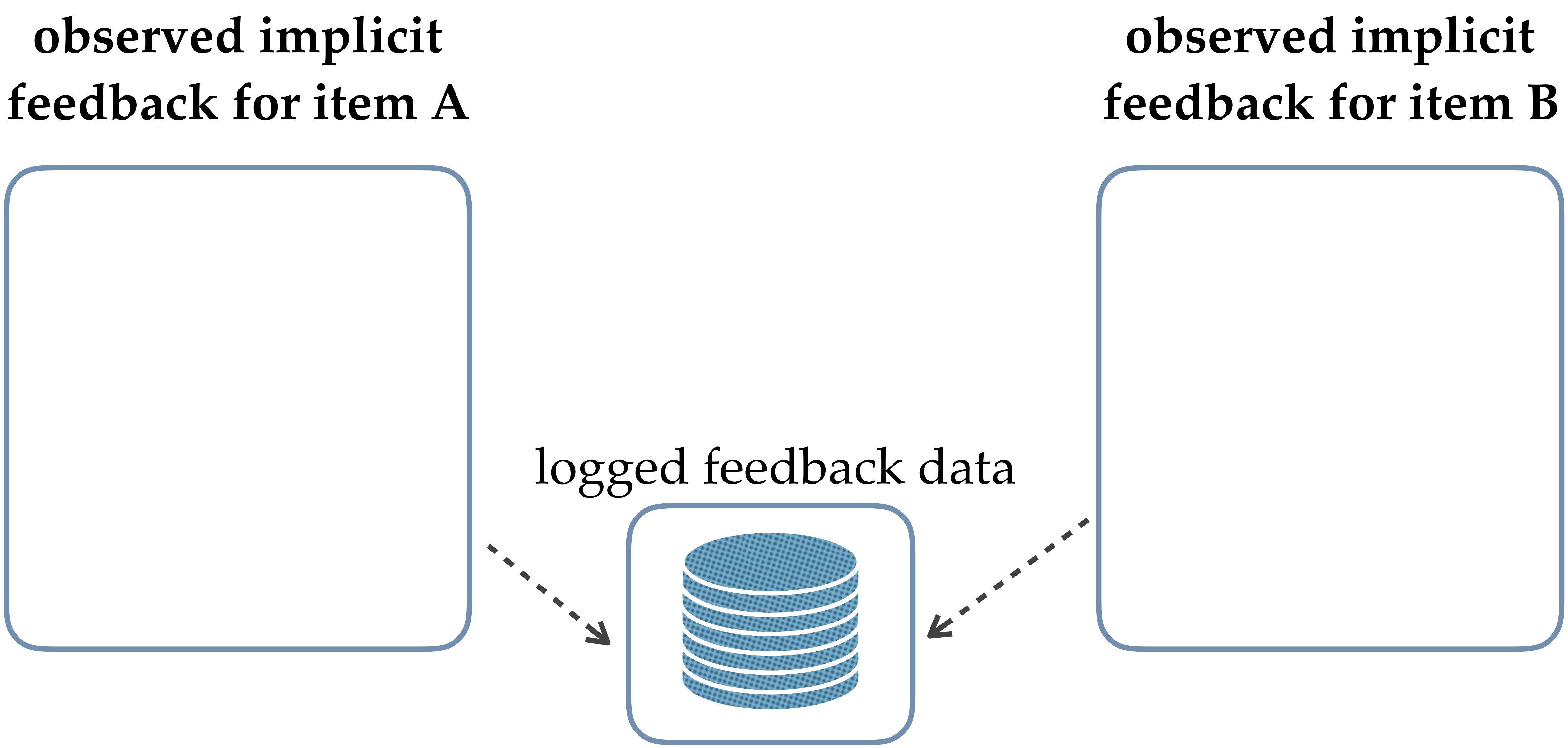


Standard collaborative filtering methods are limited because they can only exploit or ignore



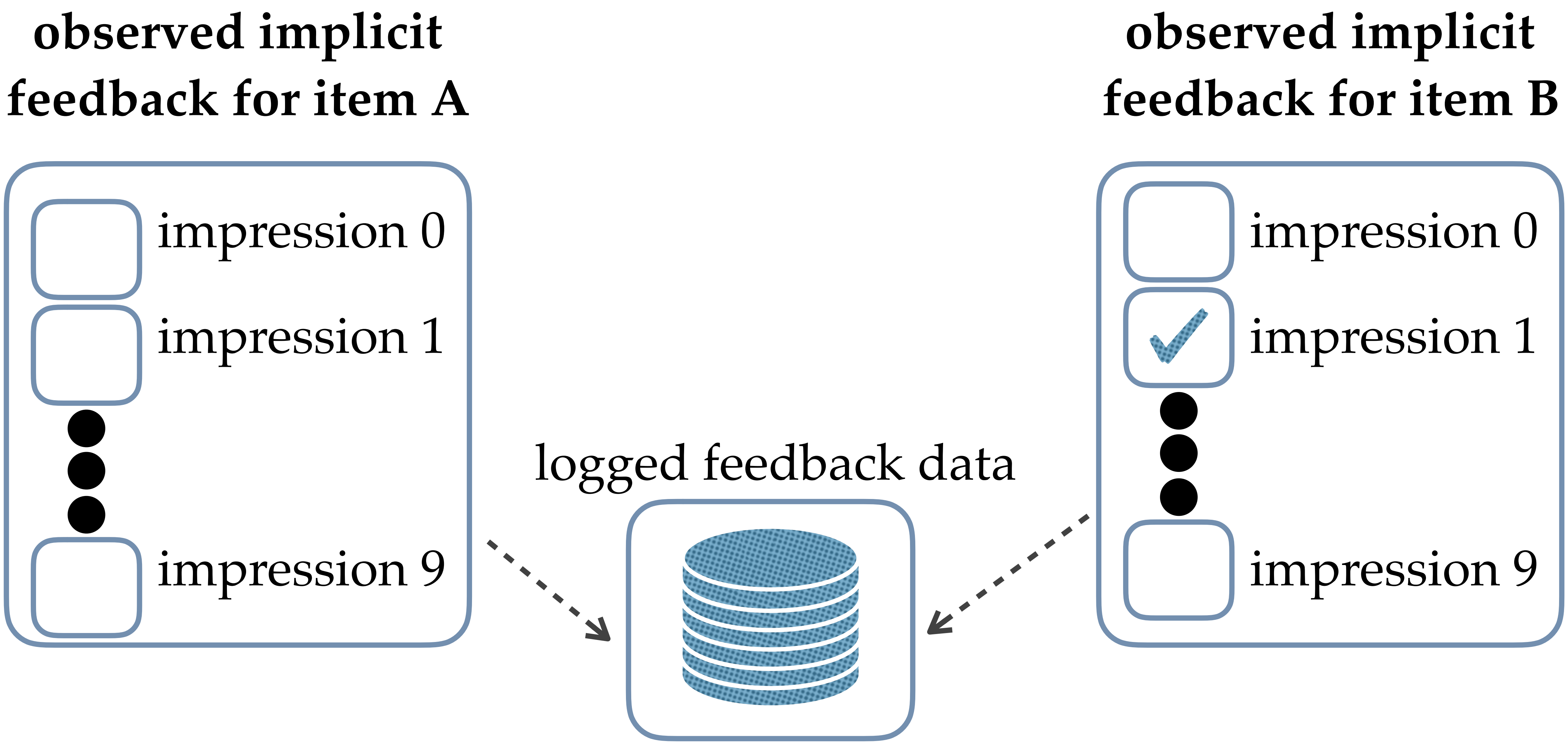
Standard collaborative filtering methods are limited because they can only exploit or ignore

- e.g. two items, A and B, with the same click rate = 0.1



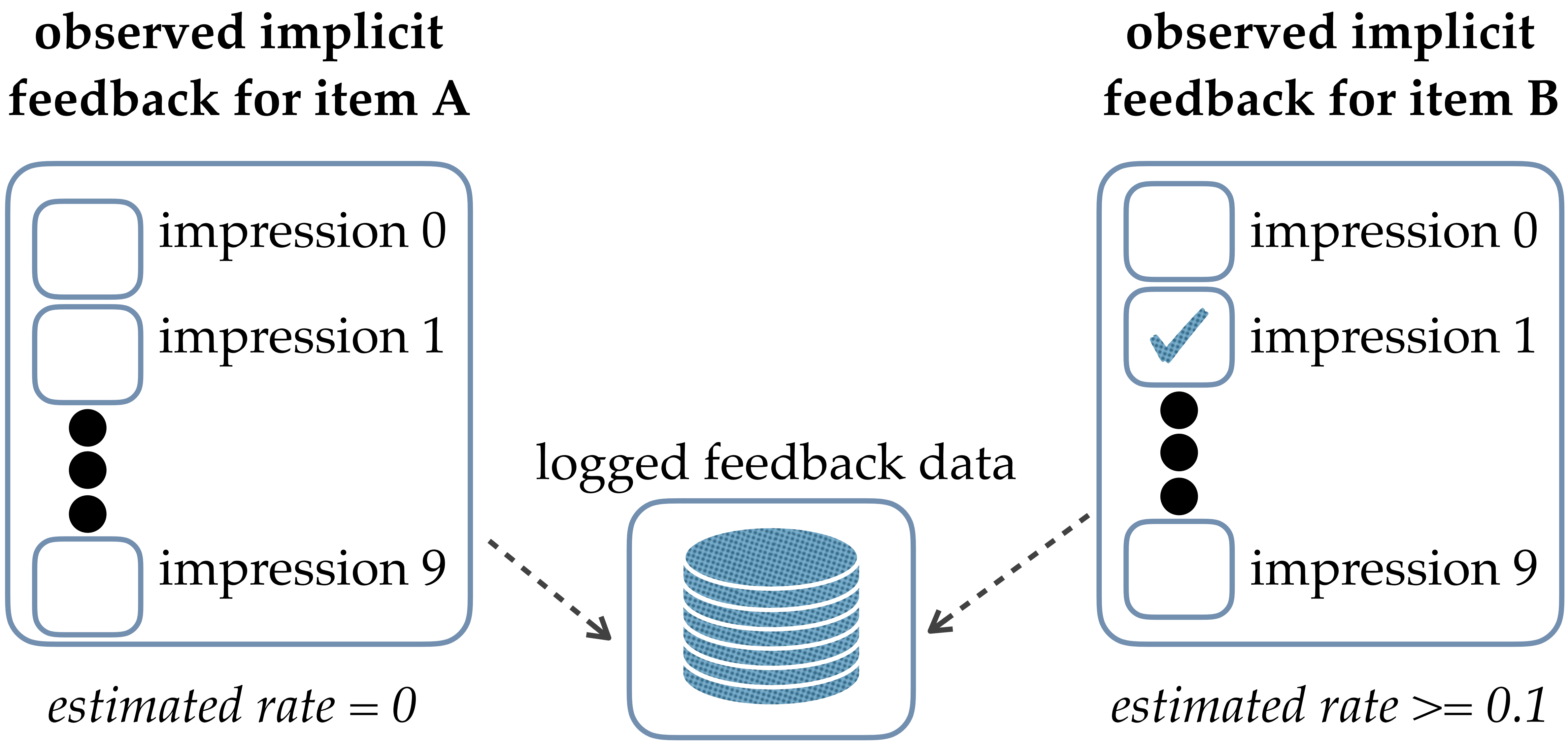
Standard collaborative filtering methods are limited because they can only exploit or ignore

- e.g. two items, A and B, with the same click rate = 0.1



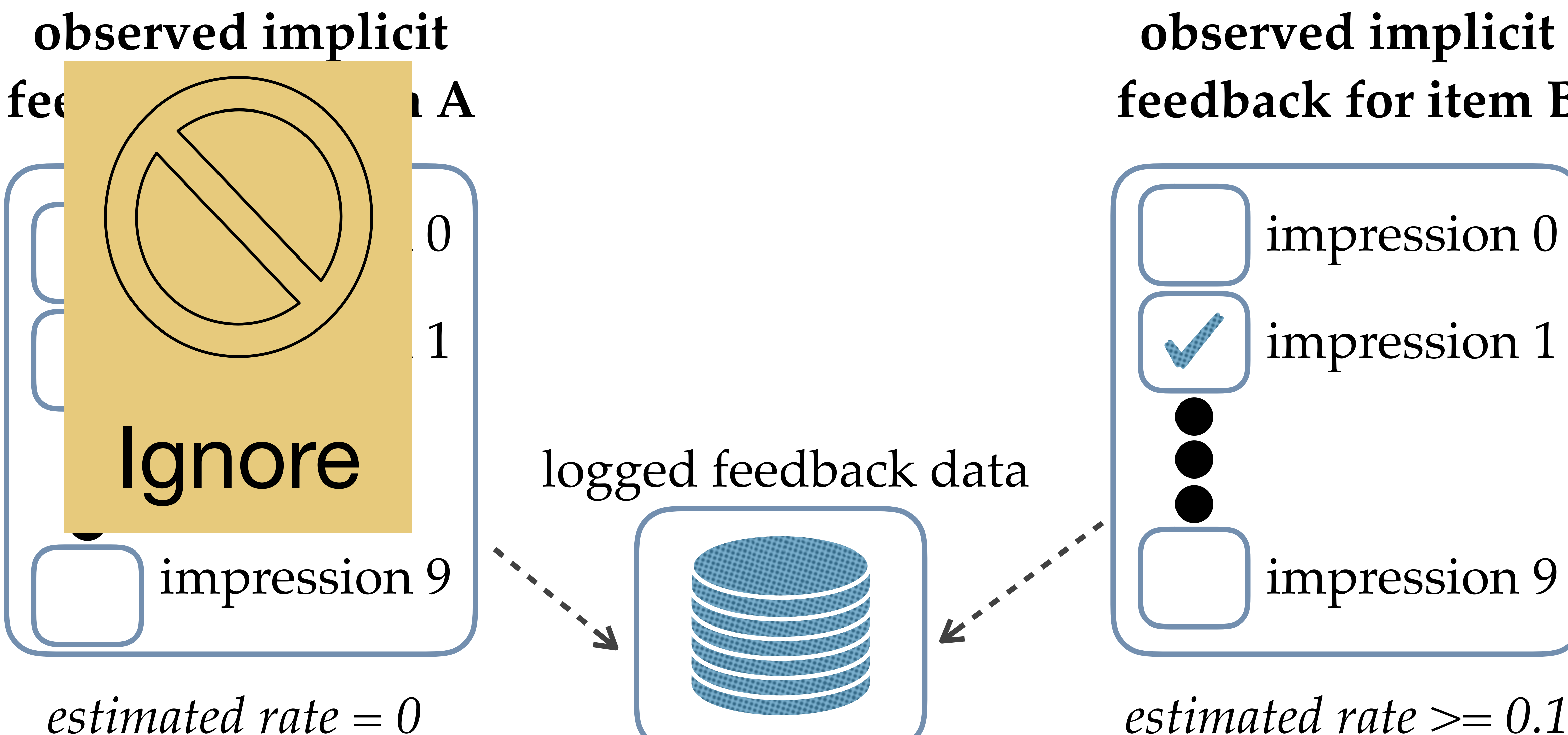
Standard collaborative filtering methods are limited because they can only exploit or ignore

- e.g. two items, A and B, with the same click rate = 0.1



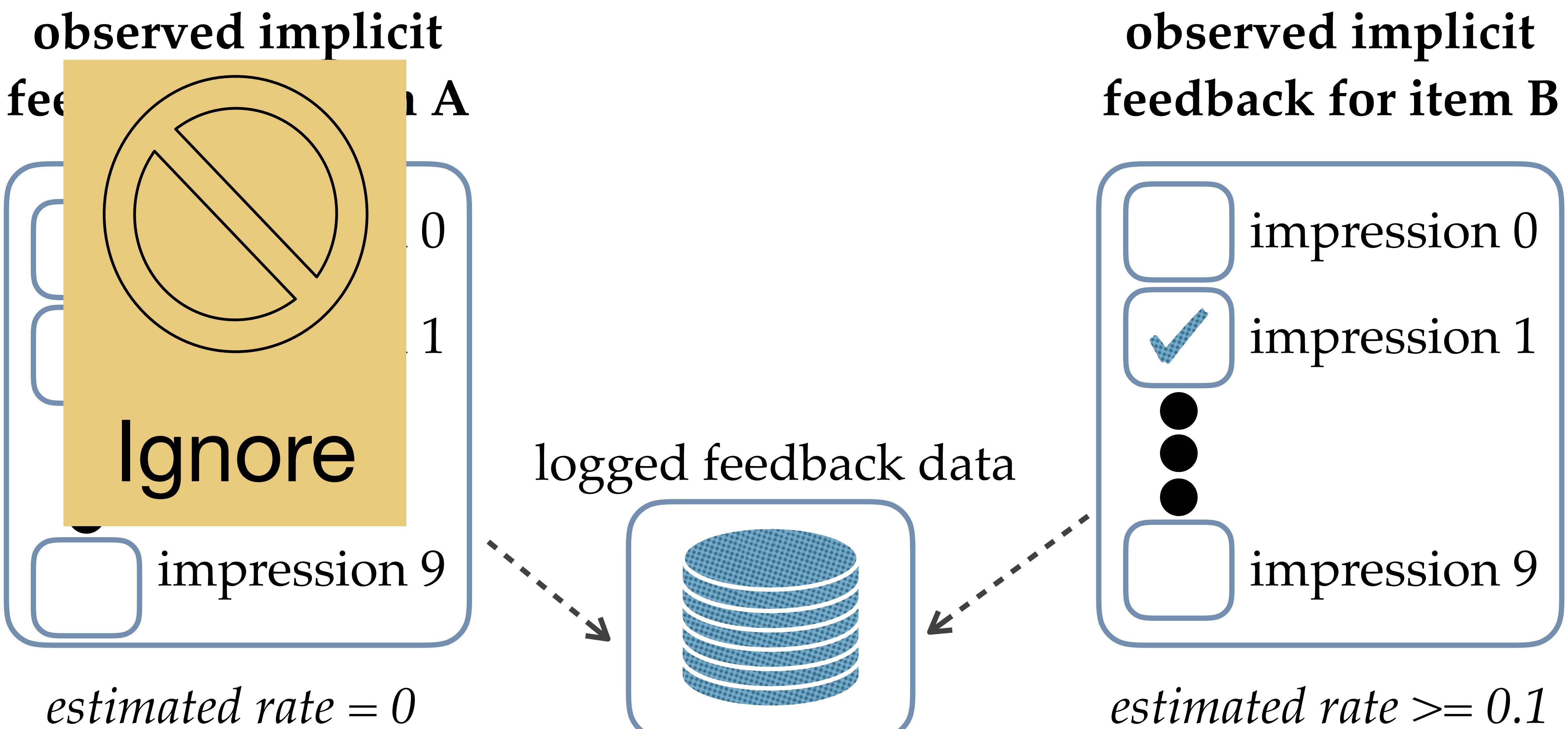
Standard collaborative filtering methods are limited because they can only exploit or ignore

- e.g. two items, A and B, with the same click rate = 0.1



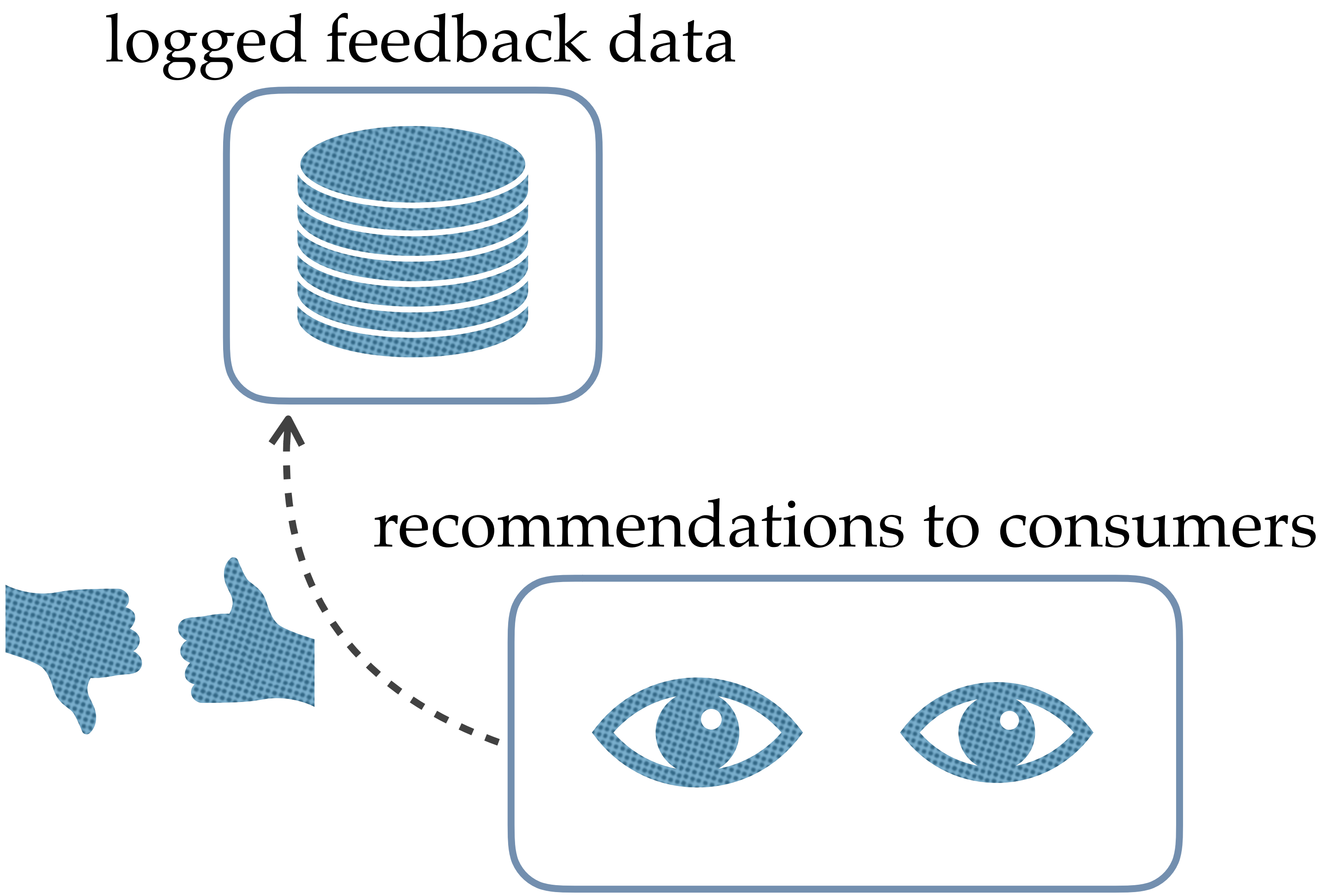
Standard collaborative filtering methods are limited because they can only exploit or ignore

- e.g. two items, A and B, with the same click rate = 0.1

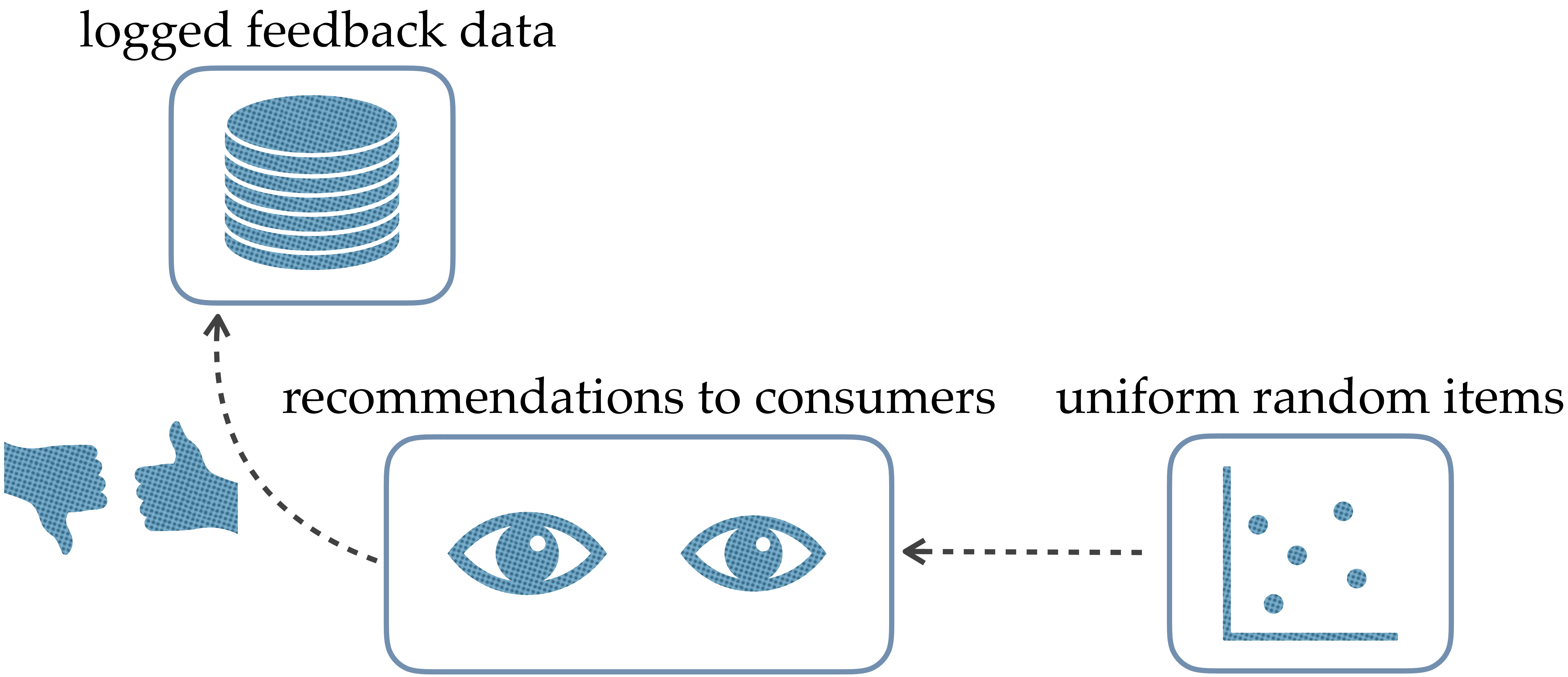


- for this example, this outcome happens 22.7% of the time.

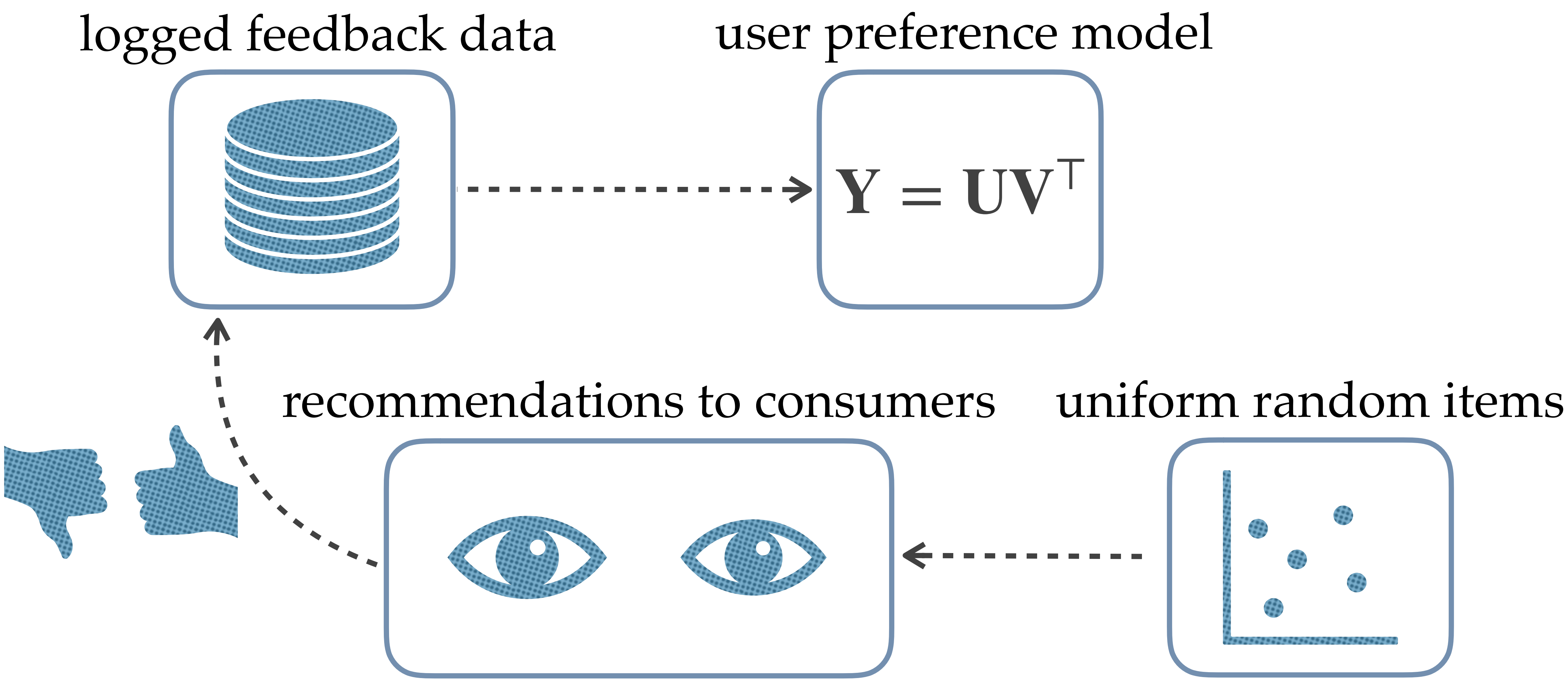
Let's restart from the basic ideal of randomized controlled trials



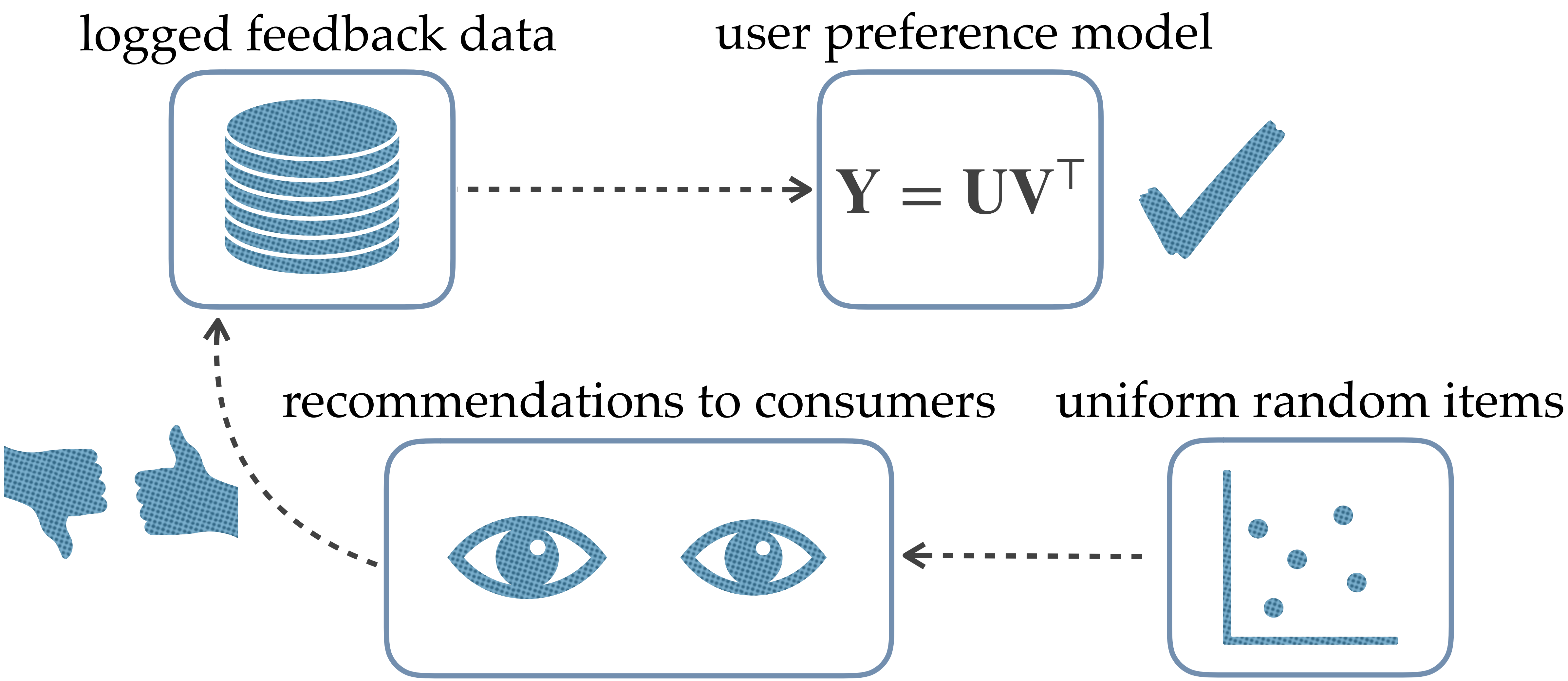
Let's restart from the basic ideal of randomized controlled trials



Let's restart from the basic ideal of randomized controlled trials



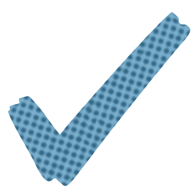
Let's restart from the basic ideal of randomized controlled trials



Let's restart from the basic ideal of randomized controlled trials

$$\checkmark \quad = \quad \mathbb{E}_{X,A\sim\mathrm{Uniform}(\mathscr{A}),Y}[\log p_\theta(Y\,|\,A,X)]$$

Let's restart from the basic ideal of randomized controlled trials



$$= \mathbb{E}_{X, A \sim \mathrm{Uniform}(\mathscr{A}), Y}[\log p_\theta(Y \mid A, X)]$$

random item recommended

set of all items

model parameters

context

Let's restart from the basic ideal of randomized controlled trials

$$\checkmark = \mathbb{E}_{X,A\sim\text{Uniform}(\mathscr{A}),Y}[\log p_\theta(Y\,|\,A,X)]$$

random item recommended

set of all items

model parameters

context

$\arg_\theta \max$ with finite data set is maximum likelihood

Let's restart from the basic ideal of randomized controlled trials

$$\checkmark \;=\; \mathbb{E}_{X,A\sim\mathrm{Uniform}(\mathscr{A}),Y}[\log p_\theta(Y|A,X)]$$

random item recommended

set of all items

model parameters

context

$\arg_\theta \max$ with finite data set is maximum likelihood

- aside: matrix factorization is a special case when the context is the user index vector.

But we don't want to just recommend random stuff all the time ⚡⚡⚡

But we don’t want to just recommend random stuff all the time ⚡⚡⚡

- Enter exploration-exploitation [Sutton & Barto, 1998]

But we don’t want to just recommend random stuff all the time

- Enter exploration-exploitation [Sutton & Barto, 1998]

recommender system relevance certainty

ground truth item relevance

	Low certainty	High certainty
Low relevance	Explore	Ignore
High relevance	Explore	Exploit

But we don't want to just recommend random stuff all the time ⚡⚡⚡

- Enter exploration-exploitation [Sutton & Barto, 1998]

recommender system relevance certainty

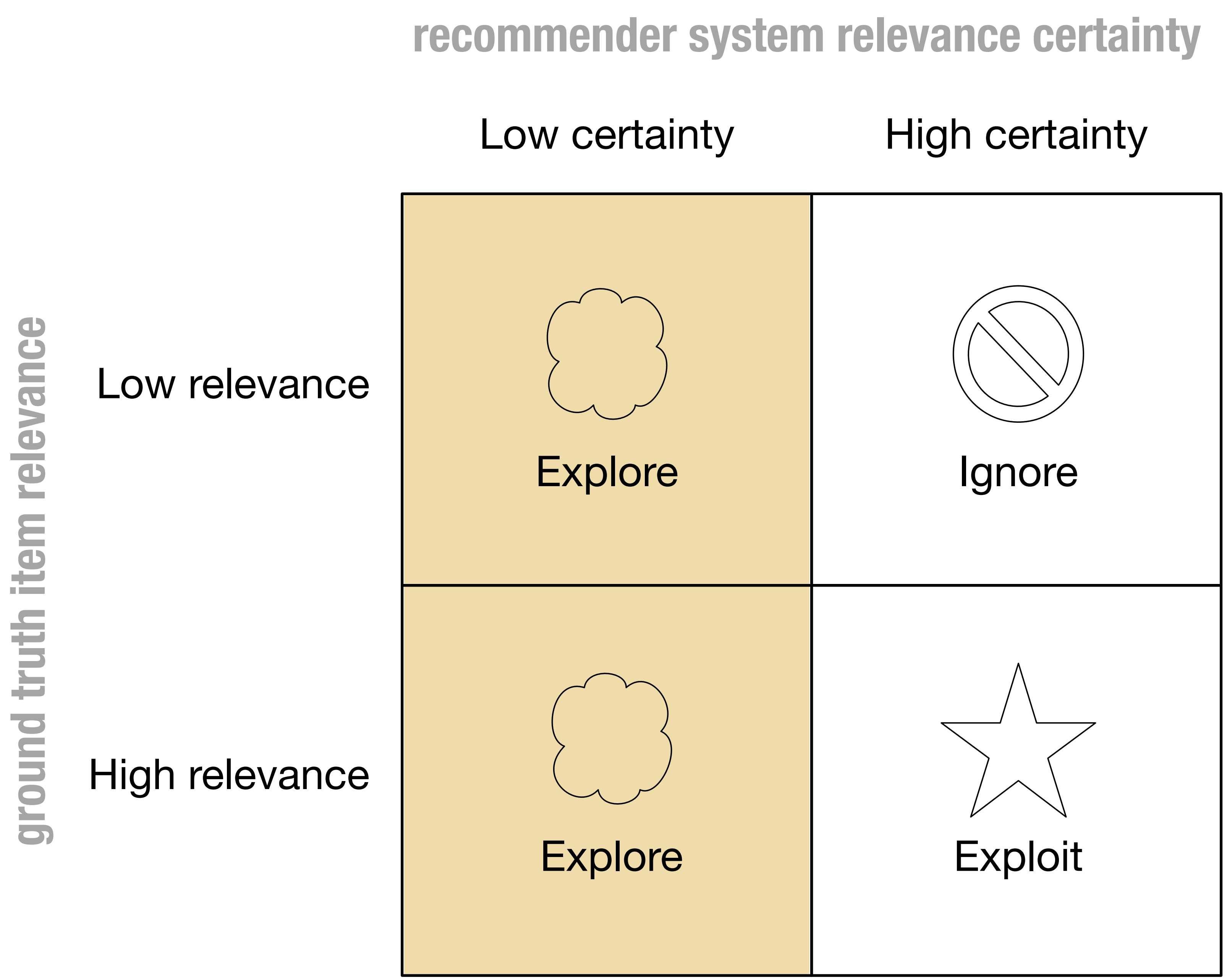
ground truth item relevance

	Low certainty	High certainty
Low relevance	Explore	Ignore
High relevance	Explore	Exploit

- When the recommender is certain it has a bad item, it ignores it.
- When the recommender is certain it has a good item, it recommends it.

But we don’t want to just recommend random stuff all the time

- Enter exploration-exploitation [Sutton & Barto, 1998]



- When the recommender is certain it has a bad item, it ignores it.
- When the recommender is certain it has a good item, it recommends it.

How to balance exploration and exploitation?

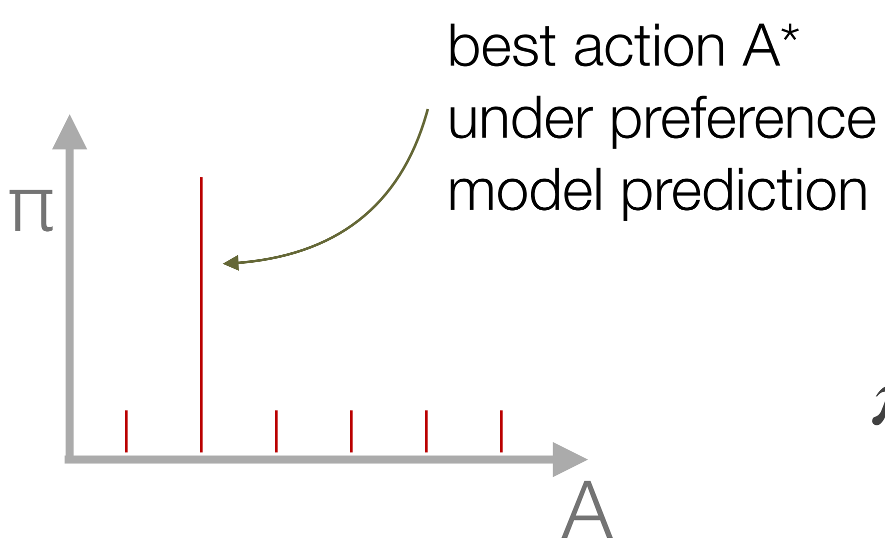
How to balance exploration and exploitation?

- the central question of contextual multi-armed bandits
- standard methods include epsilon-greedy, Thompson sampling, and upper confidence bounds

How to balance exploration and exploitation?

- the central question of contextual multi-armed bandits
- standard methods include epsilon-greedy, Thompson sampling, and upper confidence bounds

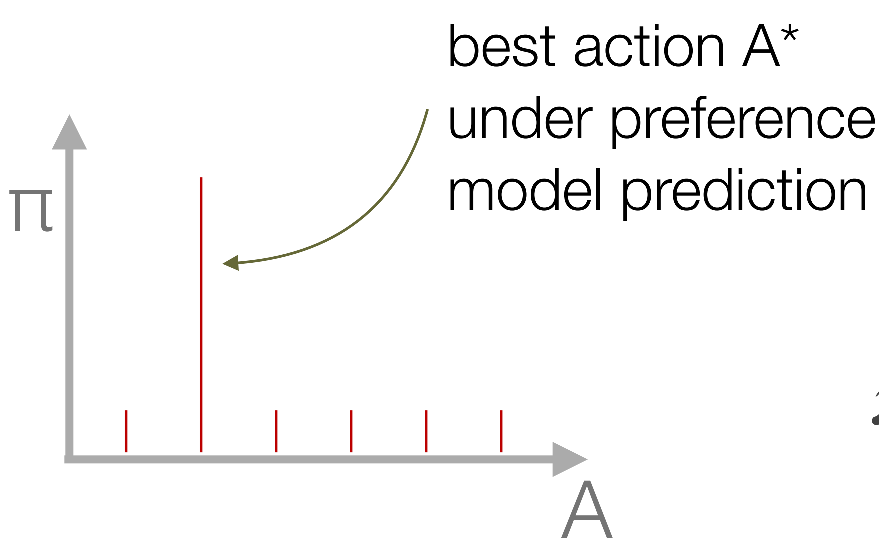
ε-greedy



$$\pi(A \mid X) = \begin{cases} (1-\epsilon) + \frac{\epsilon}{|\mathscr{A}|} & \text{when } A = A^* \\ \frac{\epsilon}{|\mathscr{A}|} & \text{otherwise} \end{cases}$$

How to balance exploration and exploitation?

- the central question of contextual multi-armed bandits
- standard methods include epsilon-greedy, Thompson sampling, and upper confidence bounds

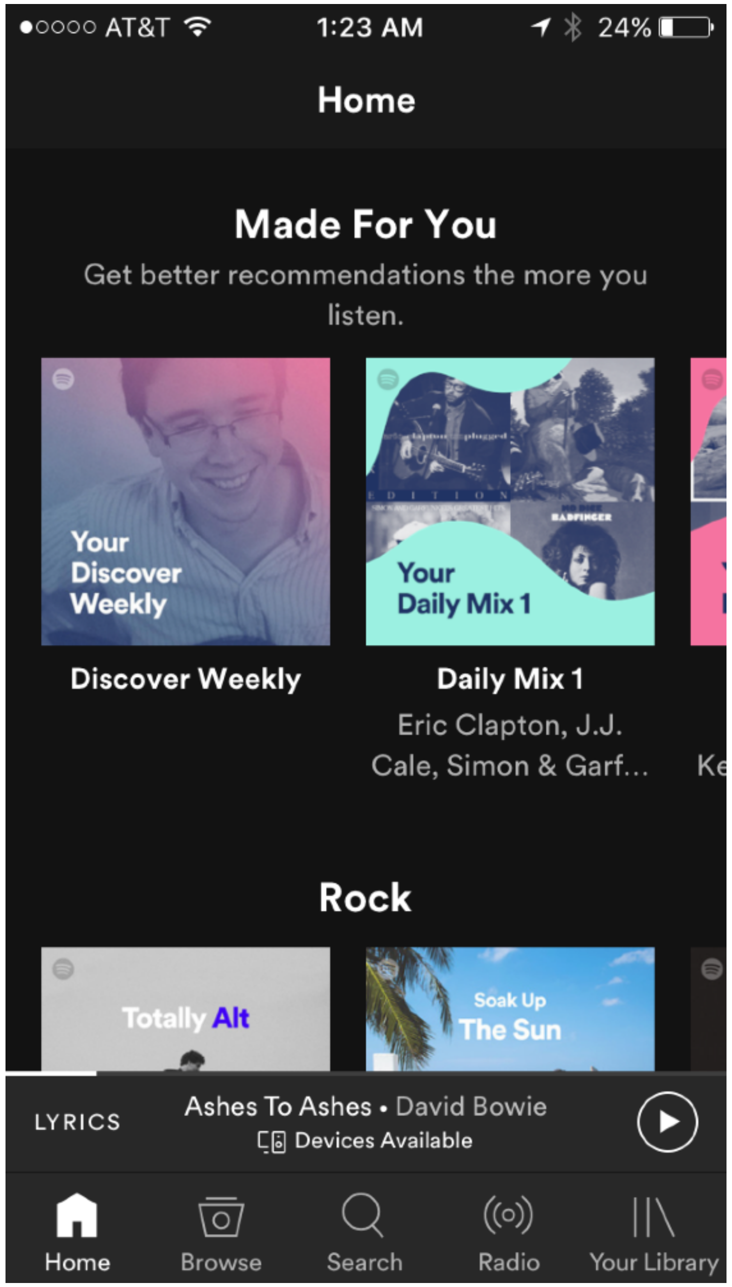


ε-greedy

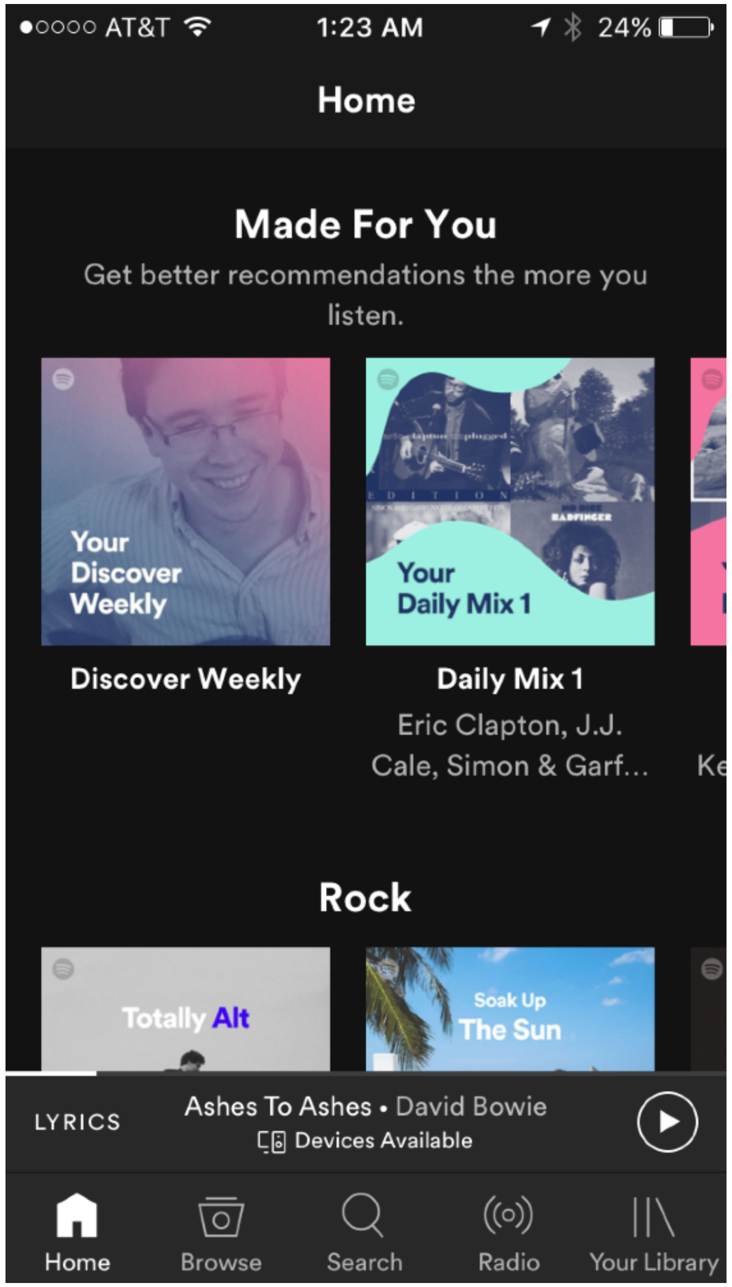
exploration parameter (when fixed -> crude exploitation; can also decay over time)

$$\pi(A \mid X) = \begin{cases} (1-\epsilon) + \dfrac{\epsilon}{|\mathscr{A}|} & \text{when } A = A^* \\ \dfrac{\epsilon}{|\mathscr{A}|} & \text{otherwise} \end{cases}$$

Research question: how to explore-exploit over explainable recommendations?



Research question: how to explore-exploit over explainable recommendations?



- e.g. home page of Spotify, YouTube, or Netflix
- items arranged into shelves, each shelf has a title or <u>explanation</u> for the associated recommendation
- naively, the bandit has to try every possible combination of item and explanation many times before being able to exploit the best combinations

Bart

- Bart (bandits for recommendations as treatments) consists of:
  - a user preference model conditioned on the context
  - a ranking procedure
  - a training procedure

Bart

- Bart (bandits for recommendations as treatments) consists of:
  - a user preference model conditioned on the context
  - a ranking procedure
  - a training procedure

factorization machine capturing interactions between features in a parameter efficient manner [Rendle, 2010]

Bart

- Bart (bandits for recommendations as treatments) consists of:
  - a user preference model conditioned on the context
  - a ranking procedure
  - a training procedure

anything we know about the user and item, including region, age group, recent listening patterns, time of day

factorization machine capturing interactions between features in a parameter efficient manner [Rendle, 2010]

Bart

- Bart (bandits for recommendations as treatments) consists of:
  - a user preference model conditioned on the context
  - a ranking procedure
  - a training procedure

counterfactual maximum likelihood [Joachims & Swaminathan, 2016]

anything we know about the user and item, including region, age group, recent listening patterns, time of day

factorization machine capturing interactions between features in a parameter efficient manner [Rendle, 2010]

Bart

- Bart (bandits for recommendations as treatments) consists of:
  - a user preference model conditioned on the context
  - a ranking procedure
  - a training procedure

anything we know about the user and item, including region, age group, recent listening patterns, time of day

counterfactual maximum likelihood
[Joachims & Swaminathan, 2016]

factorization machine capturing interactions between features in a parameter efficient manner [Rendle, 2010]

(details in a publication under conference review; soon to be on arXiv)

Bart

- Bart (bandits for recommendations as treatments) consists of:
  - a user preference model conditioned on the context
  - a ranking procedure
  - a training procedure

anything we know about the user and item, including region, age group, recent listening patterns, time of day

factorization machine capturing interactions between features in a parameter efficient manner [Rendle, 2010]

counterfactual maximum likelihood
[Joachims & Swaminathan, 2016]

(details in a publication under conference review; soon to be on arXiv)

Animation of ranking procedure

Assumptions of shelf browsing model

Horizontal scrolling

Animation of ranking procedure

Assumptions of shelf browsing model

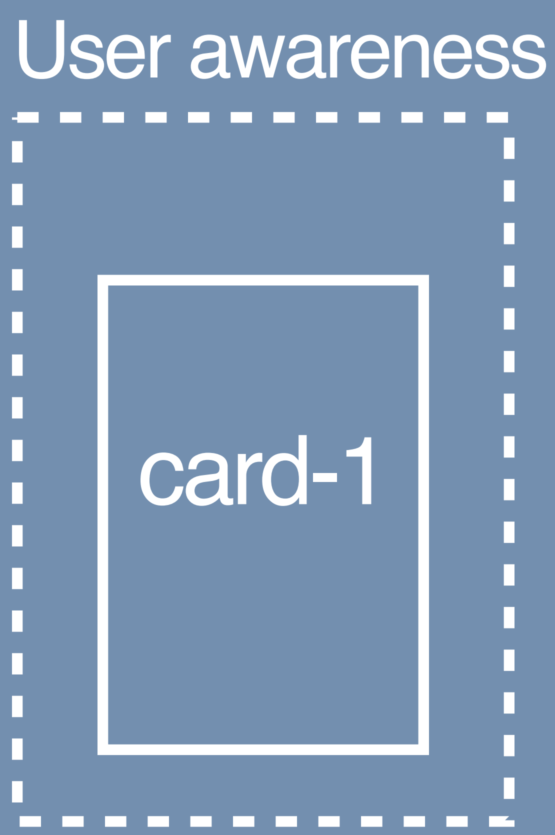
Horizontal scrolling

User awareness

Animation of ranking procedure

Assumptions of shelf browsing model

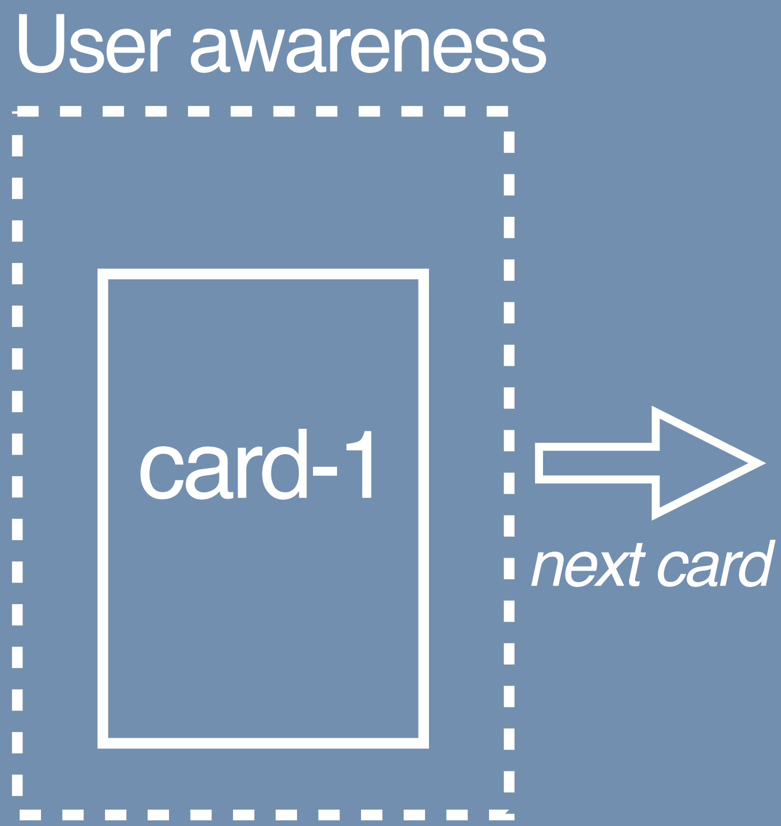
Horizontal scrolling



Animation of ranking procedure

Assumptions of shelf browsing model

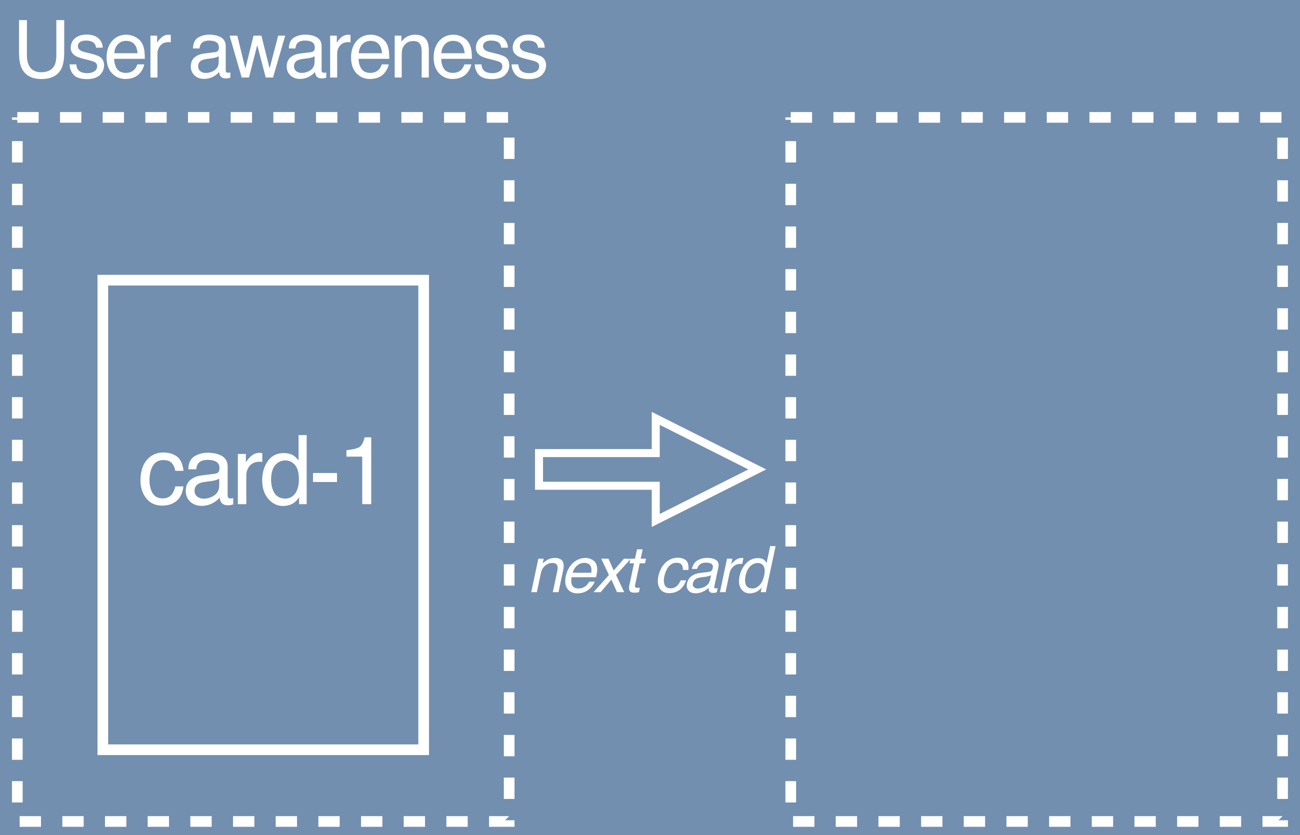
Horizontal scrolling



Animation of ranking procedure

Assumptions of shelf browsing model

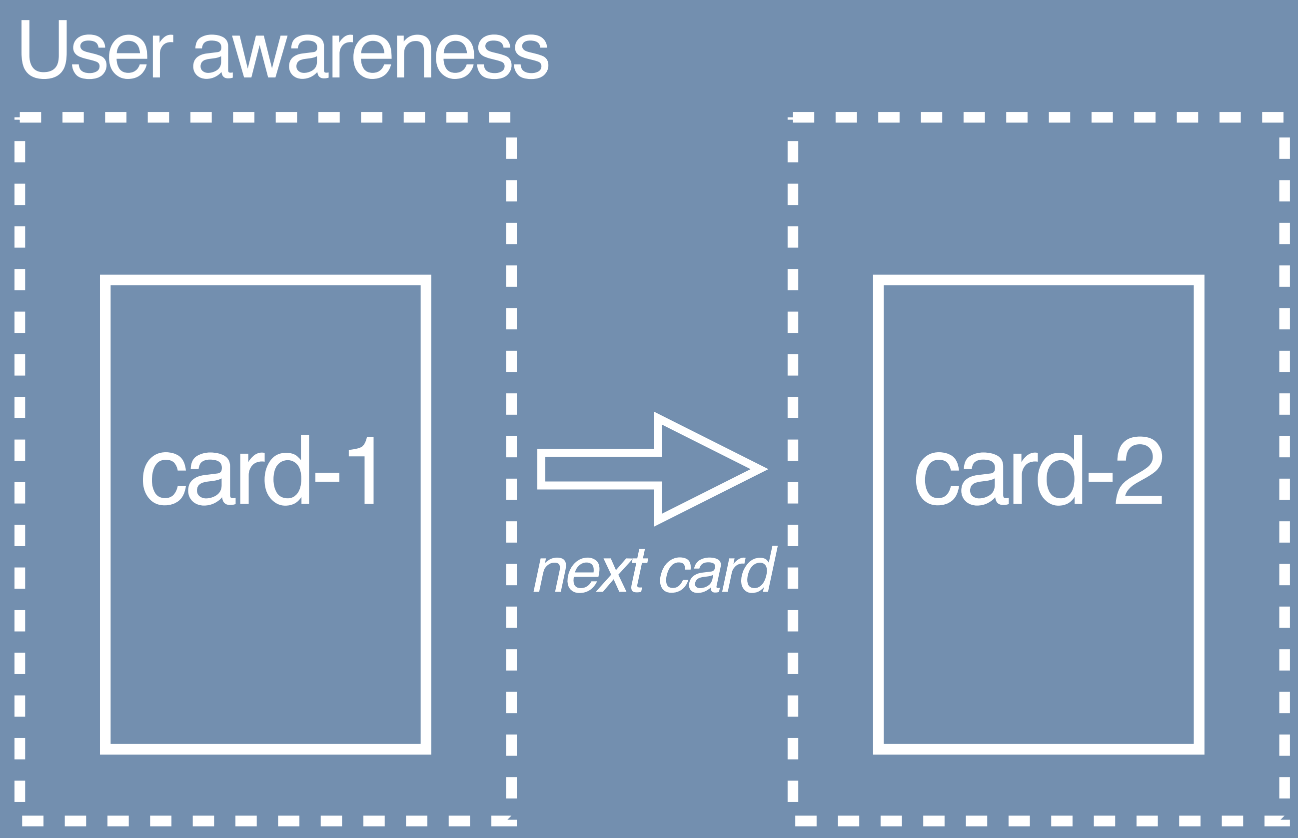
Horizontal scrolling



Animation of ranking procedure

Assumptions of shelf browsing model

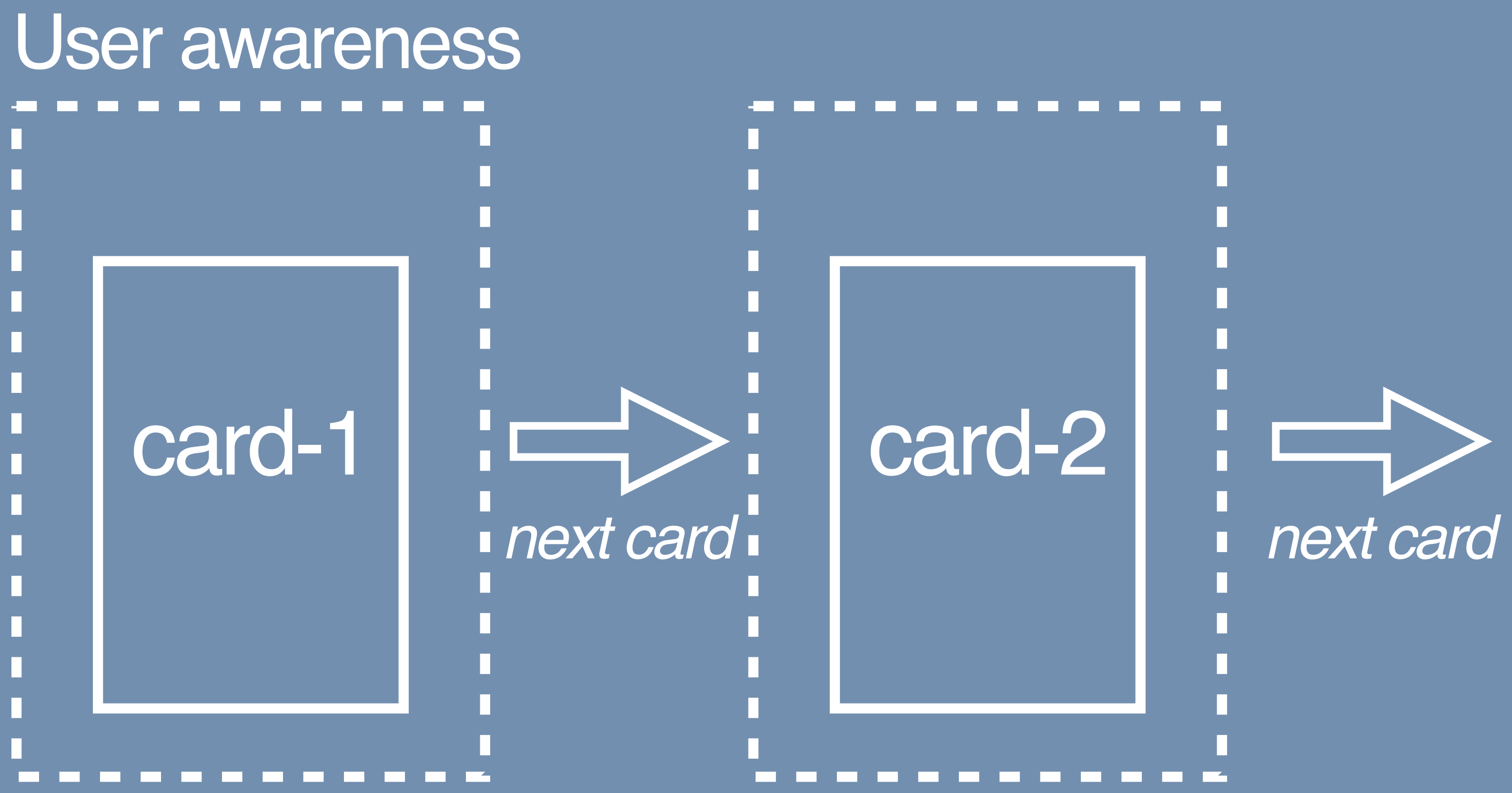
Horizontal scrolling



Animation of ranking procedure

Assumptions of shelf browsing model

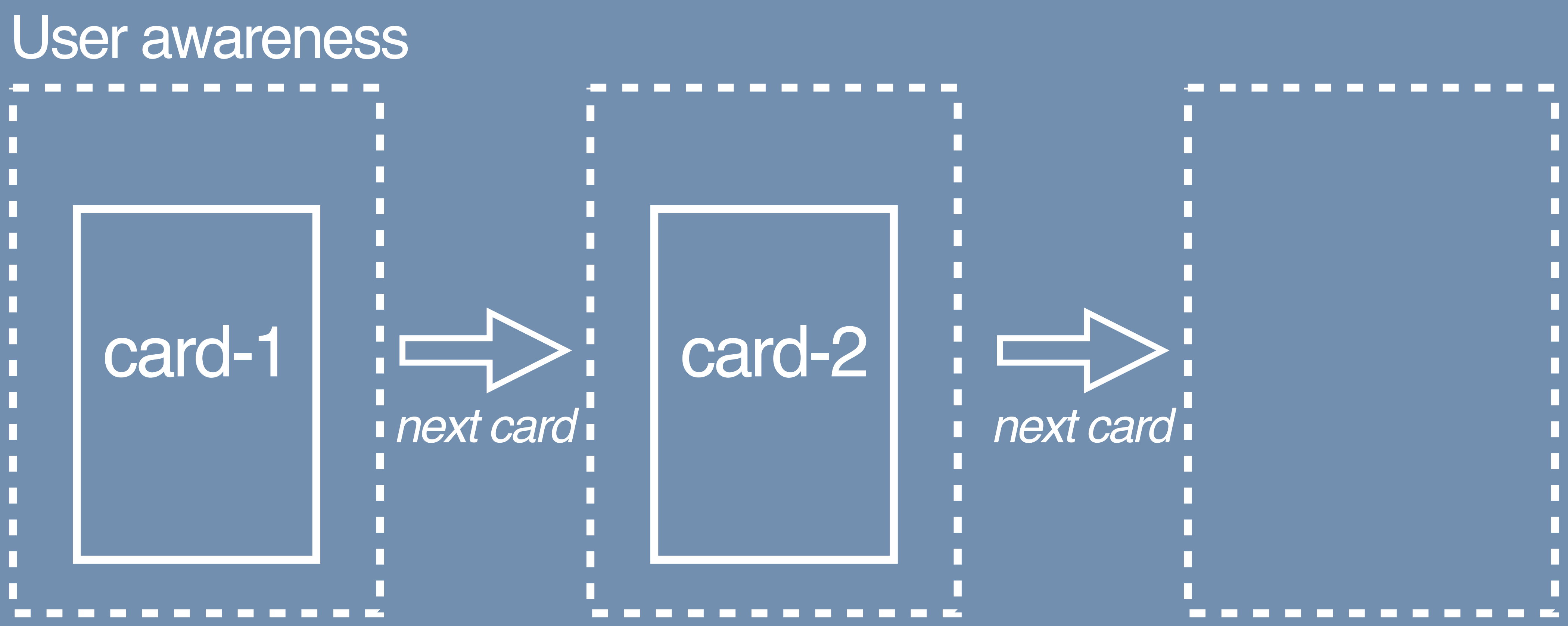
Horizontal scrolling



Animation of ranking procedure

Assumptions of shelf browsing model

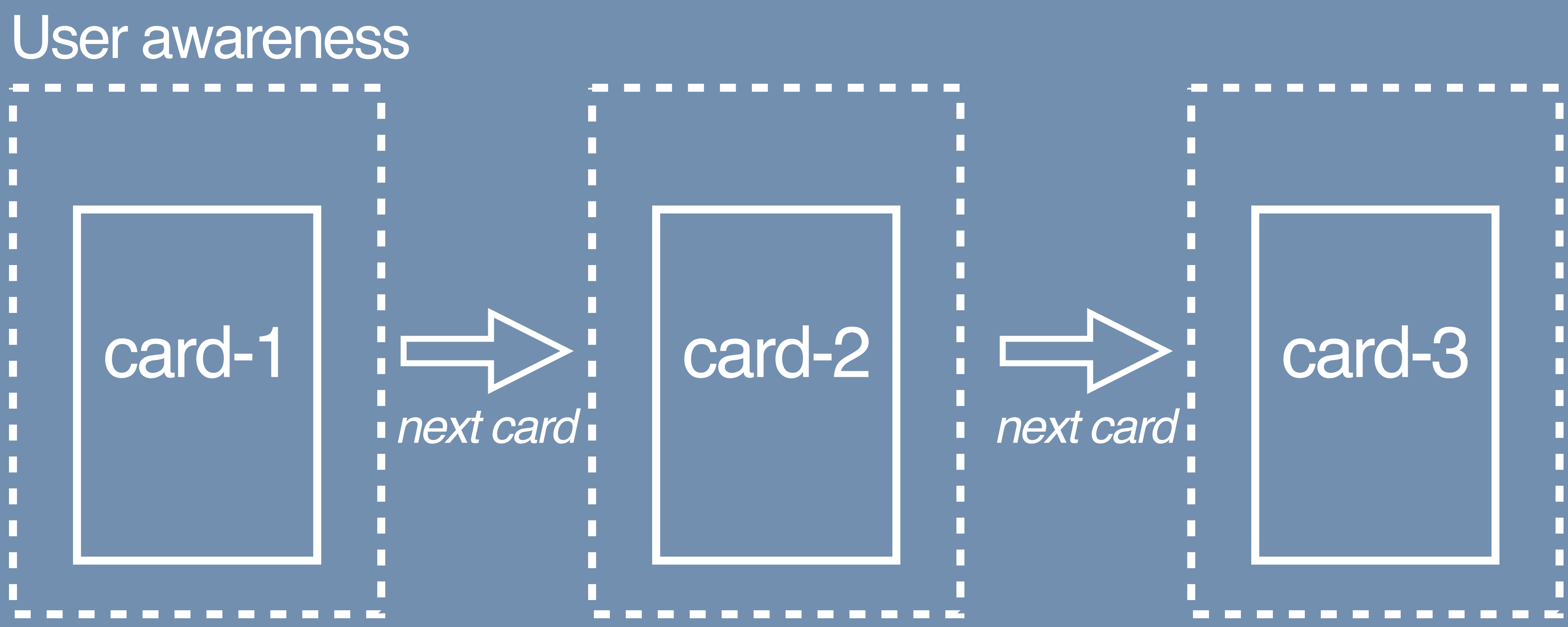
Horizontal scrolling



Animation of ranking procedure

Assumptions of shelf browsing model

Horizontal scrolling



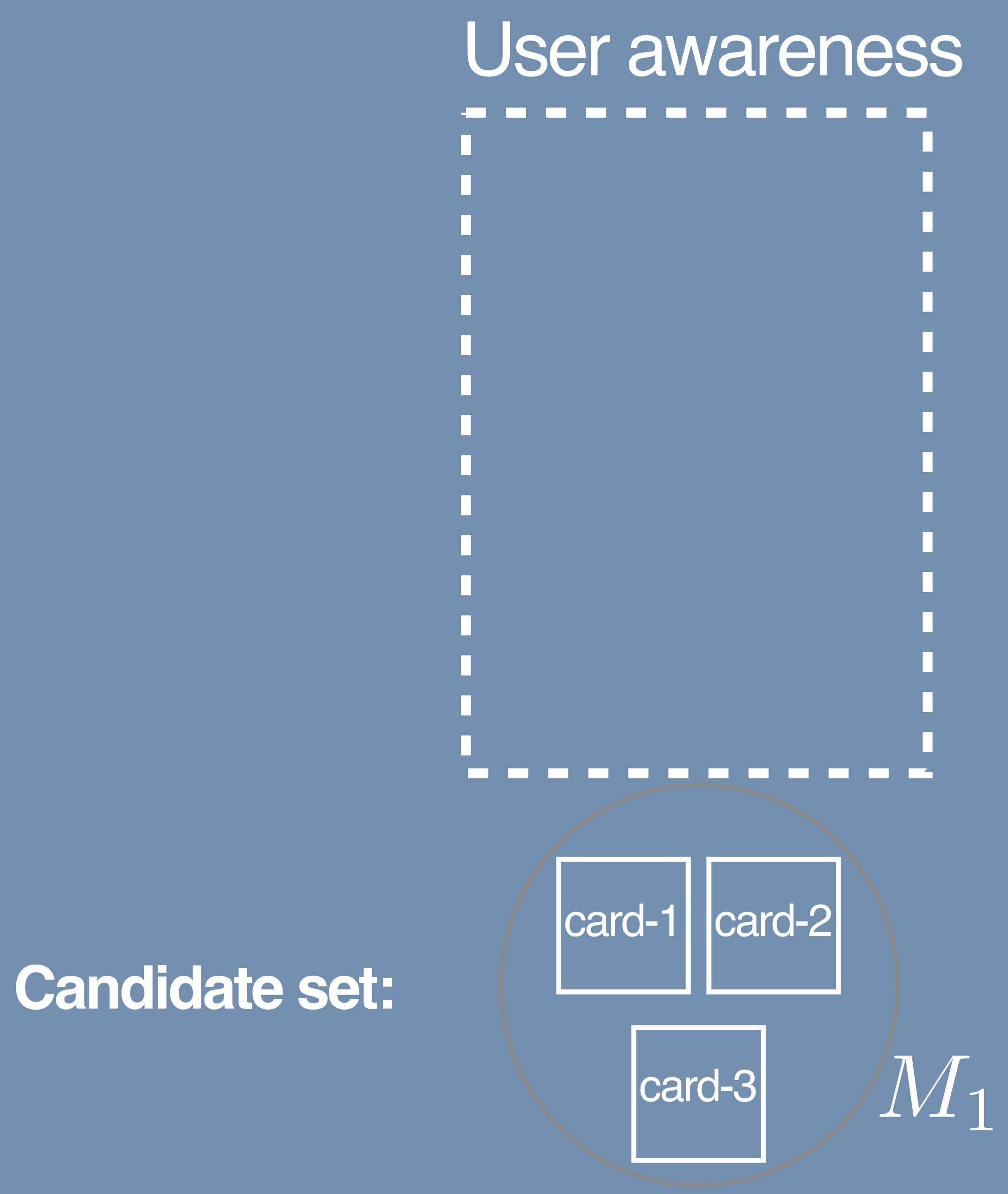
Animation of ranking procedure with bandit

Horizontal scrolling

User awareness

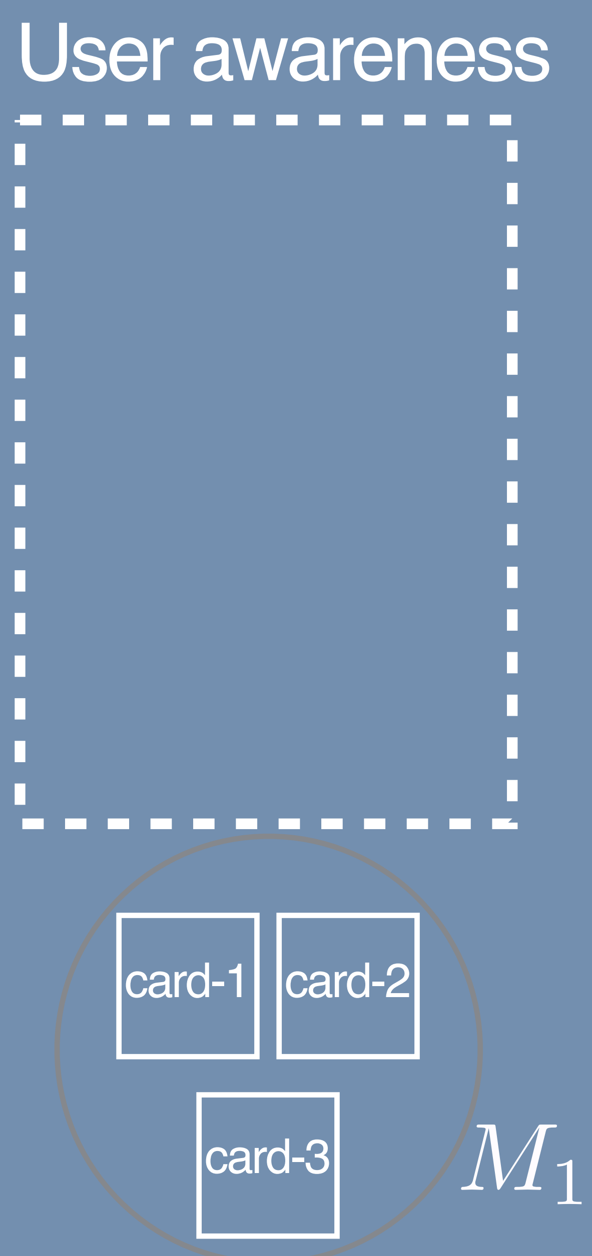
Animation of ranking procedure with bandit

Horizontal scrolling



Animation of ranking procedure with bandit

Horizontal scrolling

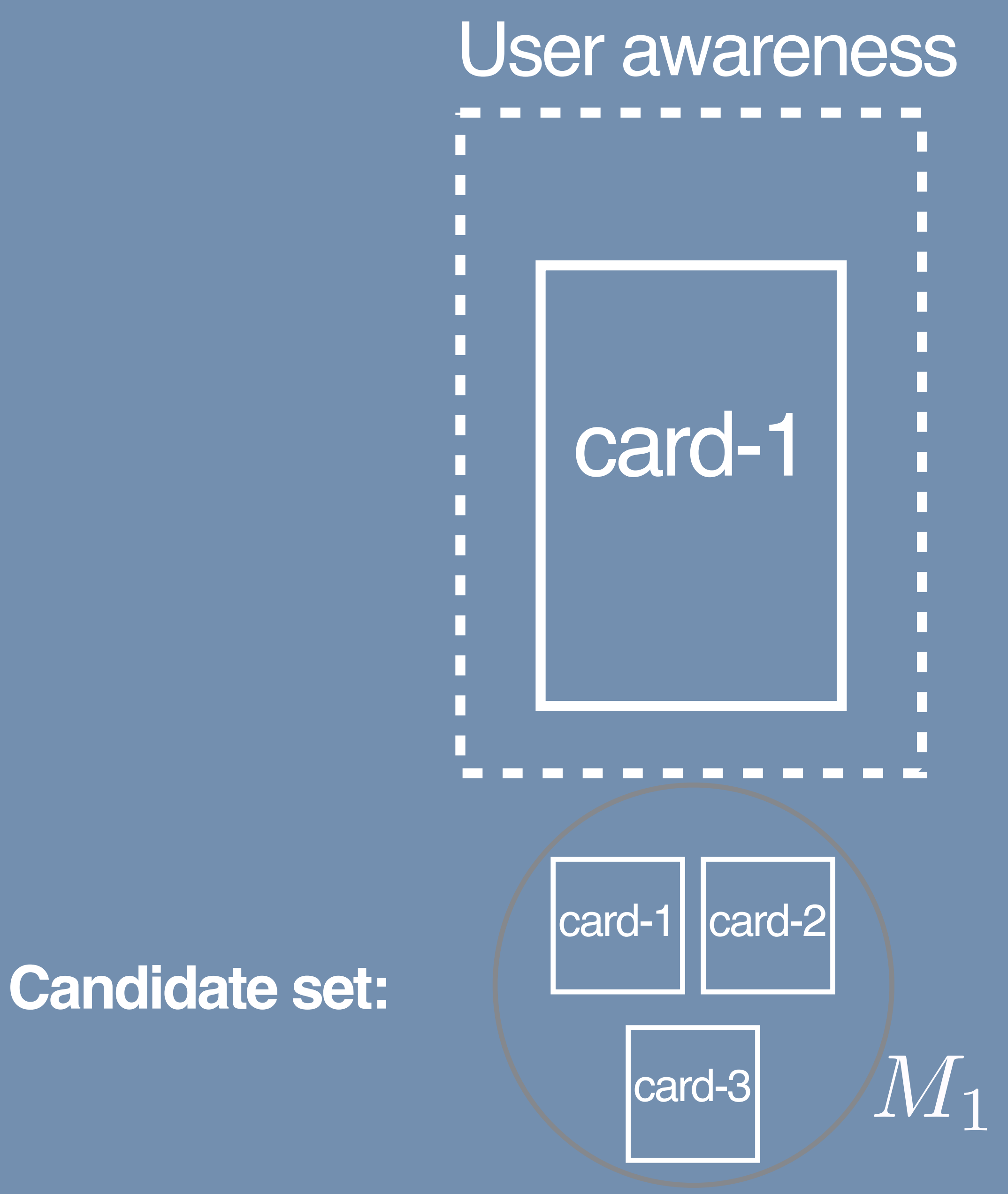


Candidate set:

Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$

Animation of ranking procedure with bandit

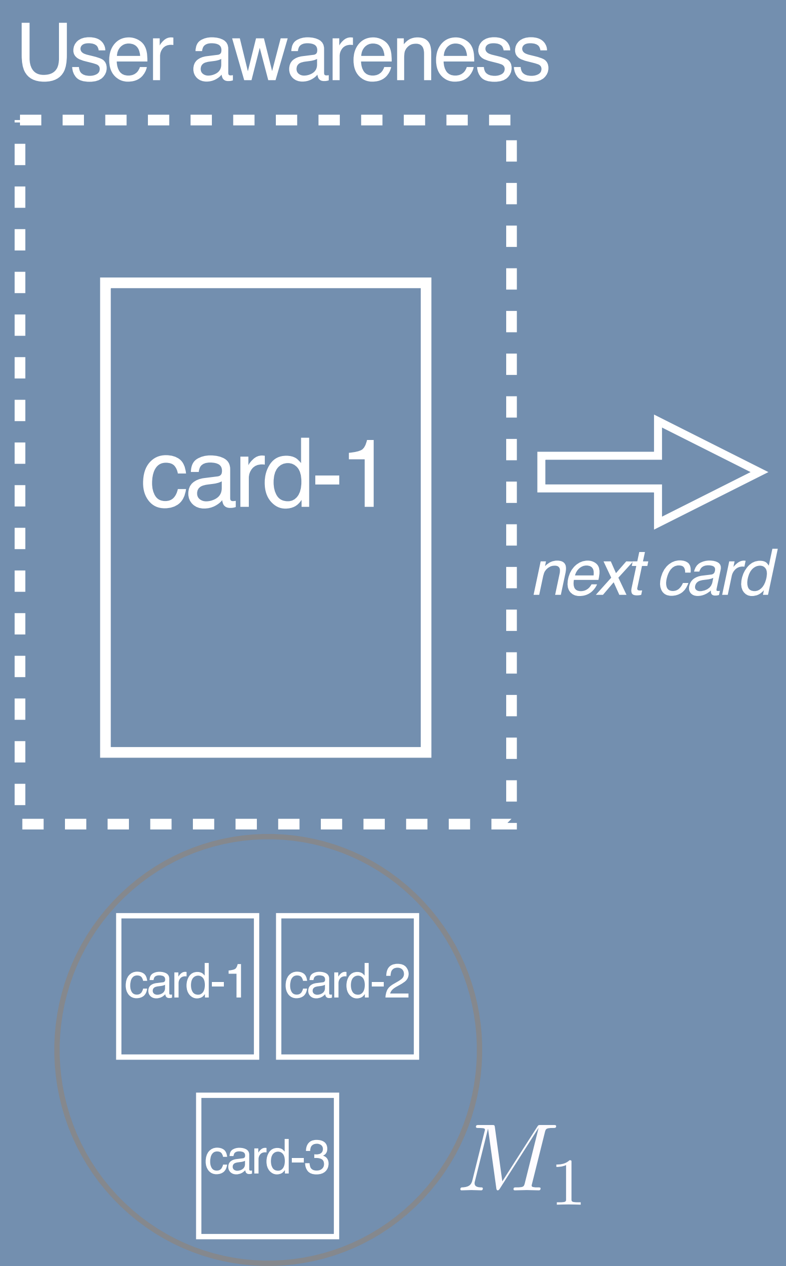
Horizontal scrolling



Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$

Animation of ranking procedure with bandit

Horizontal scrolling

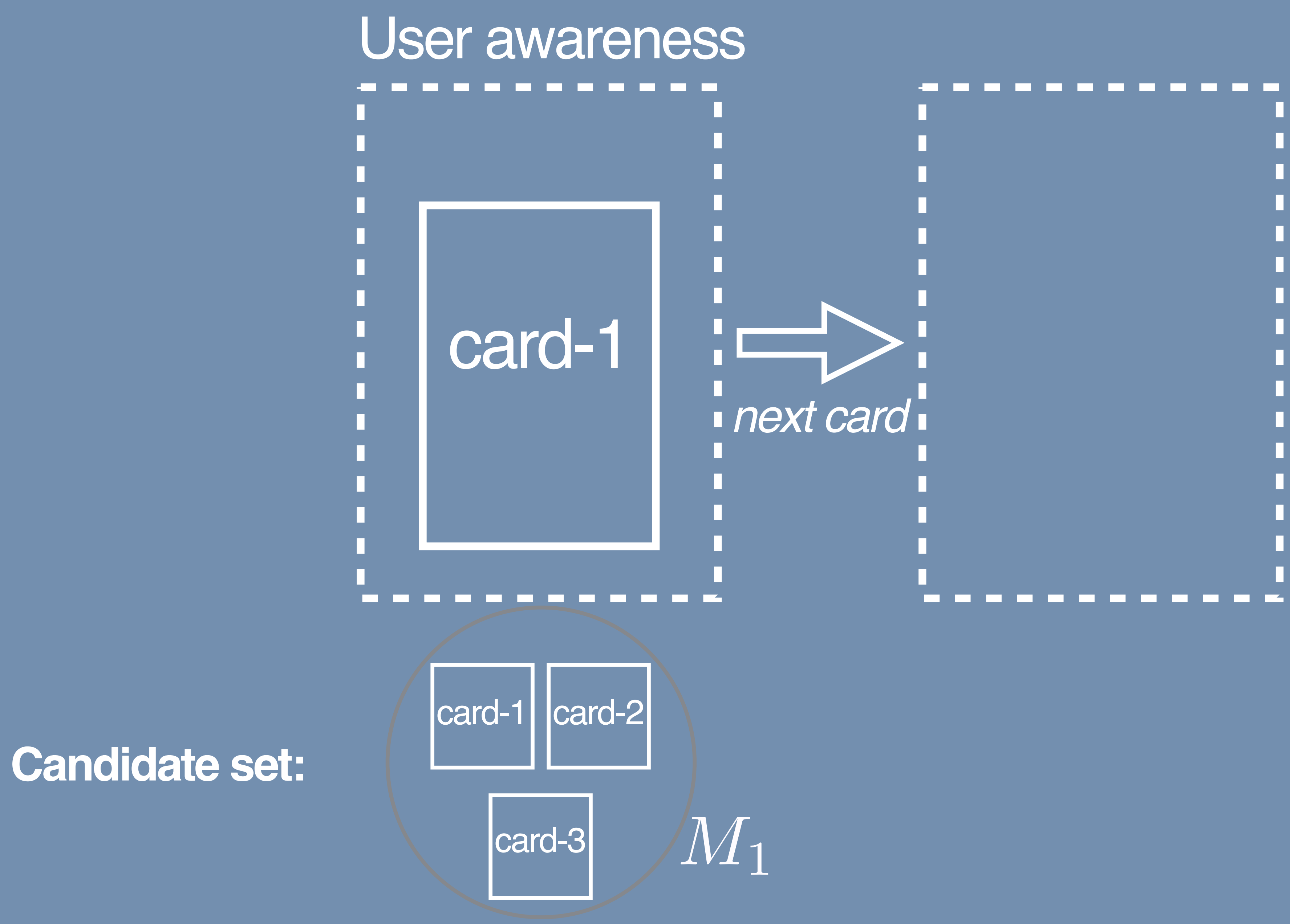


Candidate set:

Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$

Animation of ranking procedure with bandit

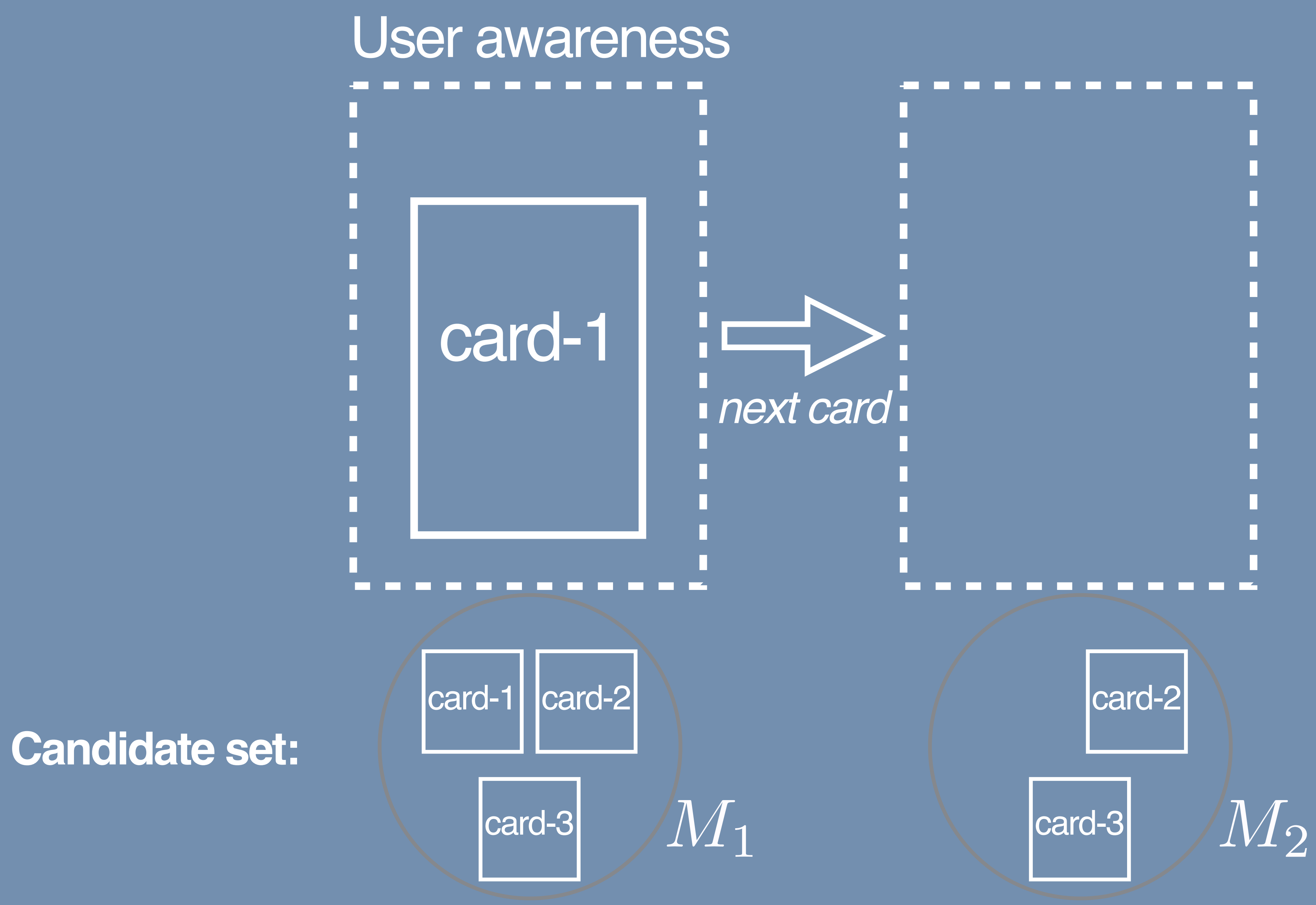
Horizontal scrolling



Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$

Animation of ranking procedure with bandit

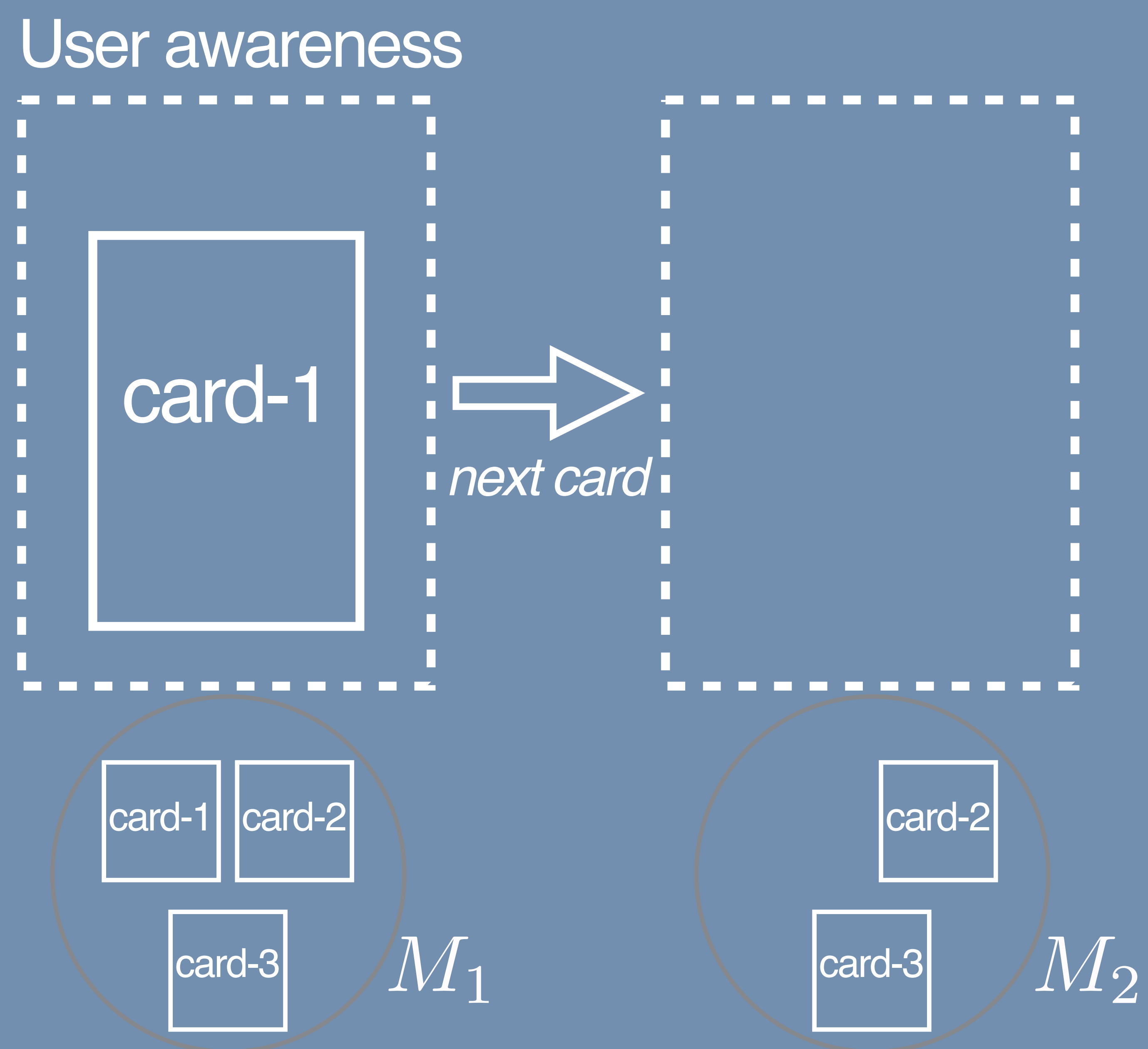
Horizontal scrolling



Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$

Animation of ranking procedure with bandit

Horizontal scrolling

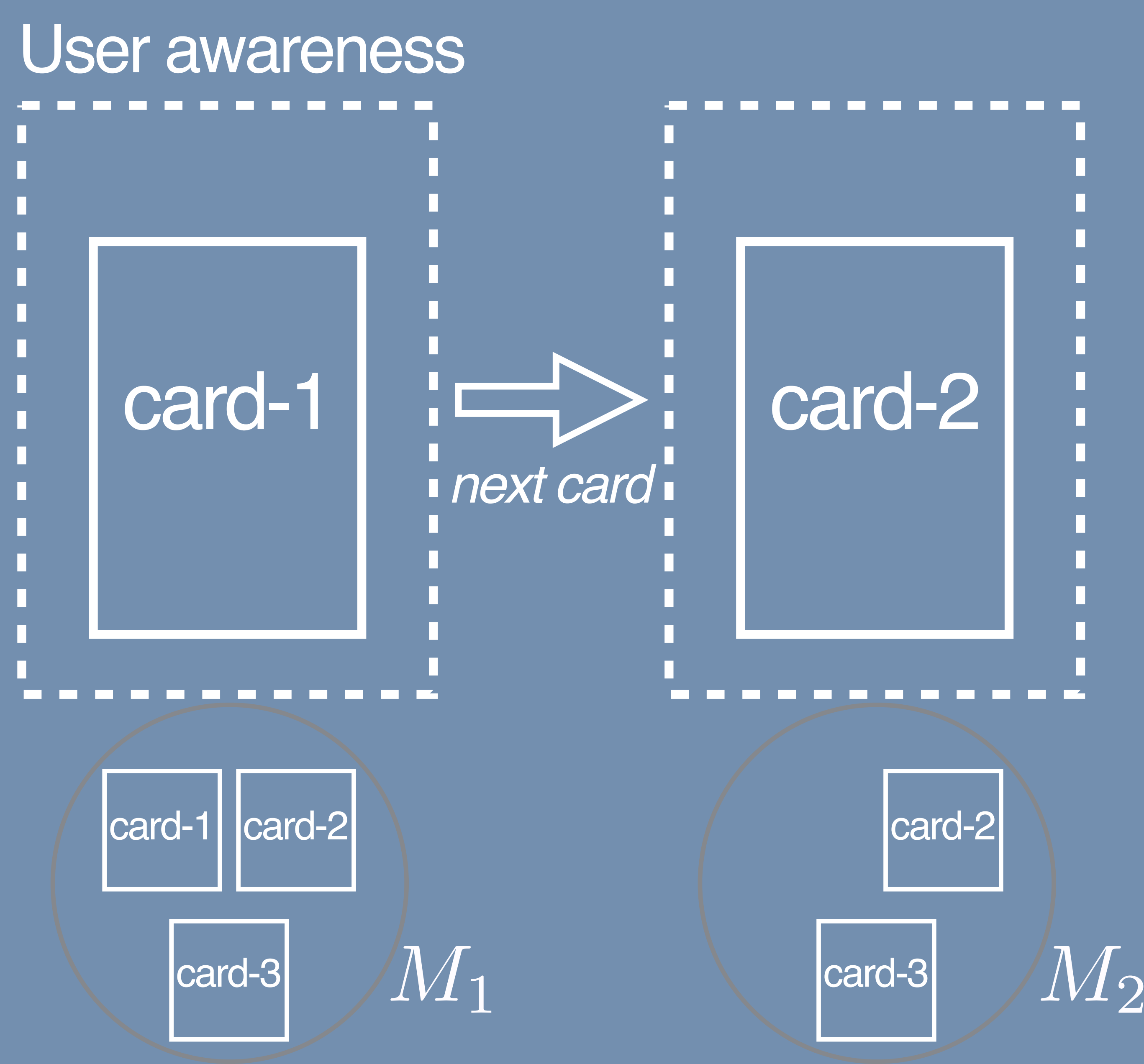


Candidate set:

Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$ $\mathrm{card}_2 \sim \pi_{s,r}(M_2)$

Animation of ranking procedure with bandit

Horizontal scrolling

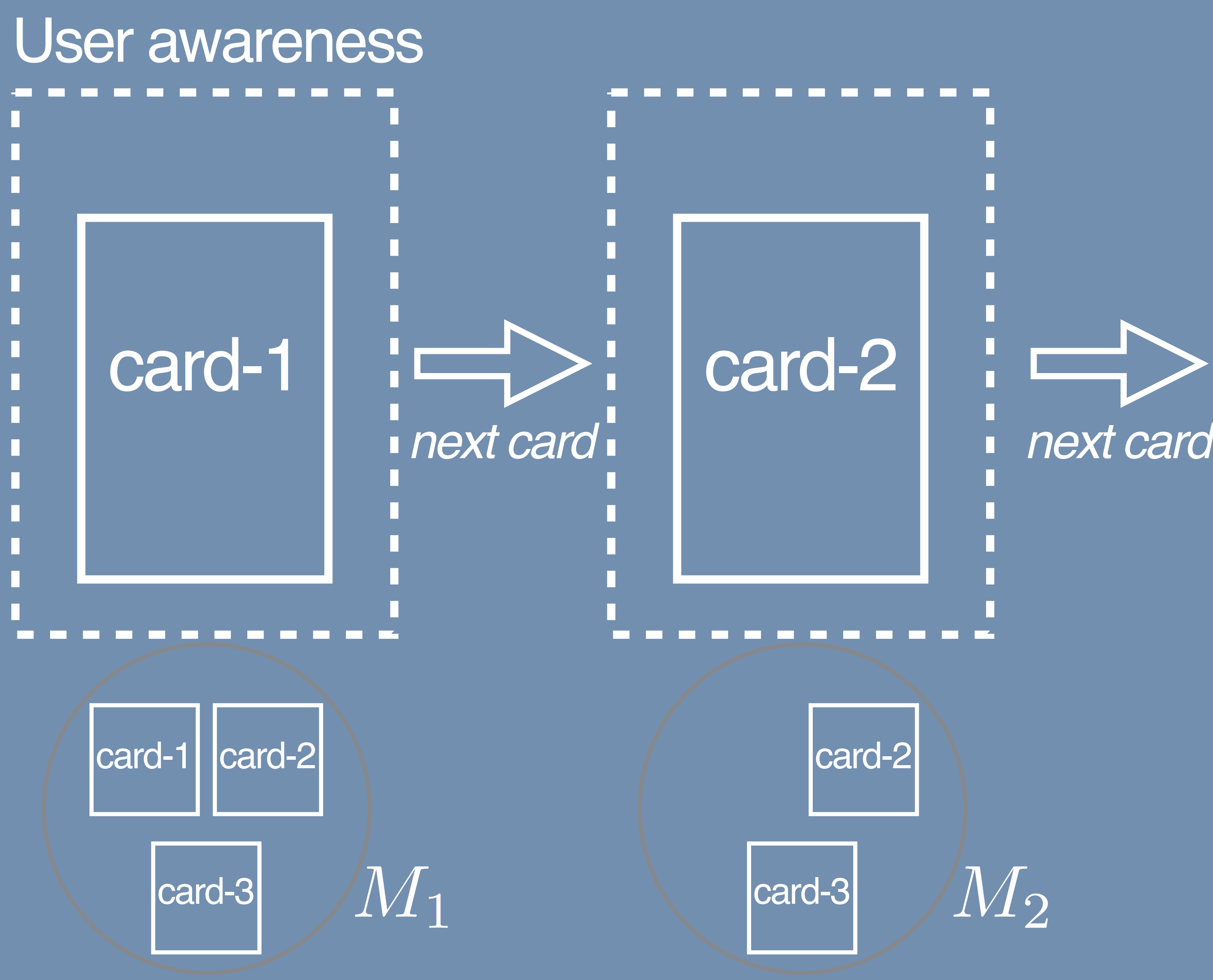


Candidate set:

Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$ $\mathrm{card}_2 \sim \pi_{s,r}(M_2)$

Animation of ranking procedure with bandit

Horizontal scrolling

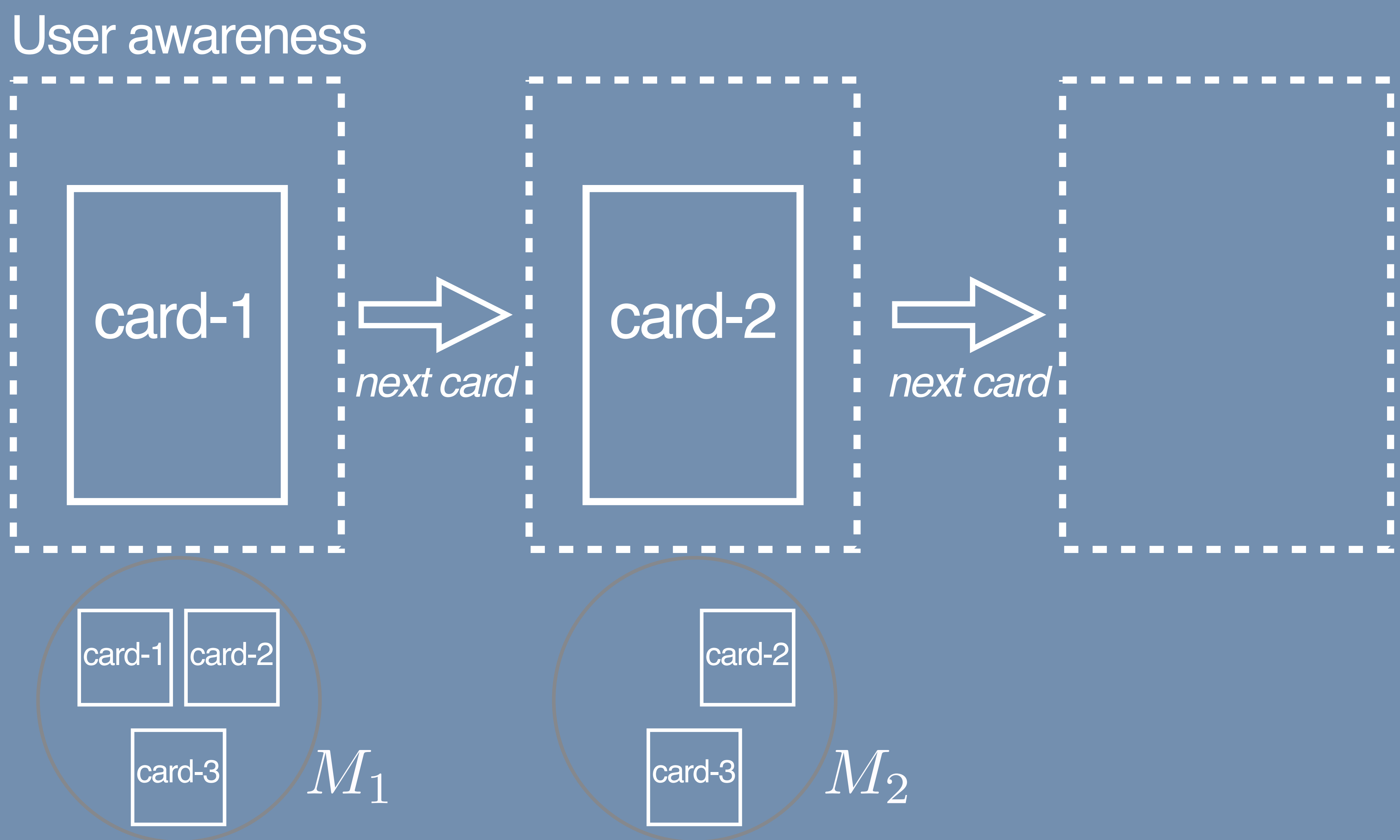


Candidate set:

Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$ $\quad$ $\mathrm{card}_2 \sim \pi_{s,r}(M_2)$

Animation of ranking procedure with bandit

Horizontal scrolling

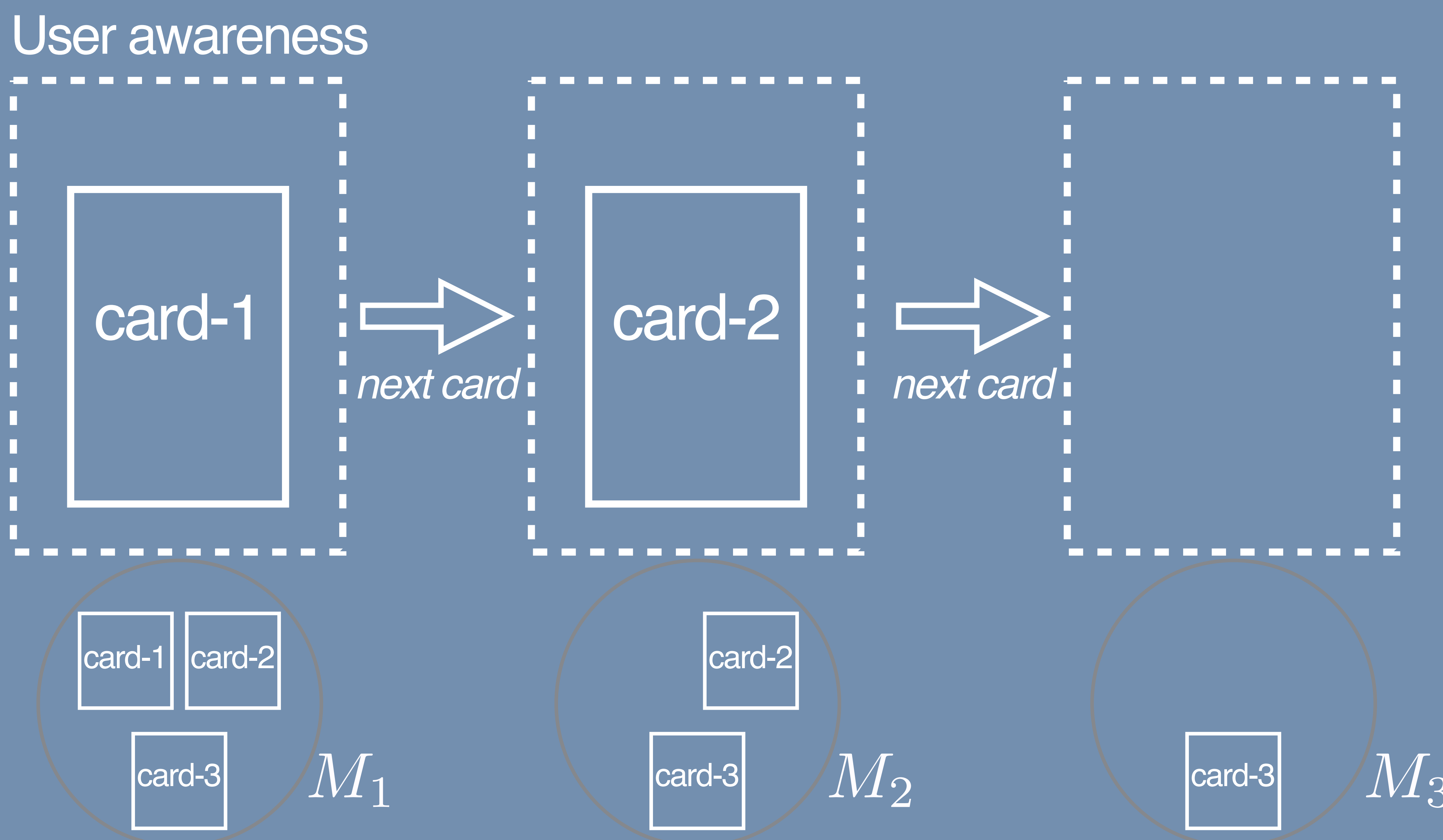


Candidate set:

Action select: $\text{card}_1 \sim \pi_{s,r}(M_1)$ $\text{card}_2 \sim \pi_{s,r}(M_2)$

Animation of ranking procedure with bandit

Horizontal scrolling

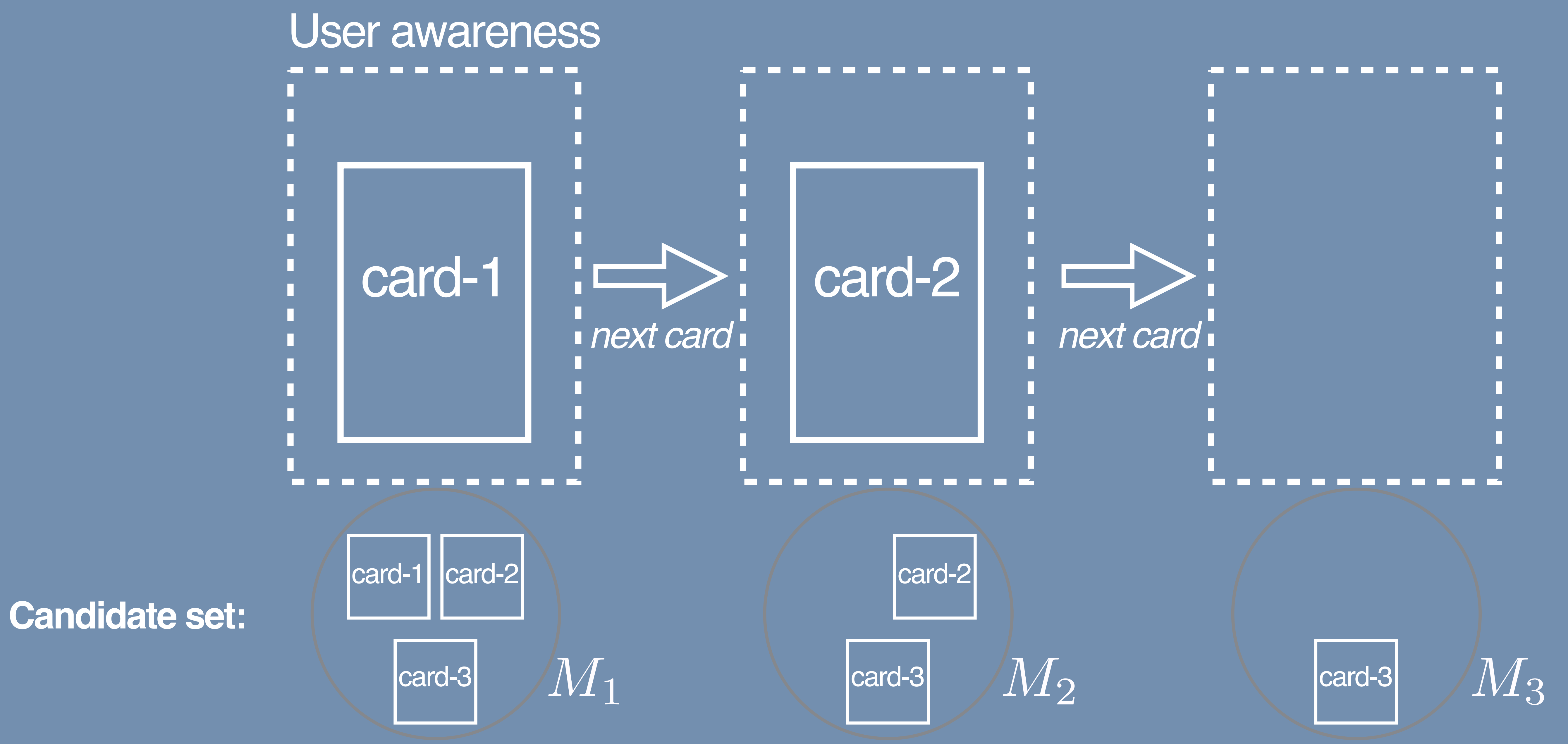


Candidate set:

Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$ $\mathrm{card}_2 \sim \pi_{s,r}(M_2)$

Animation of ranking procedure with bandit

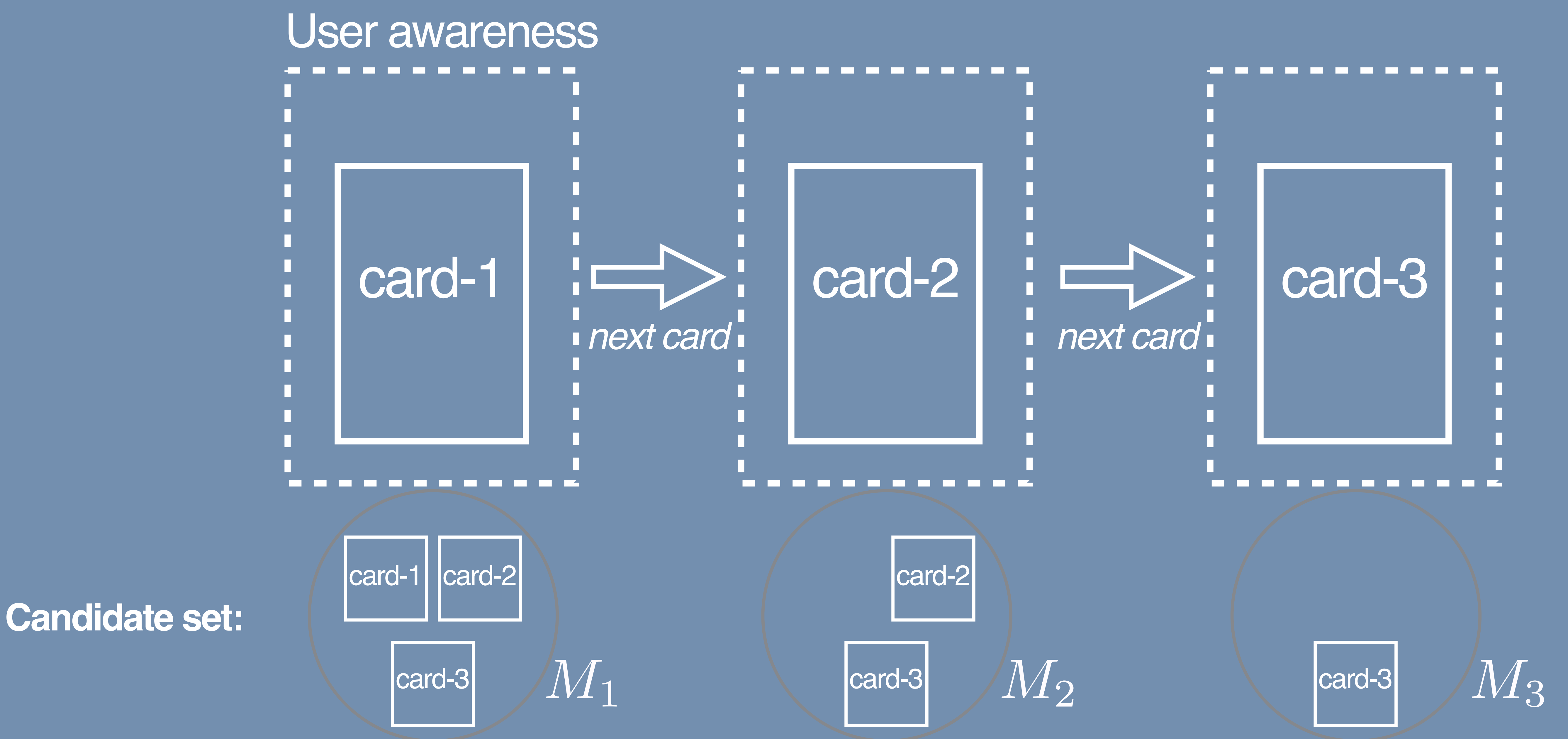
Horizontal scrolling



Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$ $\mathrm{card}_2 \sim \pi_{s,r}(M_2)$ $\mathrm{card}_3 \sim \pi_{s,r}(M_3)$

Animation of ranking procedure with bandit

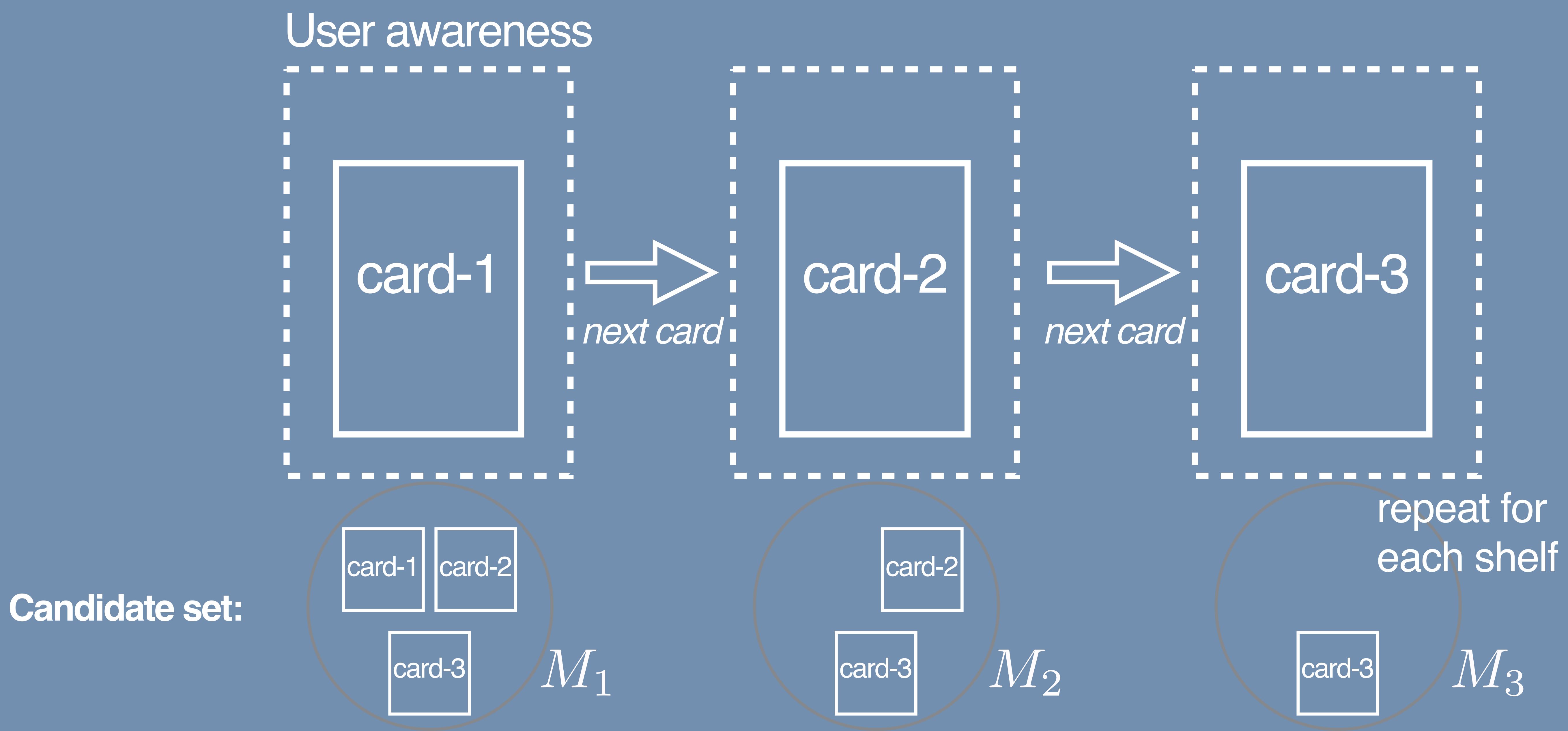
Horizontal scrolling



Action select: $\text{card}_1 \sim \pi_{s,r}(M_1)$ $\text{card}_2 \sim \pi_{s,r}(M_2)$ $\text{card}_3 \sim \pi_{s,r}(M_3)$

Animation of ranking procedure with bandit

Horizontal scrolling



Action select: $\mathrm{card}_1 \sim \pi_{s,r}(M_1)$ $\mathrm{card}_2 \sim \pi_{s,r}(M_2)$ $\mathrm{card}_3 \sim \pi_{s,r}(M_3)$

Animation of ranking procedure with bandit

Vertical scrolling

User awareness

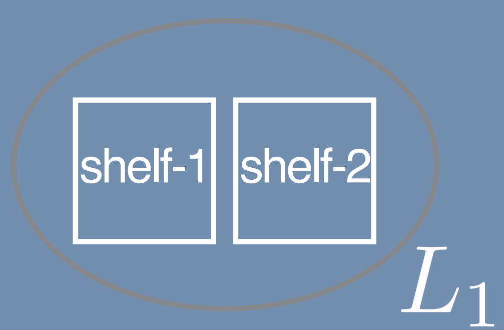
Animation of ranking procedure with bandit

Vertical scrolling

Animation of ranking procedure with bandit

Vertical scrolling

Candidate set



Action select

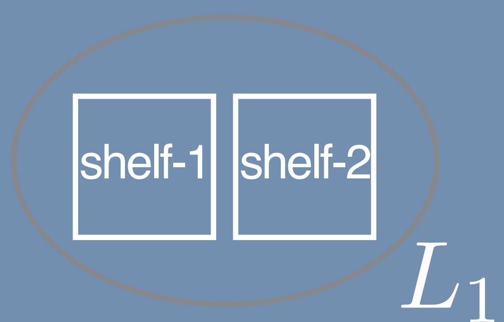
$$\text{shelf}_1 \sim \pi_{s,r'}(L_1)$$

User awareness

Animation of ranking procedure with bandit

Vertical scrolling

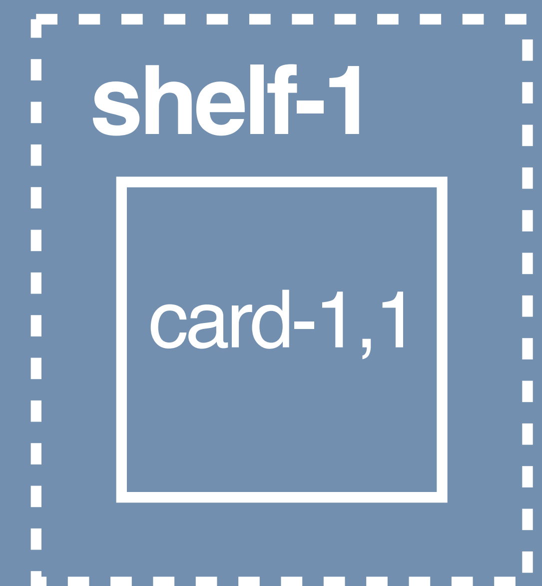
Candidate set



Action select

$$\text{shelf}_1 \sim \pi_{s,r'}(L_1)$$

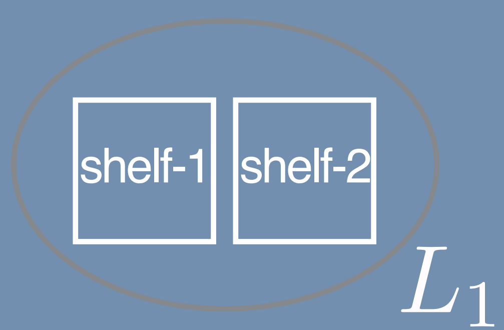
User awareness



Animation of ranking procedure with bandit

Vertical scrolling

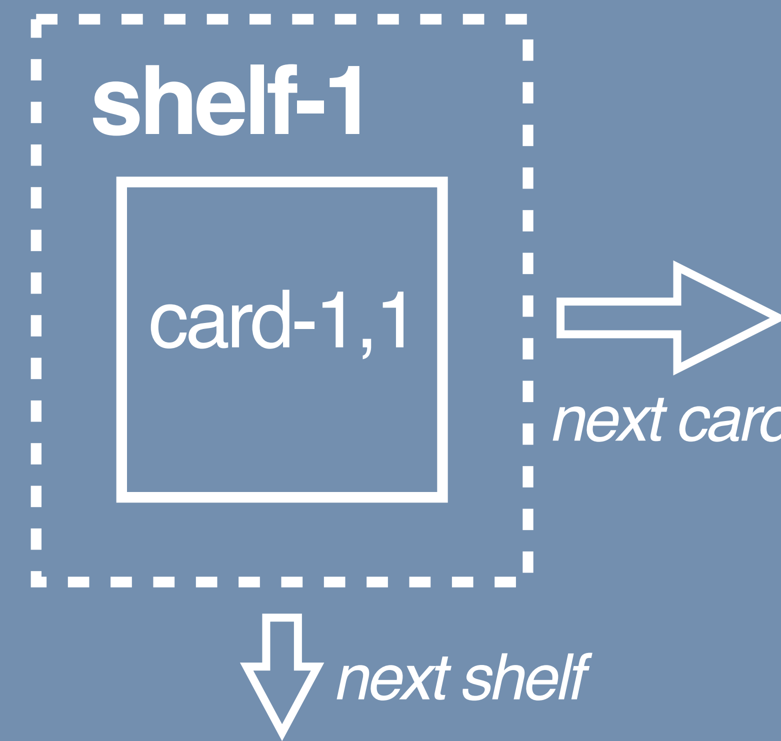
Candidate set



Action select

$$\mathrm{shelf}_1 \sim \pi_{s,r'}(L_1)$$

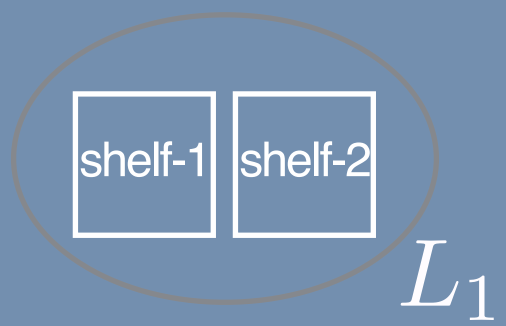
User awareness



Animation of ranking procedure with bandit

Vertical scrolling

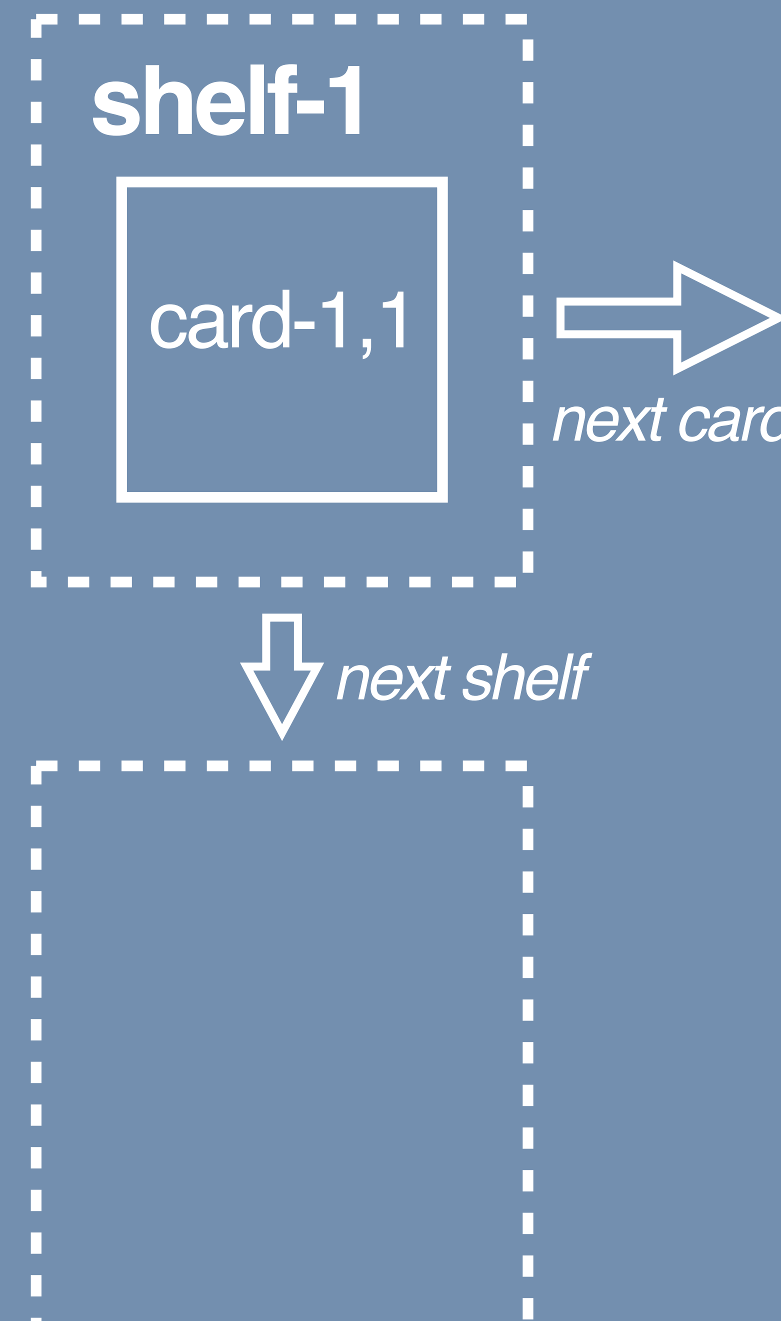
Candidate set



Action select

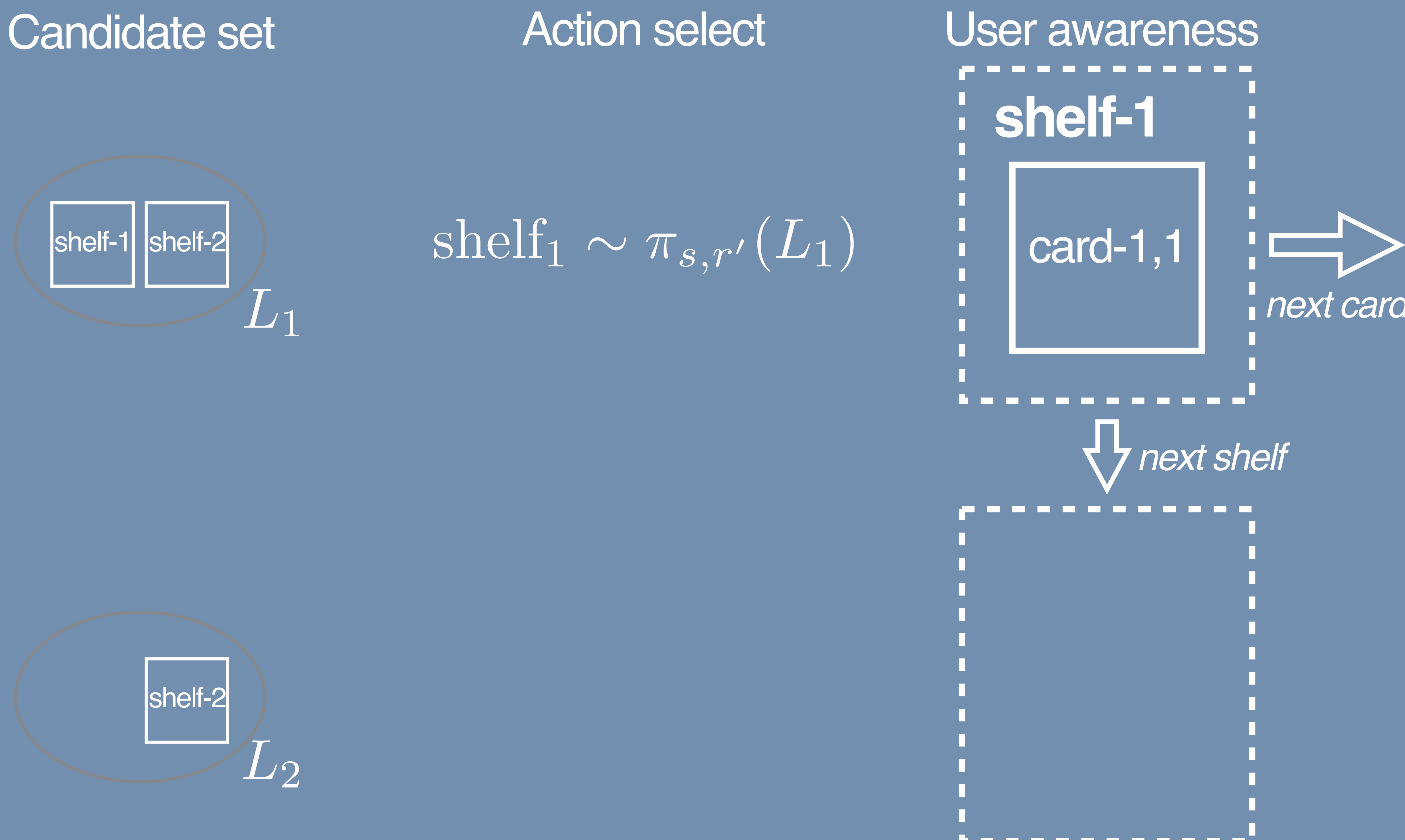
$$\text{shelf}_1 \sim \pi_{s,r'}(L_1)$$

User awareness



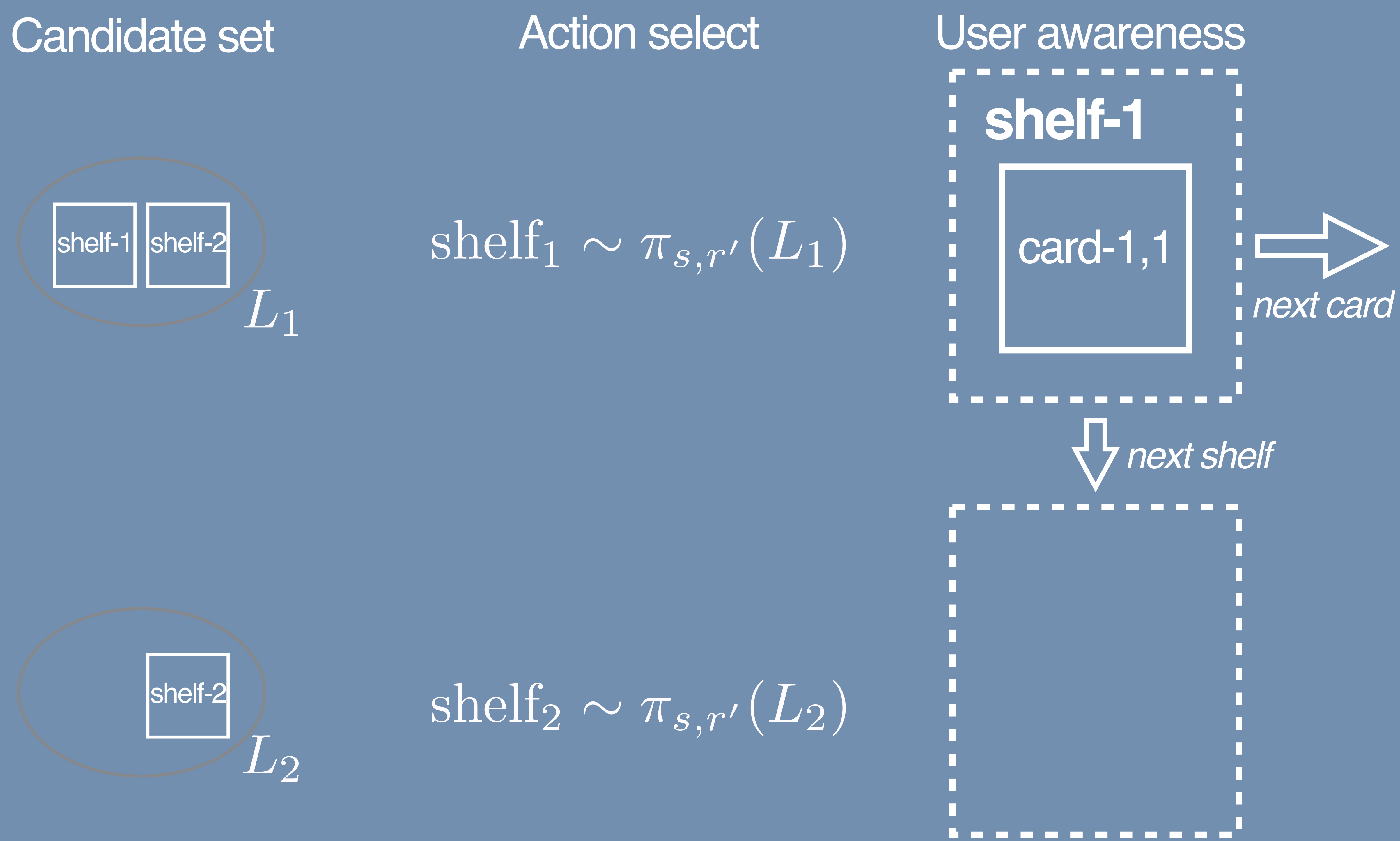
Animation of ranking procedure with bandit

Vertical scrolling



Animation of ranking procedure with bandit

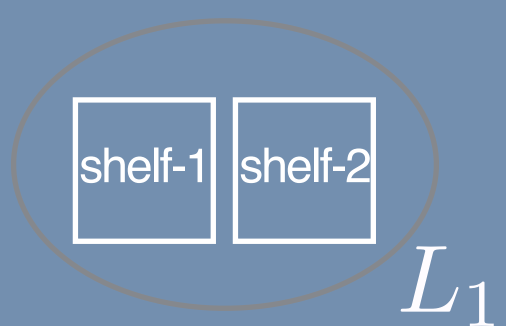
Vertical scrolling



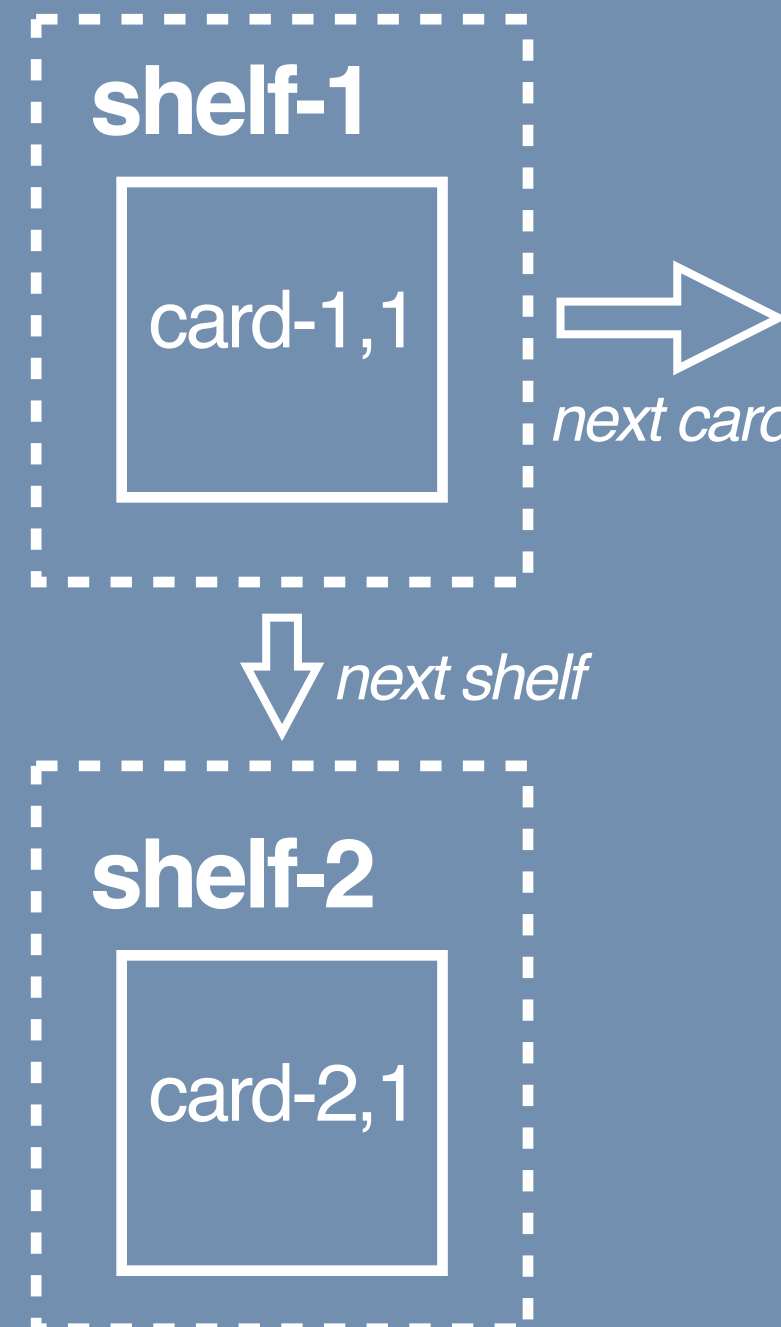
Animation of ranking procedure with bandit

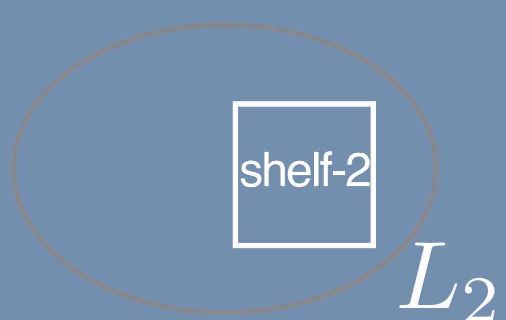
Vertical scrolling

Candidate set | Action select | User awareness



$$\text{shelf}_1 \sim \pi_{s,r'}(L_1)$$

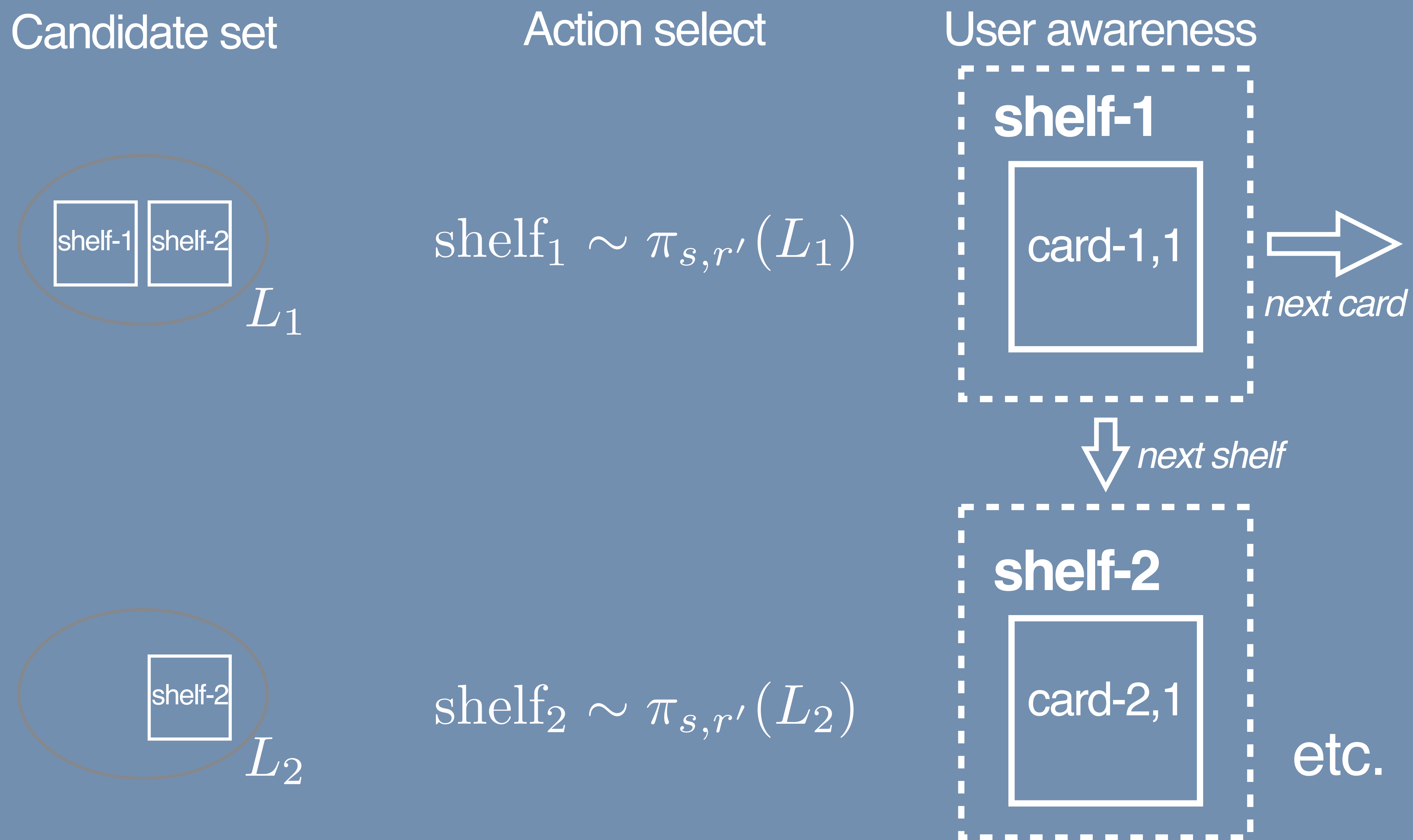




$$\text{shelf}_2 \sim \pi_{s,r'}(L_2)$$

Animation of ranking procedure with bandit

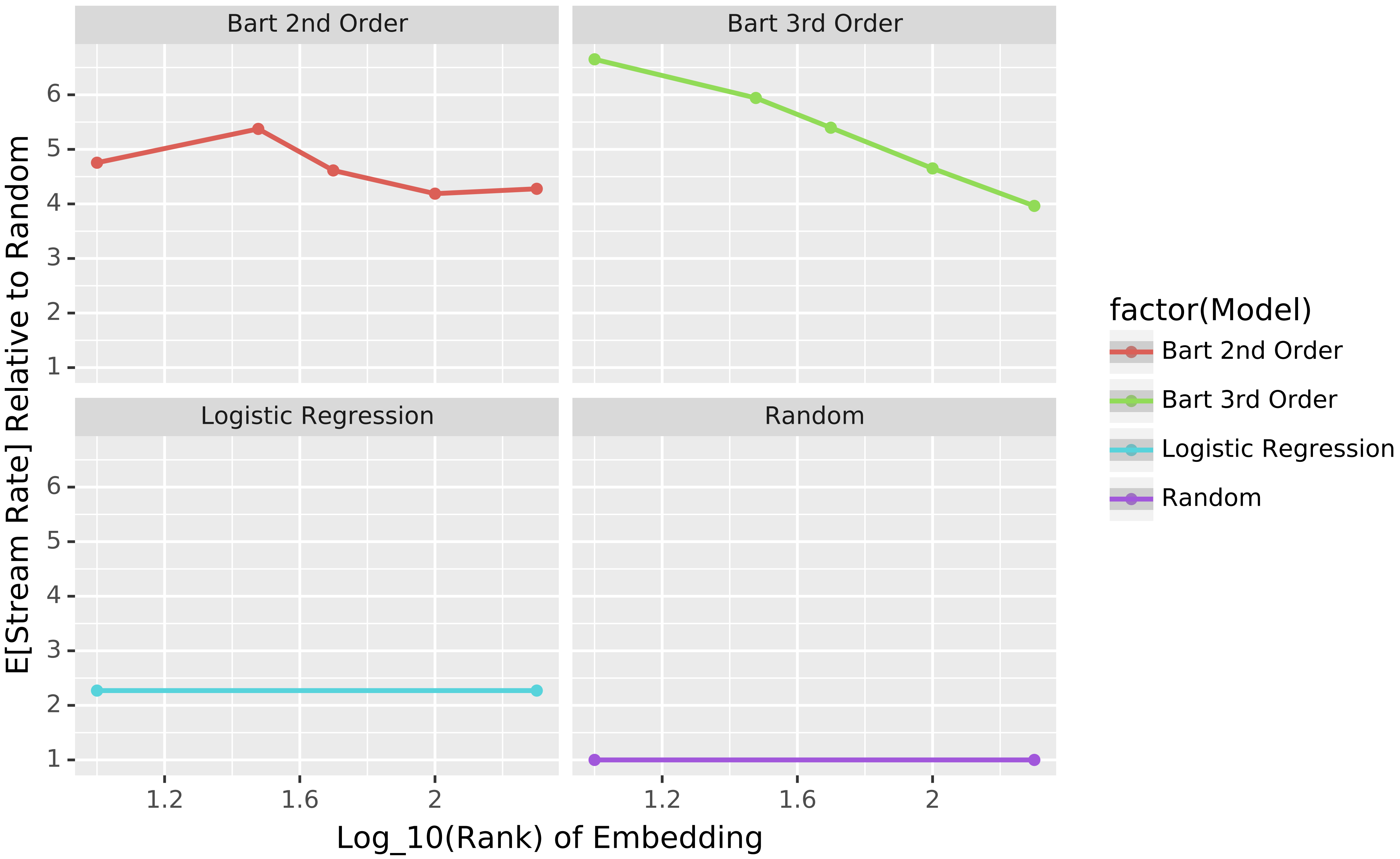
Vertical scrolling



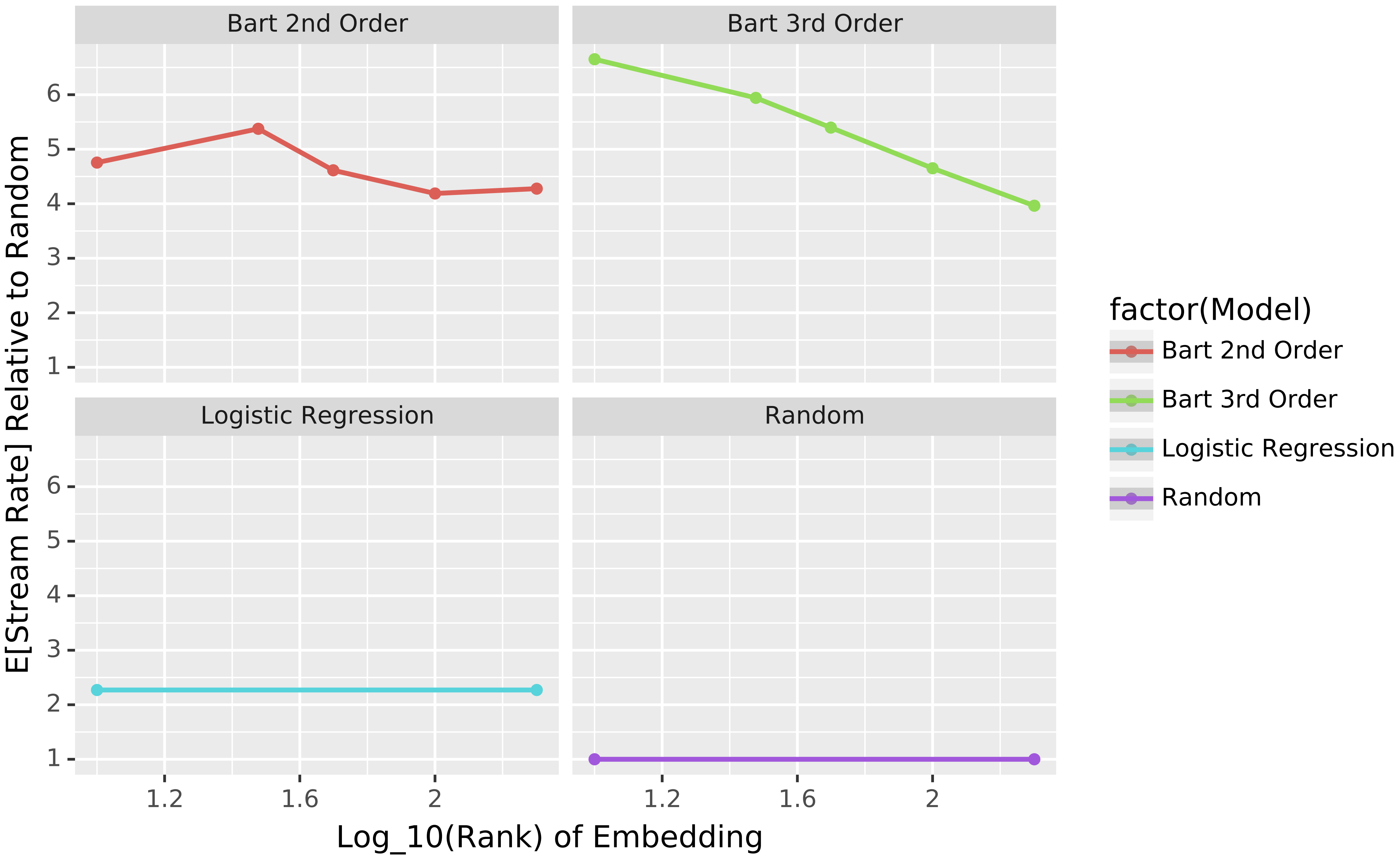
Experimental evaluation

- we collected randomized recommendation data
- offline experiments:
    - counterfactual estimation of A/B test performance using importance sampling reweighting
- online A/B test experiments

Offline experiments

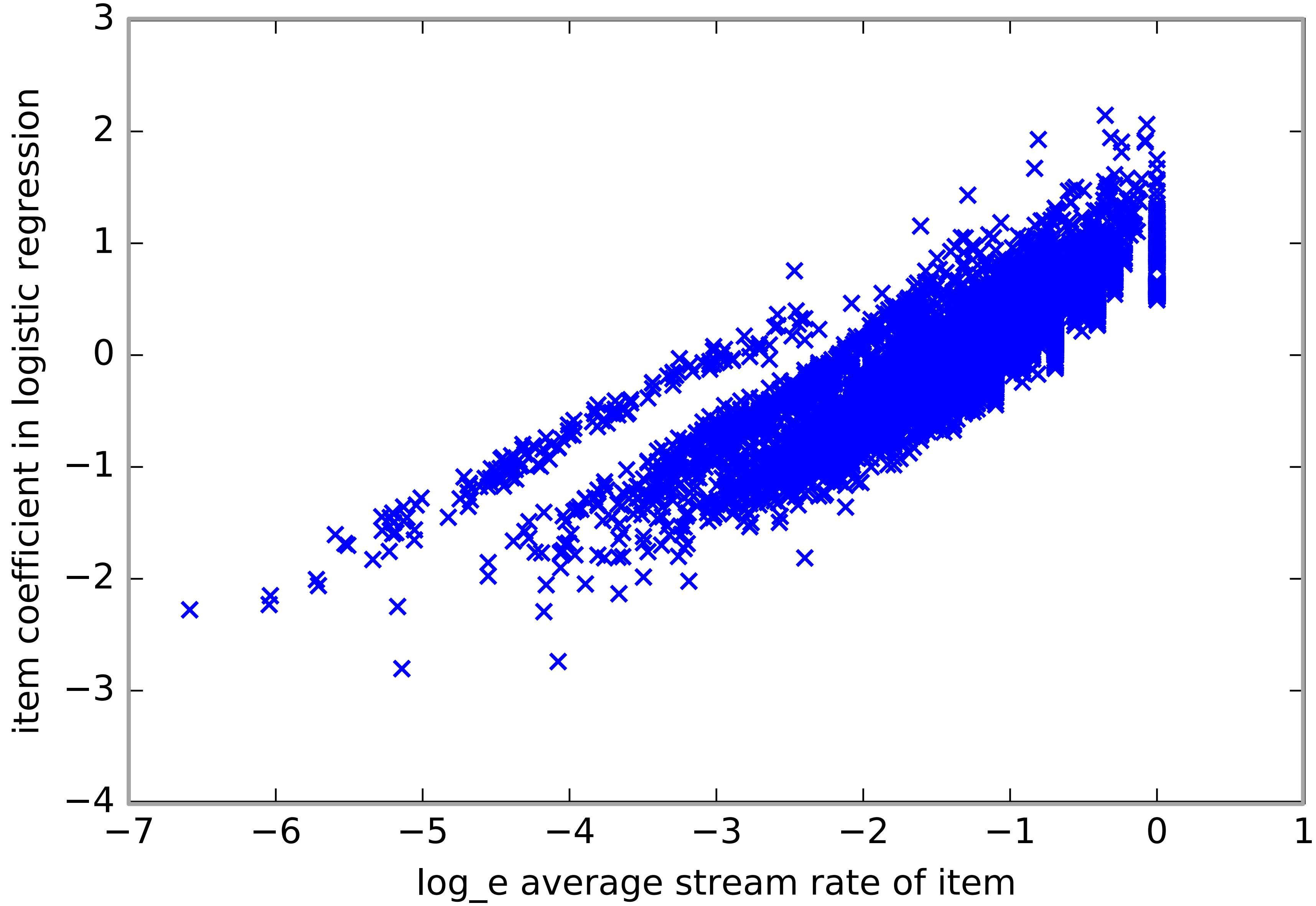


Offline experiments

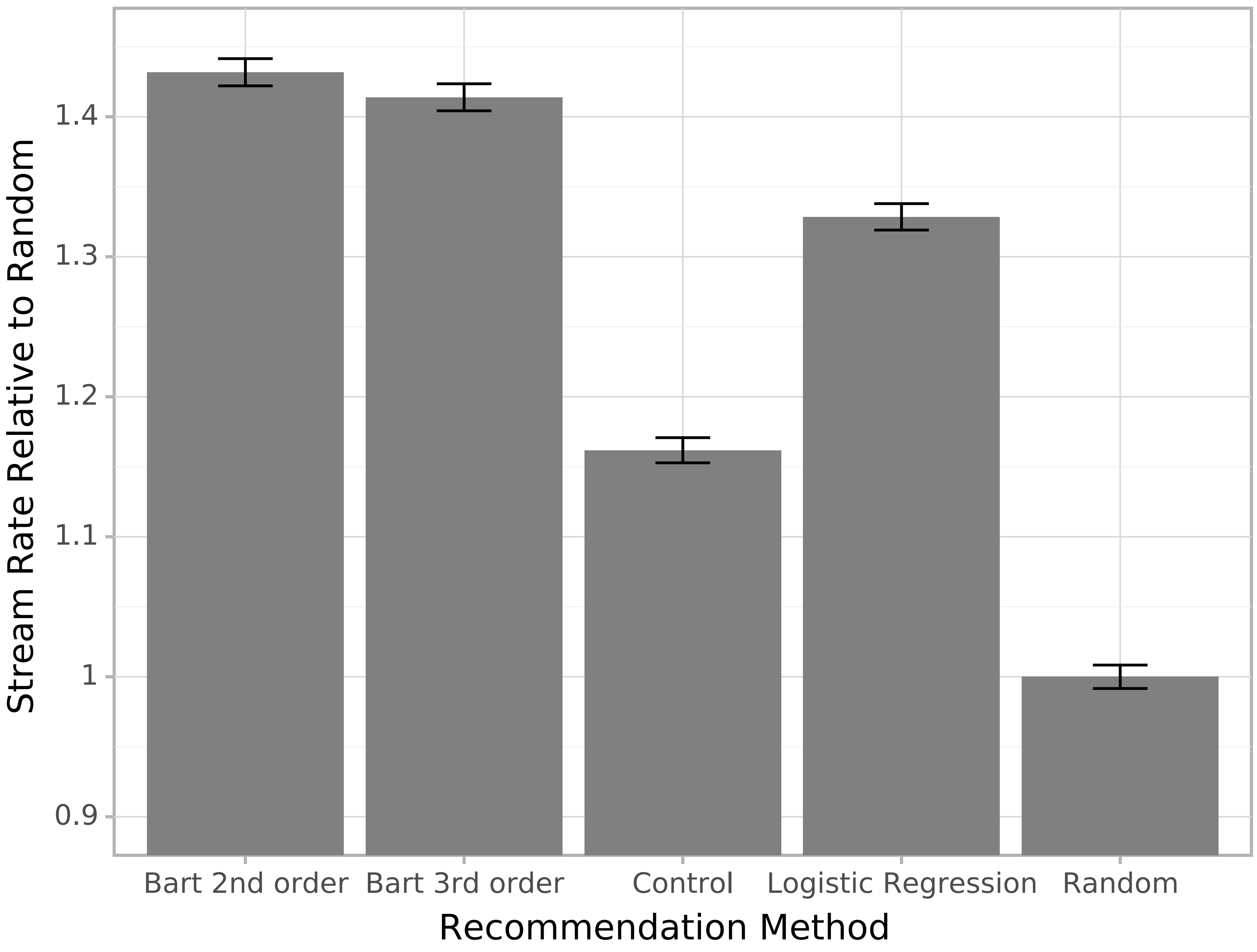


(similar conclusions as NDCG@10 for the metric)

Offline experiments



Online A/B test



Bart limitations and future work

- user preference model:
  - assumes independence of impression outcomes
  - attempts to estimate absolute reward, competitive pairwise model might improve predictions
  - maximizes our defined reward, does it approximate user satisfaction?
- ranking model not defined to promote diversity, slate recommendation could be incorporated
- exploration-exploitation over a <u>candidate set</u> not the full item set

Thank you, any questions?

email: jamesm@spotify.com