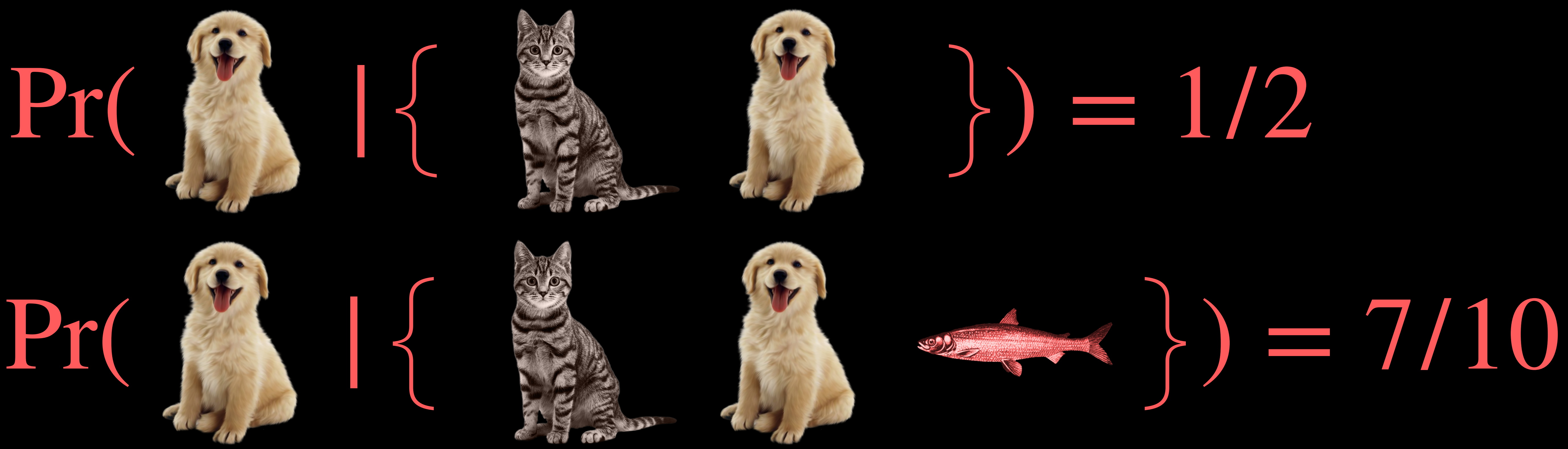
Code: bit.ly/csc-kdd-code
Slides: bit.ly/csc-kdd-slides



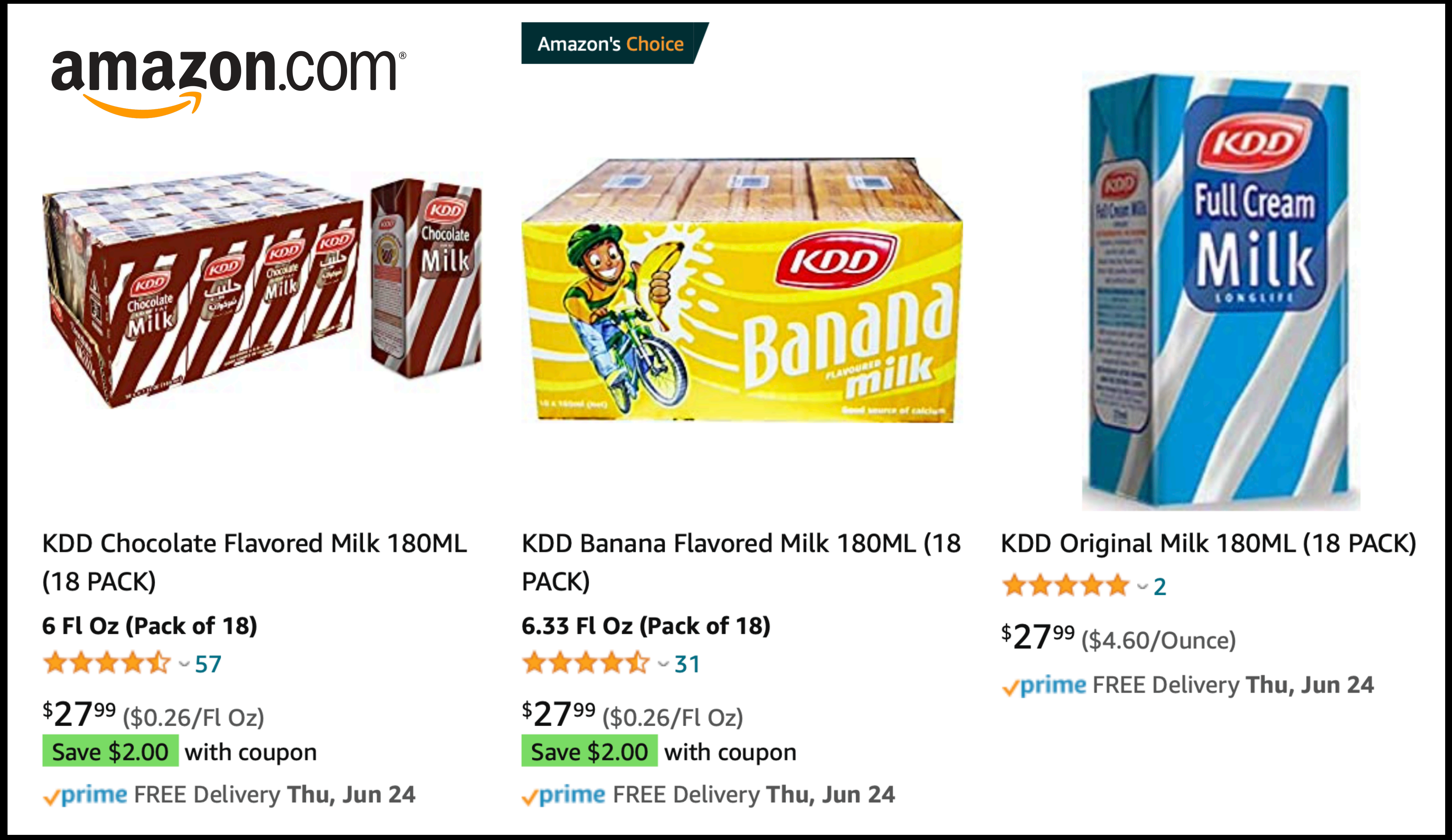
Choice Set Confounding in Discrete Choice

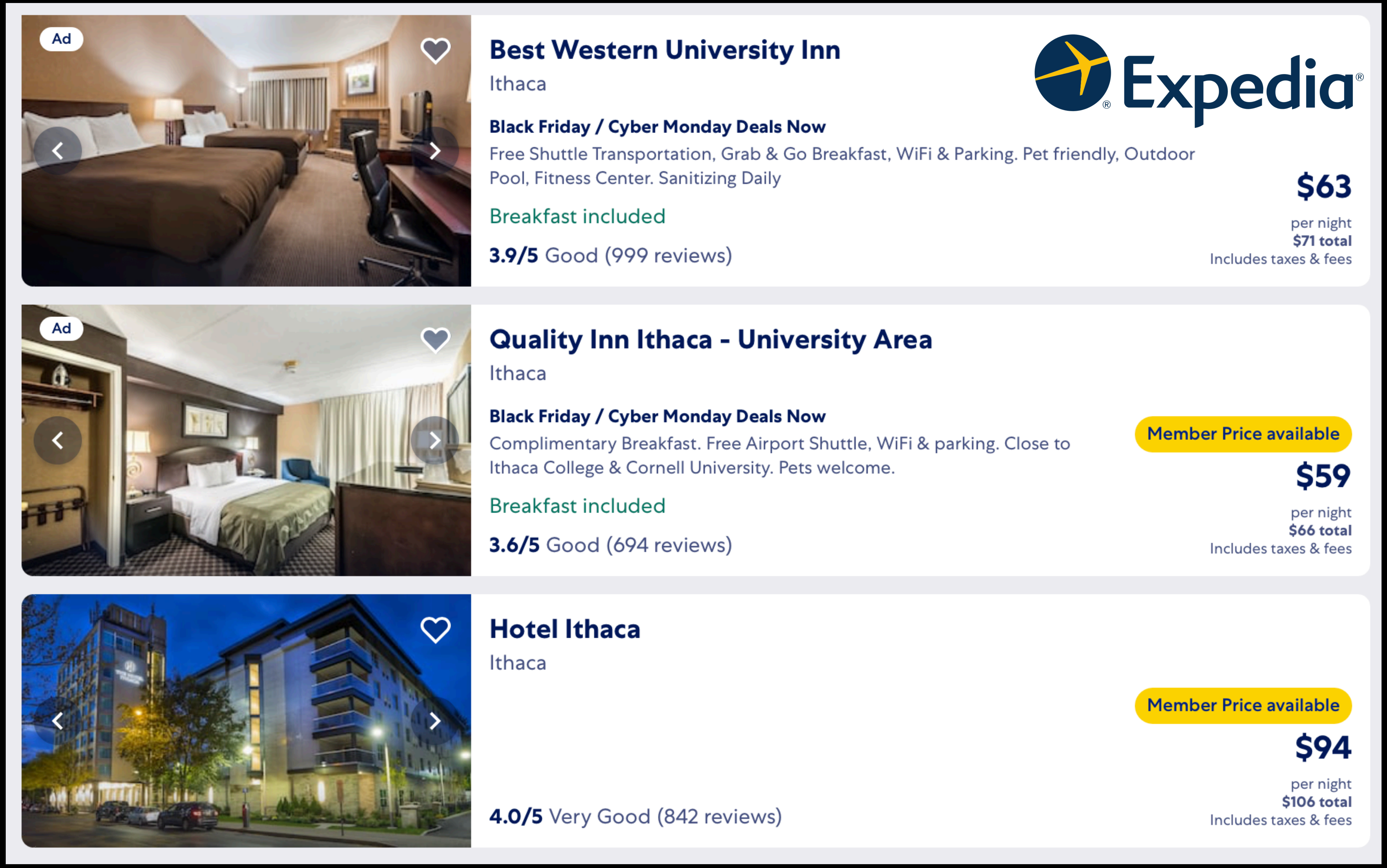
Kiran Tomlinson
PhD Student, Cornell

with Johan Ugander & Austin R. Benson

Choices and context effects

Discrete choices are everywhere

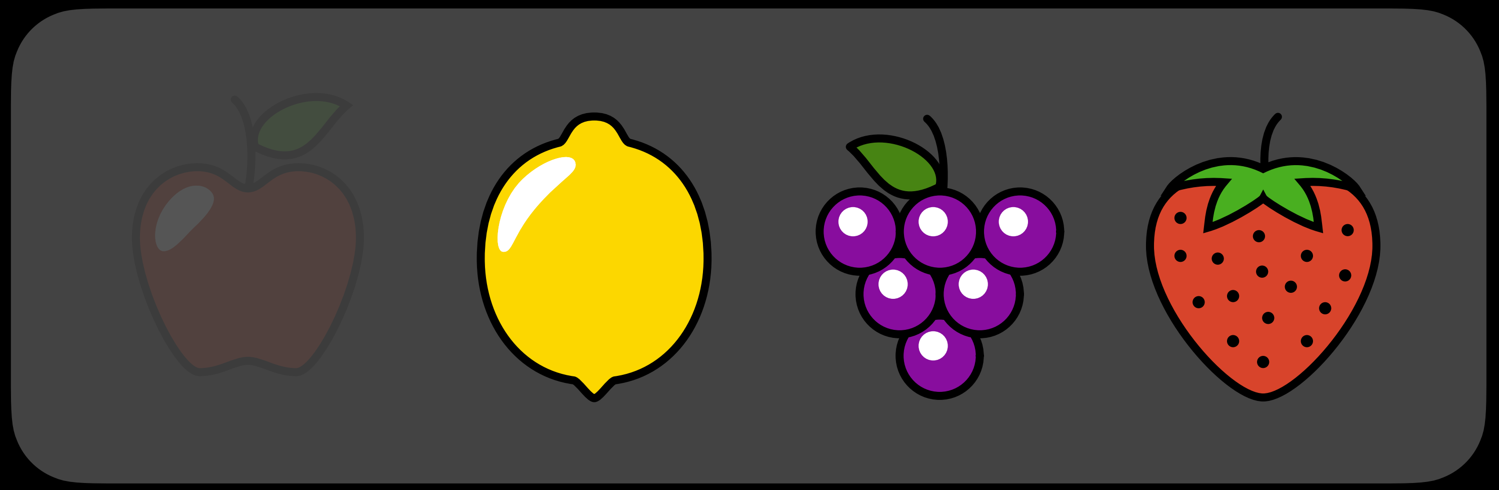




The goal of choice modeling: learn $\Pr(i \mid C)$

The goal of choice modeling: learn $\Pr(i \mid C)$

choice set

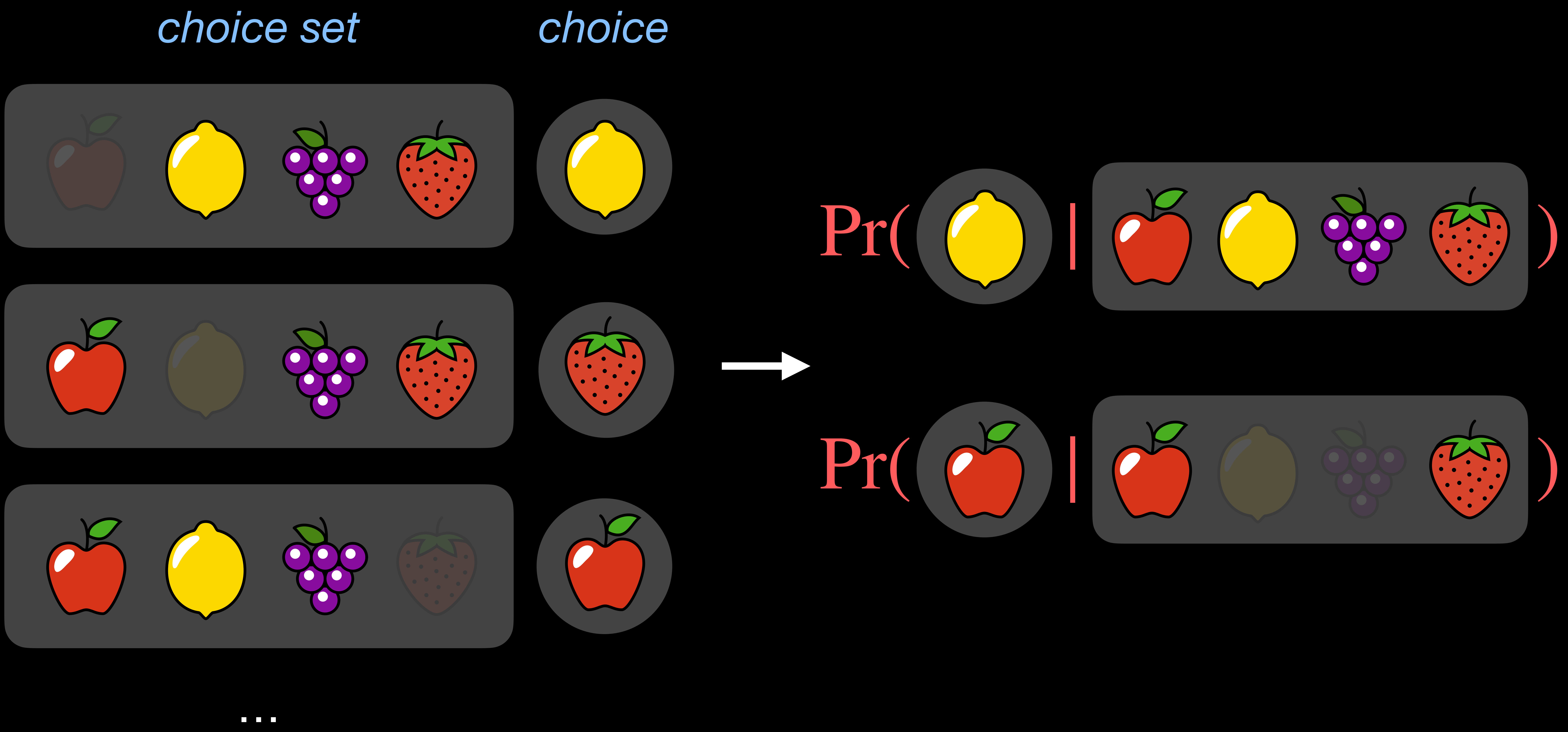


The goal of choice modeling: learn $\Pr(i \mid C)$

The goal of choice modeling: learn $\Pr(i \mid C)$

...

The goal of choice modeling: learn $\Pr(i \mid C)$



The classic model: *logit*

(McFadden, *Frontiers in Econometrics* 1973)

The classic model: *logit*

(McFadden, *Frontiers in Econometrics* 1973)

Assume *item* i has *utility* u_i

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

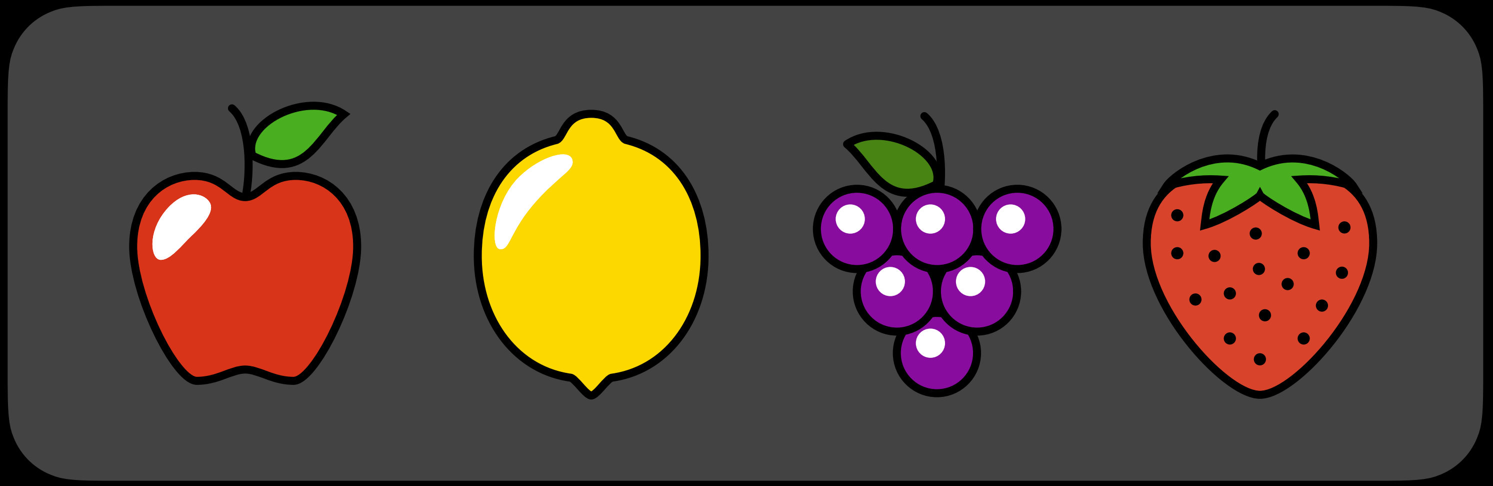
The classic model: *logit*

(McFadden, *Frontiers in Econometrics* 1973)

Assume *item* i has *utility* u_i

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

C



u_i	1	-1	0	2
$\Pr(i \mid C)$	.24	.03	.09	.64

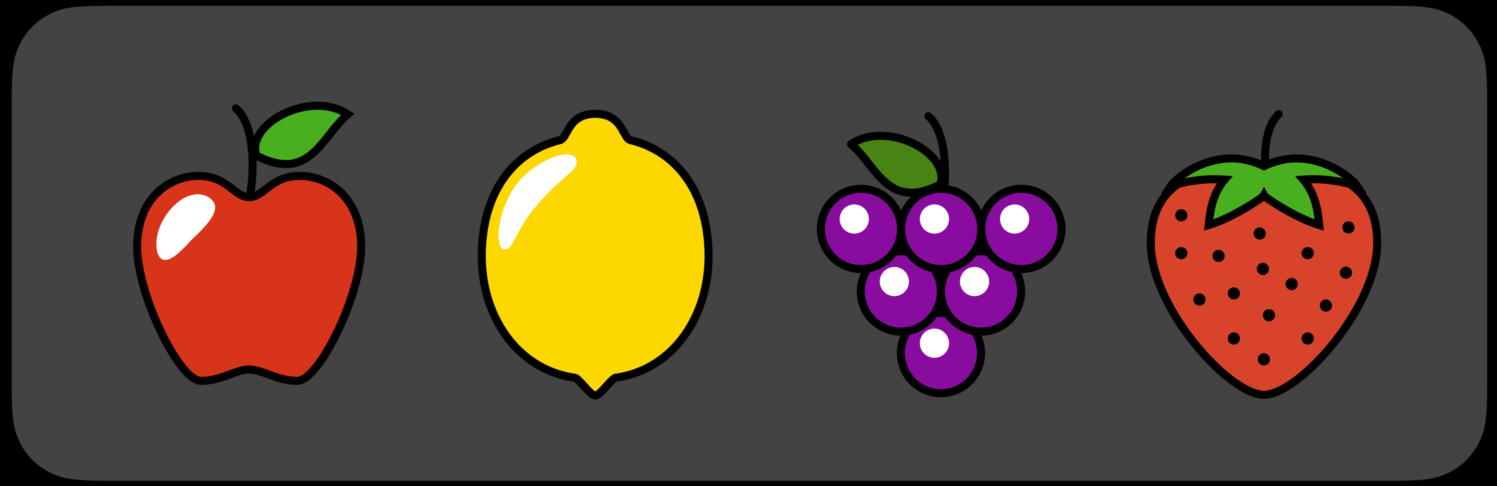
The classic model: *logit*

(McFadden, *Frontiers in Econometrics* 1973)

Assume *item* i has *utility* u_i

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

C



u_i	1	-1	0	2
$\Pr(i \mid C)$	.24	.03	.09	.64

Unique choice model satisfying
independence of irrelevant alternatives (IIA):

(Luce, *Individual Choice Behavior* 1959)

$$\frac{\Pr(i \mid C)}{\Pr(j \mid C)} = \frac{\Pr(i \mid C')}{\Pr(j \mid C')}$$

Accurate models need to account for *context effects*

Accurate models need to account for *context effects*

The choice set influences preferences.

Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)

Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)

Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)

Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise

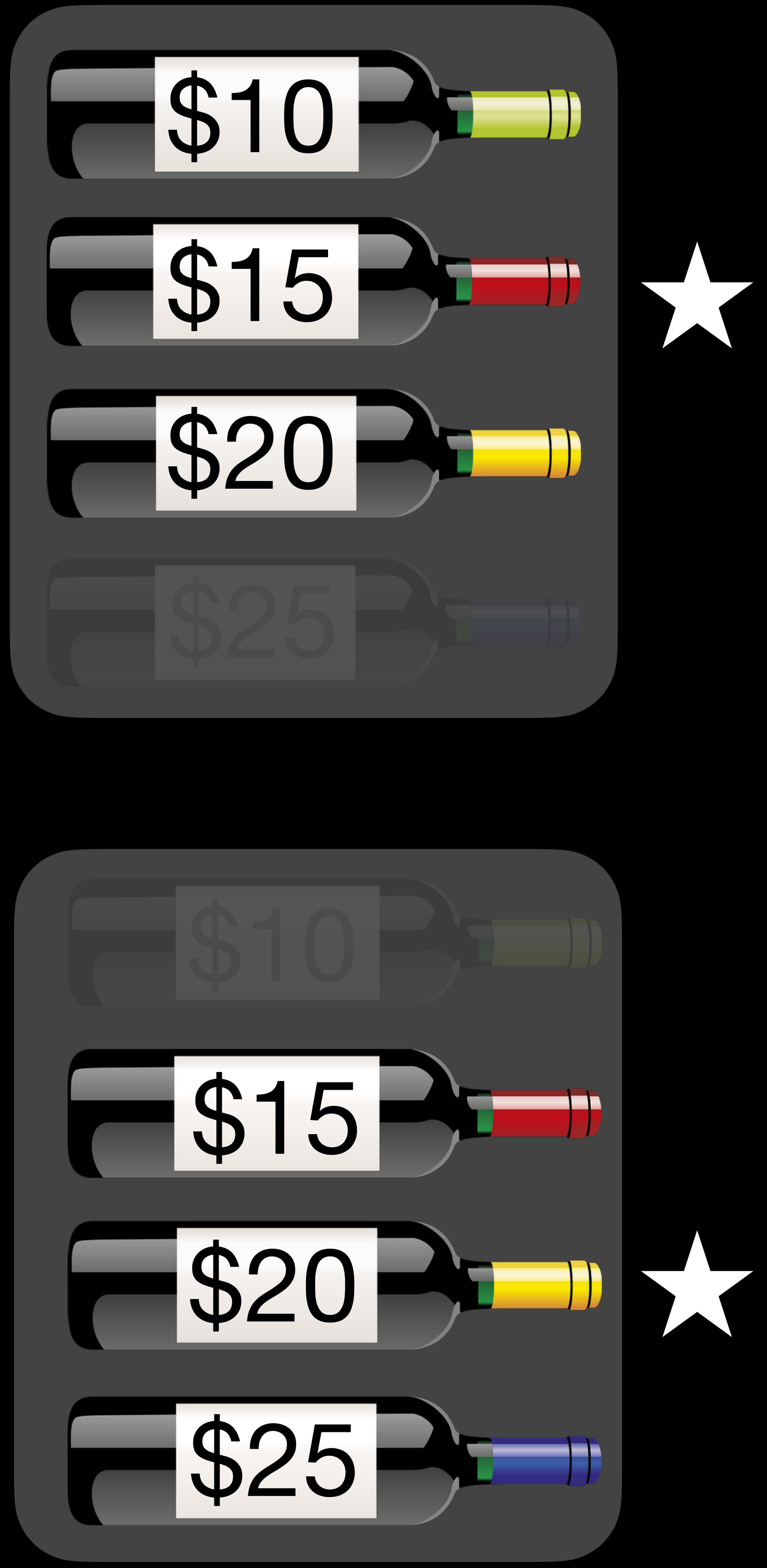
(Simonson, 1989)

Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise

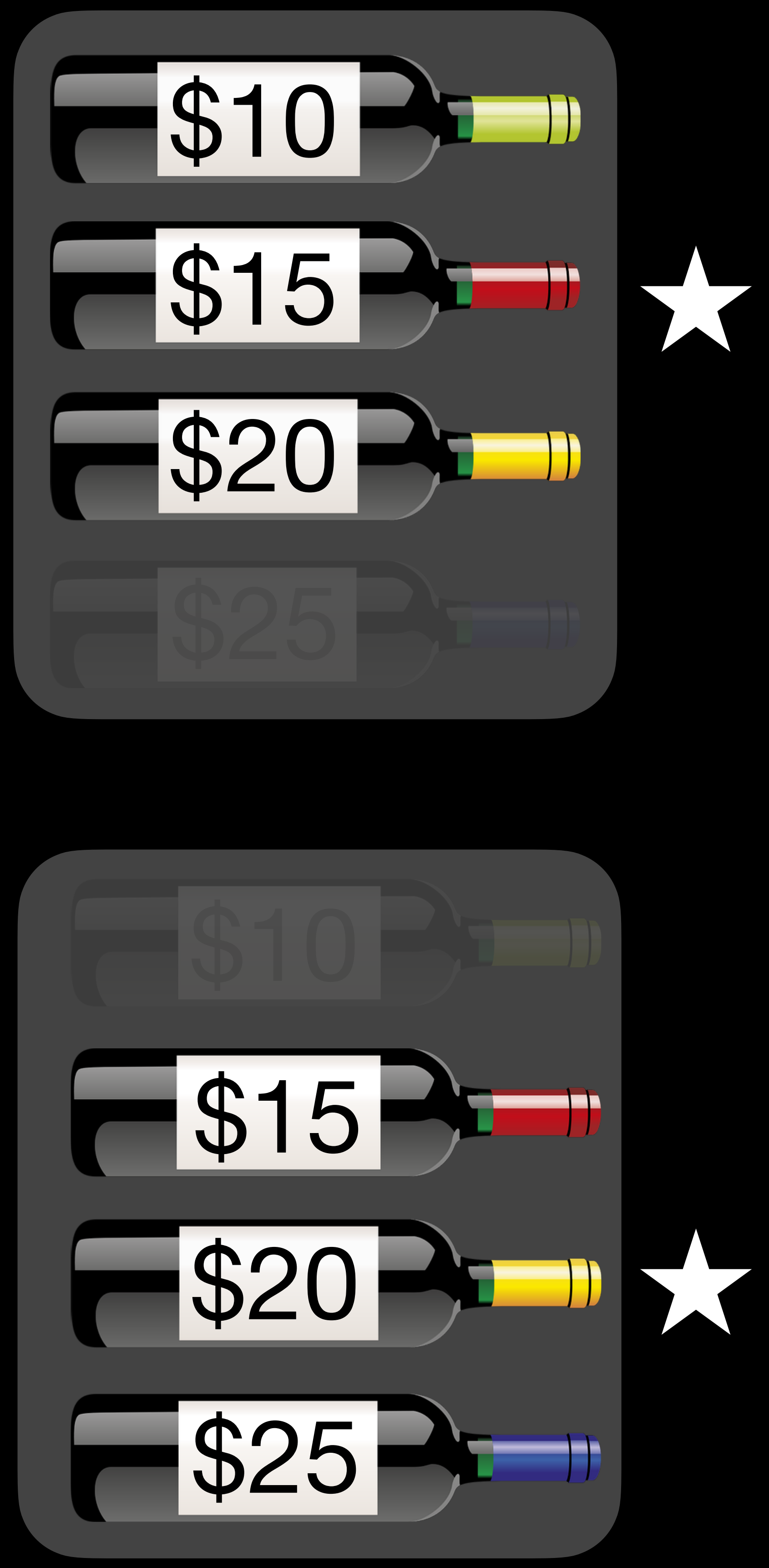
(Simonson, 1989)



Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise
(Simonson, 1989)



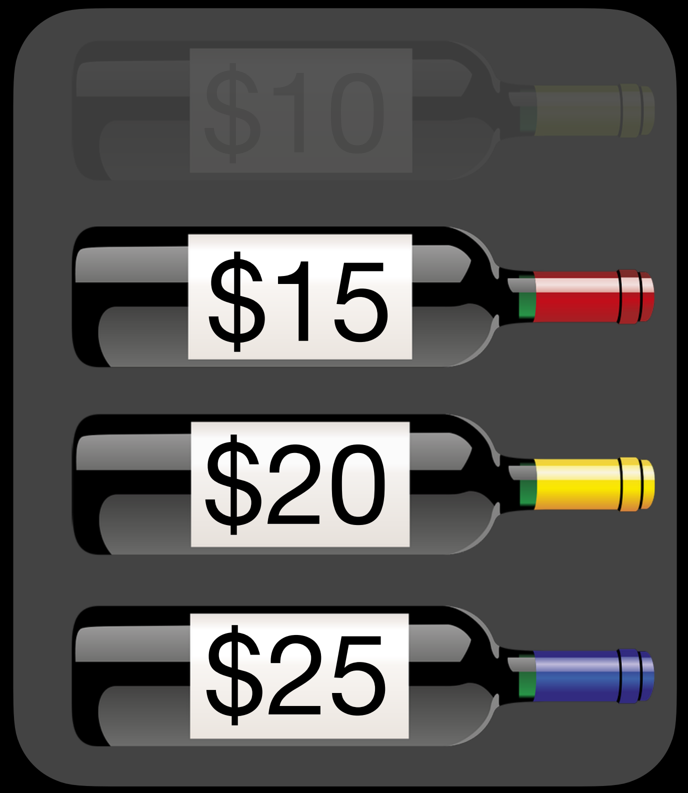
Similarity
(Tversky, 1972)

Accurate models need to account for *context effects*

The choice set influences preferences.

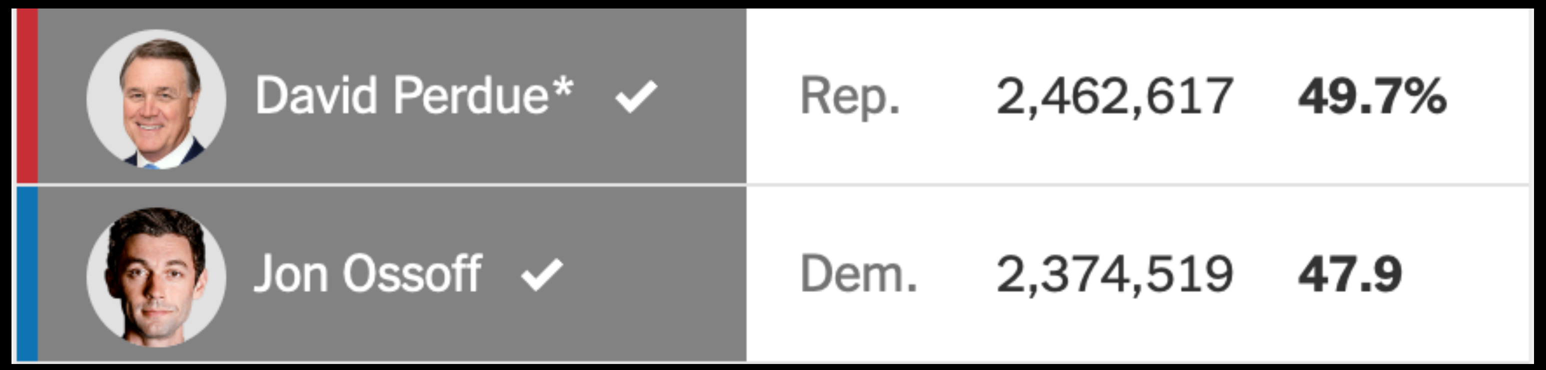
Compromise

(Simonson, 1989)



Similarity

(Tversky, 1972)

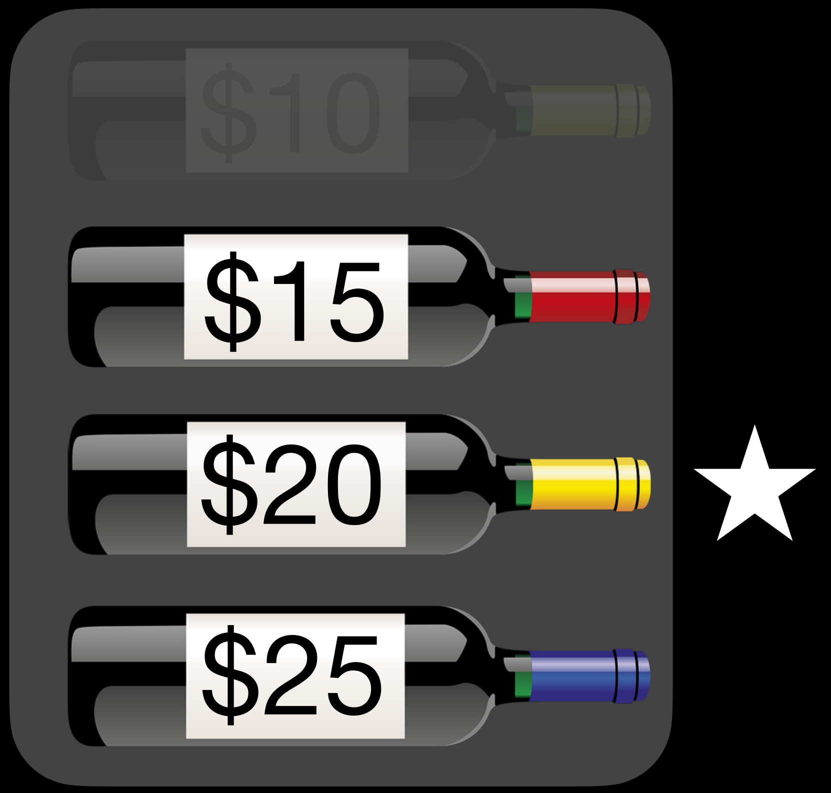


Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)



Similarity

(Tversky, 1972)

David Perdue* ✔	Rep.	2,462,617	**49.7%**
Jon Ossoff ✔	Dem.	2,374,519	**47.9**

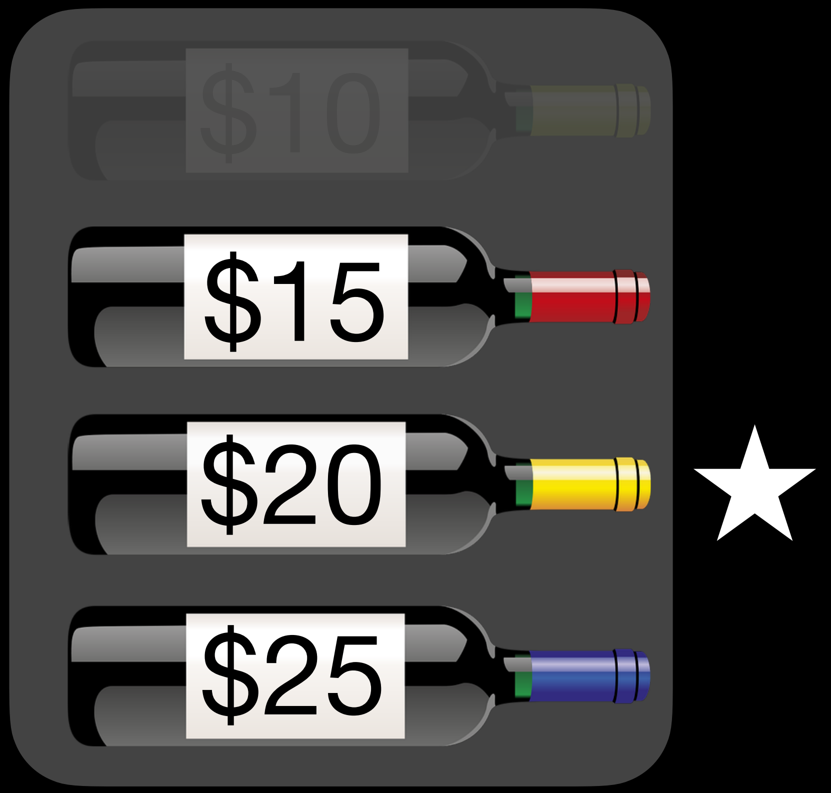
Raphael Warnock ✔	Dem.	1,617,035	**32.9%**
Kelly Loeffler* ✔	Rep.	1,273,214	**25.9**
Doug Collins	Rep.	980,454	**20.0**

Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)



Similarity

(Tversky, 1972)

David Perdue* ✔	Rep.	2,462,617	**49.7%**
Jon Ossoff ✔	Dem.	2,374,519	**47.9**

Raphael Warnock ✔	Dem.	1,617,035	**32.9%**
Kelly Loeffler* ✔	Rep.	1,273,214	**25.9**
Doug Collins	Rep.	980,454	**20.0**

Violations of IIA:

$$\frac{\Pr(i \mid C)}{\Pr(j \mid C)} \neq \frac{\Pr(i \mid C')}{\Pr(j \mid C')}$$

Accurate models need to account for *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)

★

★

Similarity

(Tversky, 1972)

David Perdue* ✔	Rep.	2,462,617	**49.7%**
Jon Ossoff ✔	Dem.	2,374,519	**47.9**

Raphael Warnock ✔	Dem.	1,617,035	**32.9%**
Kelly Loeffler* ✔	Rep.	1,273,214	**25.9**
Doug Collins	Rep.	980,454	**20.0**

Violations of IIA:

$$\frac{\Pr(i \mid C)}{\Pr(j \mid C)} \neq \frac{\Pr(i \mid C')}{\Pr(j \mid C')}$$

Recent contextual modeling

Chen & Joachims (KDD '16)
Ragain & Ugander (NeurIPS '16)
Seshadri et al. (ICML '19)
Bower & Balzano (ICML '20)
Rosenfeld et al. (ICML '20)
Tomlinson & Benson (KDD '21)

Choice set confounding

What if our data has heterogeneous preferences?

What if our data has heterogeneous preferences?

Each chooser a has their own choice probabilities: $\Pr(i \mid a, C)$

What if our data has heterogeneous preferences?

Each chooser a has their own choice probabilities: $\Pr(i \mid a, C)$

Question:
If we learn a model for $\Pr(i \mid C)$*, will this accurately reflect average choice behavior* $\mathbb{E}_a \Pr(i \mid a, C)$*?*

What if our data has heterogeneous preferences?

Each chooser a has their own choice probabilities: $\Pr(i \mid a, C)$

Question:
If we learn a model for $\Pr(i \mid C)$*, will this accurately reflect average choice behavior* $\mathbb{E}_a \Pr(i \mid a, C)$*?*

→ **Not in general.** You either need

chooser-independent preferences: $\Pr(i \mid a, C) = \Pr(i \mid C)$

or

chooser-independent choice sets: $\Pr(C) = \Pr(C \mid a)$

What if our data has heterogeneous preferences?

Each chooser a has their own choice probabilities: $\Pr(i \mid a, C)$

Question:
If we learn a model for $\Pr(i \mid C)$*, will this accurately reflect average choice behavior* $\mathbb{E}_a \Pr(i \mid a, C)$*?*

→ **Not in general.** You either need

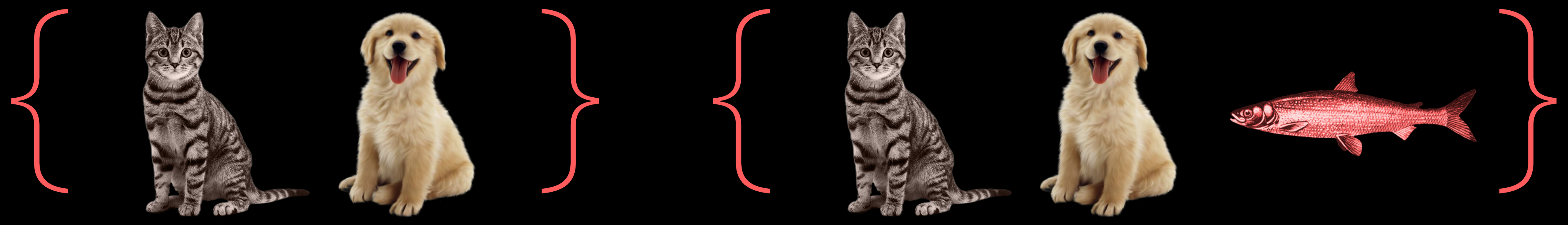
chooser-independent preferences: $\Pr(i \mid a, C) = \Pr(i \mid C)$
or
chooser-independent choice sets: $\Pr(C) = \Pr(C \mid a)$

chooser-dependent preferences **and** *chooser-dependent choice sets*
→ *choice set confounding*

Choice set confounding example

1/4

3/4



Choice set confounding **example**

1/4

3/4

Choice probabilities:

	{ 	 }	{ 		 }
	3/4	1/4	3/4	1/4	0
	1/4	3/4	1/4	3/4	0

Choice set confounding example

1/4

3/4

Choice probabilities:

					
	3/4	1/4	3/4	1/4	0
	1/4	3/4	1/4	3/4	0

Choice set assignment probabilities:

		
	3/4	1/4
	1/4	3/4

Choice set confounding example

1/4

3/4

Choice probabilities:

					
	3/4	1/4	3/4	1/4	0
	1/4	3/4	1/4	3/4	0

$\mathbb{E}_a \Pr($  $\mid a,$  $)$

$= \mathbb{E}_a \Pr($  $\mid a,$  $)$

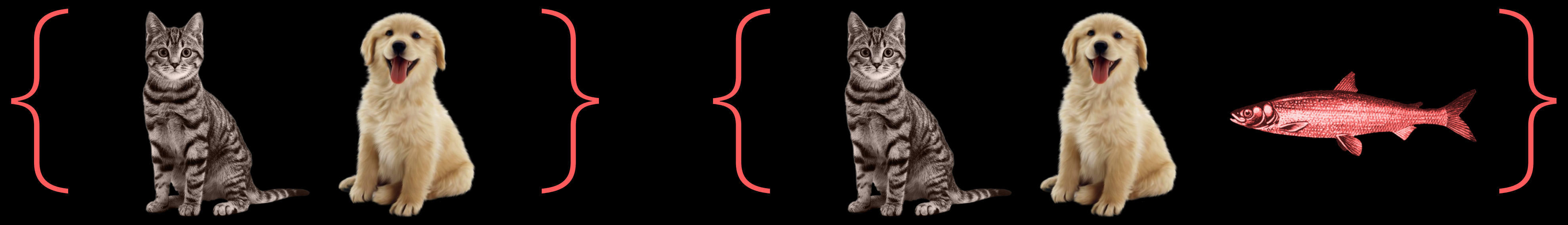
Choice set assignment probabilities:

		
	3/4	1/4
	1/4	3/4

Choice set confounding example

 1/4  3/4

Choice probabilities:

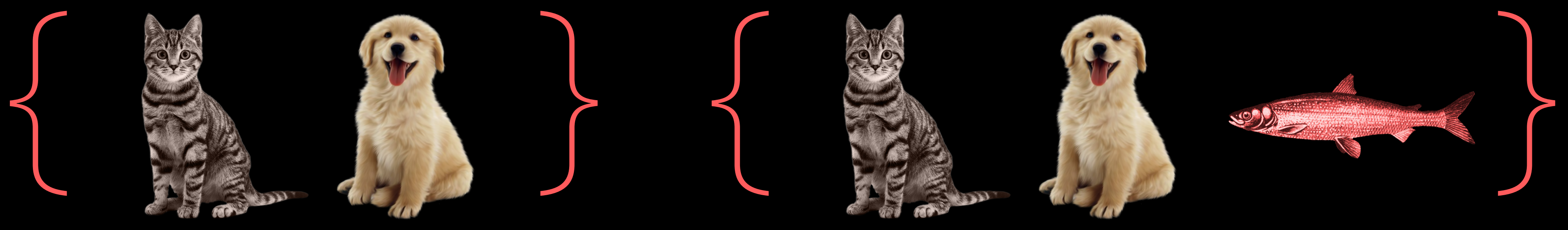
		
	3/4 1/4	3/4 1/4 0
	1/4 3/4	1/4 3/4 0

$\mathbb{E}_a \Pr($  $\mid a,$  $)$

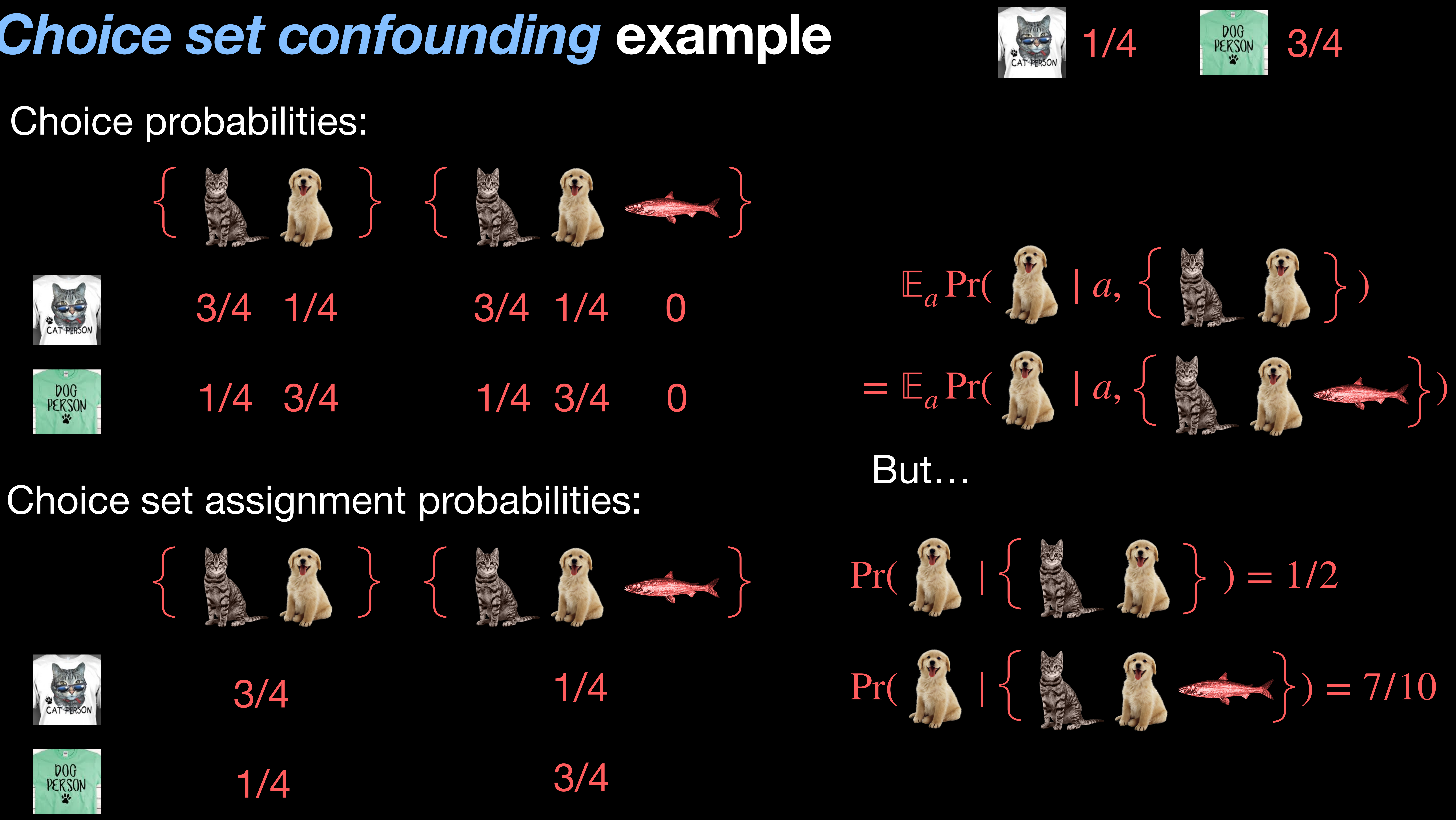
$= \mathbb{E}_a \Pr($  $\mid a,$  $)$

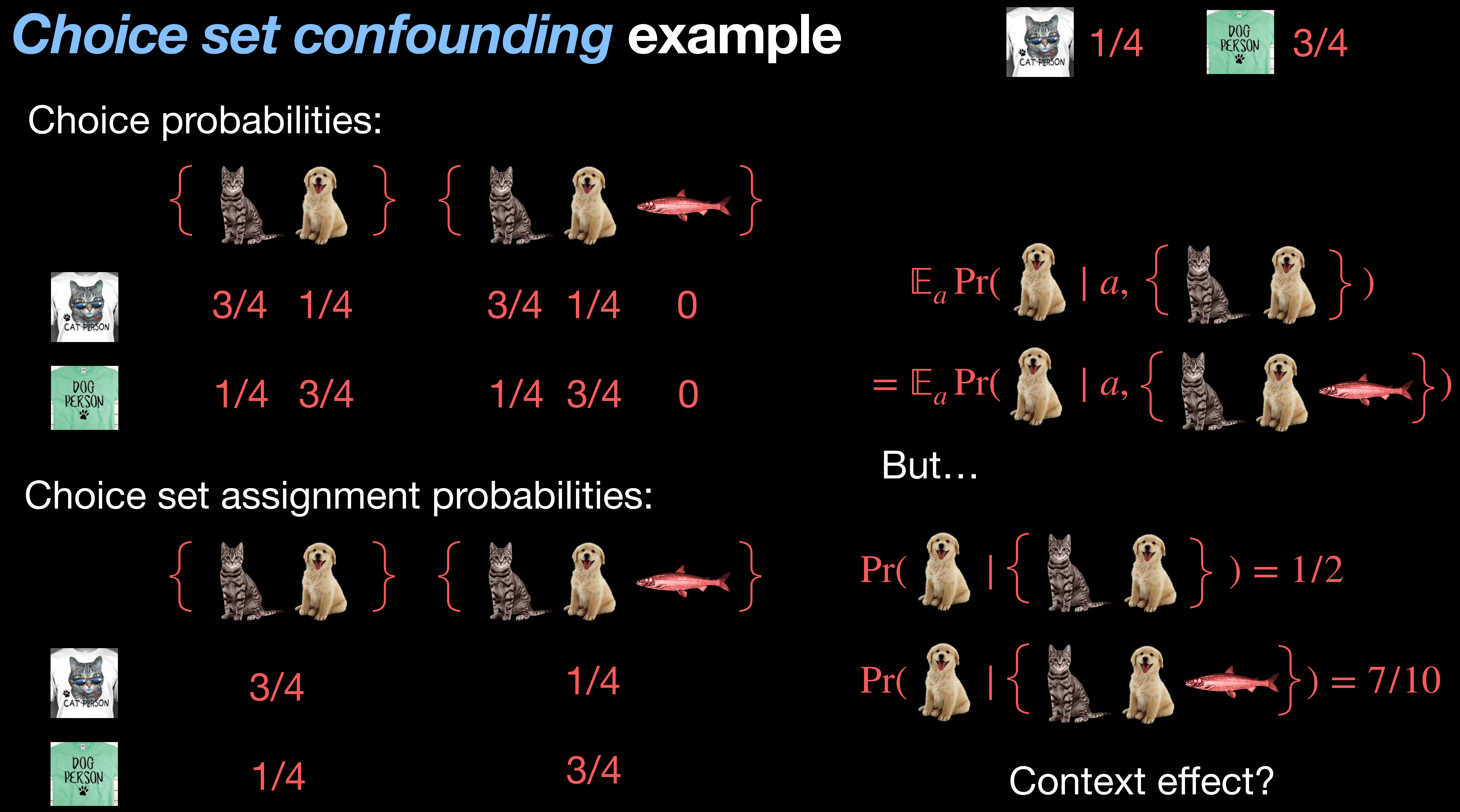
But…

Choice set assignment probabilities:

		
	3/4	1/4
	1/4	3/4

Choice set confounding example





Choice set confounding example
1/4
DOG PERSON
3/4
Choice probabilities:
CAT PERSON
3/4 1/4 3/4 1/4 0
DOG PERSON
1/4 3/4 1/4 3/4 0
$\mathbb{E}_a \mathrm{Pr}($ | a, { } $)$
$= \mathbb{E}_a \mathrm{Pr}($ | a, { } $)$
But…
Choice set assignment probabilities:
CAT PERSON
3/4
1/4
DOG PERSON
1/4
3/4
Pr(| { }) = 1/2
Pr(| { }) = 7/10
Context effect?

Choice set confounding example

1/4 3/4

Choice probabilities:

	1/4 3/	

$\mathbb{E}_a \Pr(\mid a, \{ \})$

$\mathbb{E}_a \Pr(\mid a, \{ \})$

Choice set assignment probabili

	3/4	1/4
	1/4	3/4

$\Pr($ $1/2$

$\Pr(\mid \{) = 7/10$

Context effect?

Choice set confounding in real data

Choice set confounding in real data

SFWork & SFShop

(Koppelman & Bhat, 2006)

San Francisco transportation data

Choice set confounding in real data

SFWork & SFShop

(Koppelman & Bhat, 2006)

San Francisco transportation data

Used to test context effect models:

Koppelman & Bhat (’06)
Benson et al. (WWW ’16)
Ragain & Ugander (NeurIPS ’16)
Seshadri et al. (ICML ’19)

Choice set confounding in real data

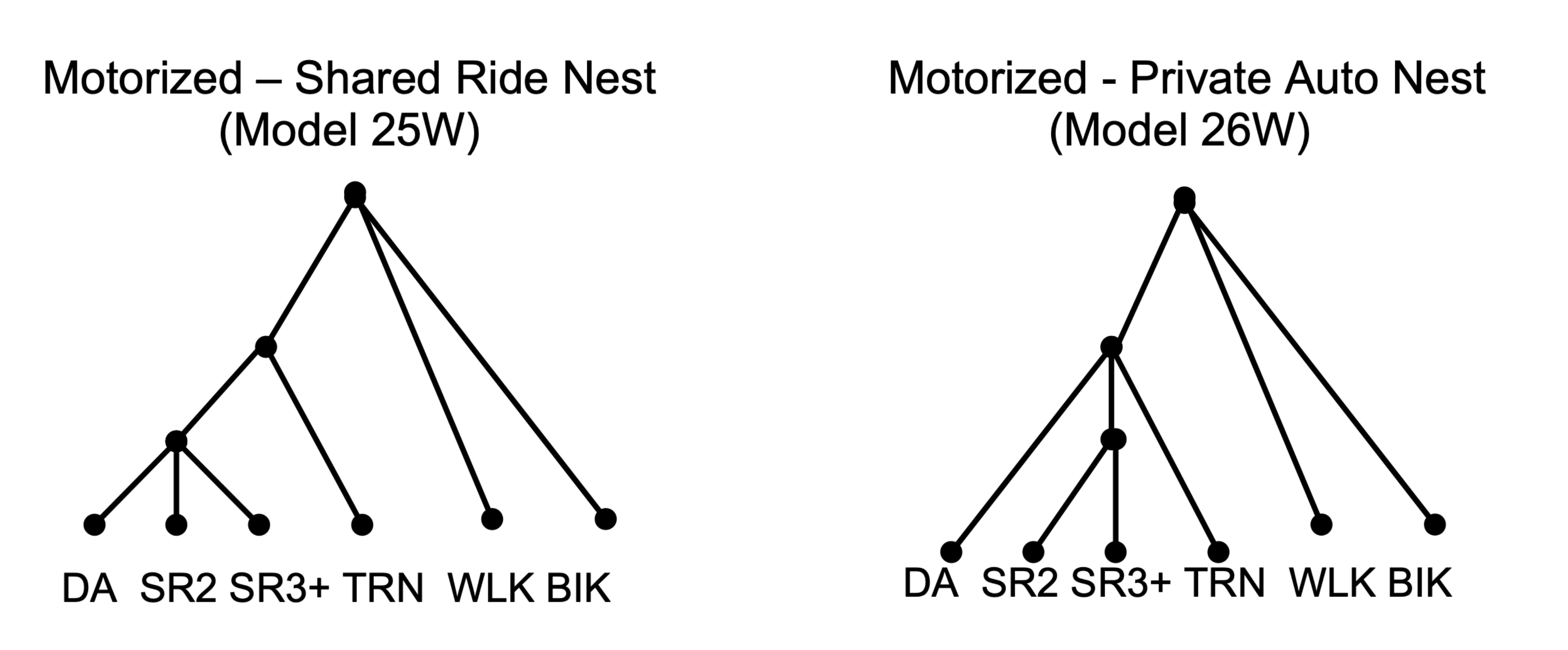
SFWork & SFShop

(Koppelman & Bhat, 2006)

San Francisco transportation data

Used to test context effect models:

Koppelman & Bhat ('06)
Benson et al. (WWW '16)
Ragain & Ugander (NeurIPS '16)
Seshadri et al. (ICML '19)



(Koppelman & Bhat, 2006)

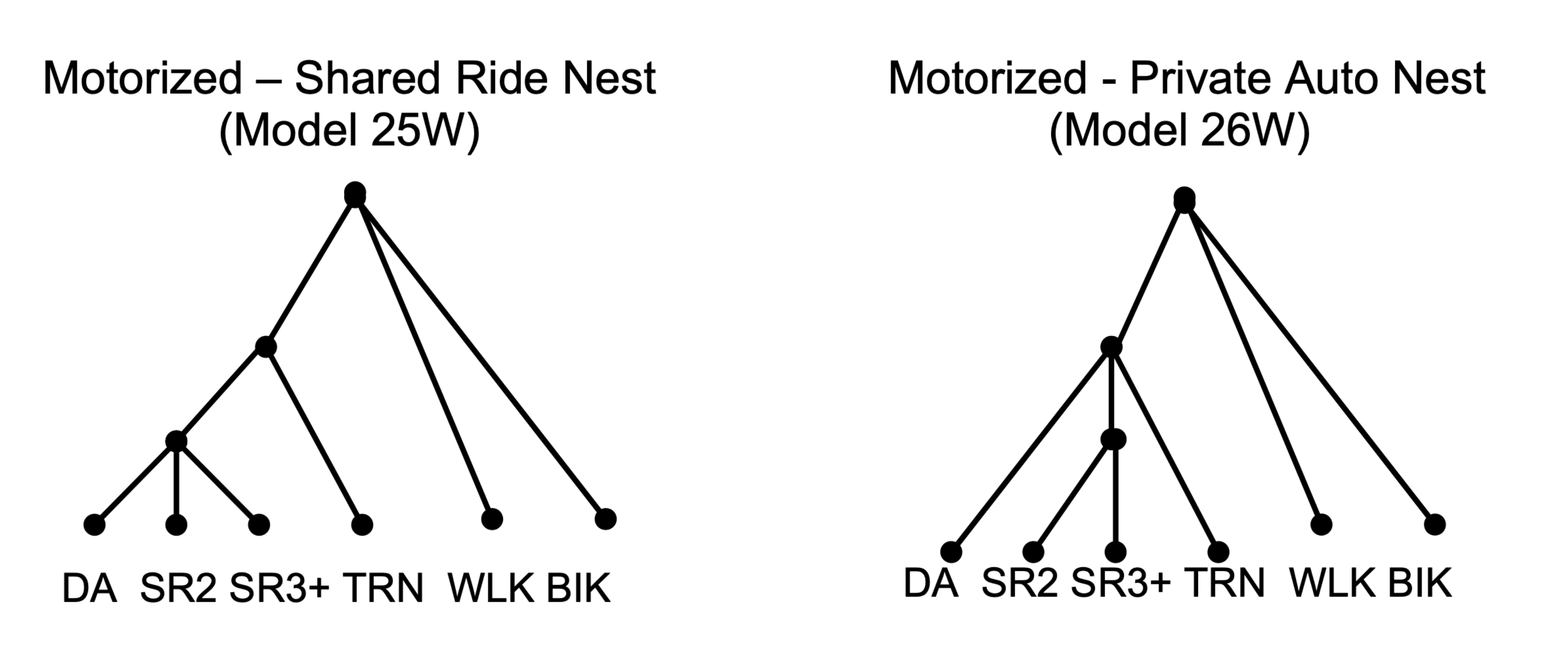
Choice set confounding in real data

SFWork & SFShop
(Koppelman & Bhat, 2006)

San Francisco transportation data

Used to test context effect models:

Koppelman & Bhat ('06)
Benson et al. (WWW '16)
Ragain & Ugander (NeurIPS '16)
Seshadri et al. (ICML '19)



(Koppelman & Bhat, 2006)

Has regularity violations!

SF-WORK

Choice set (C)	Pr(DA | C)	N
{DA, SR 2, SR 3+, Transit}	0.72	1661
{DA, SR 2, SR 3+, Transit, Bike}	0.83	829

Choice set confounding in real data

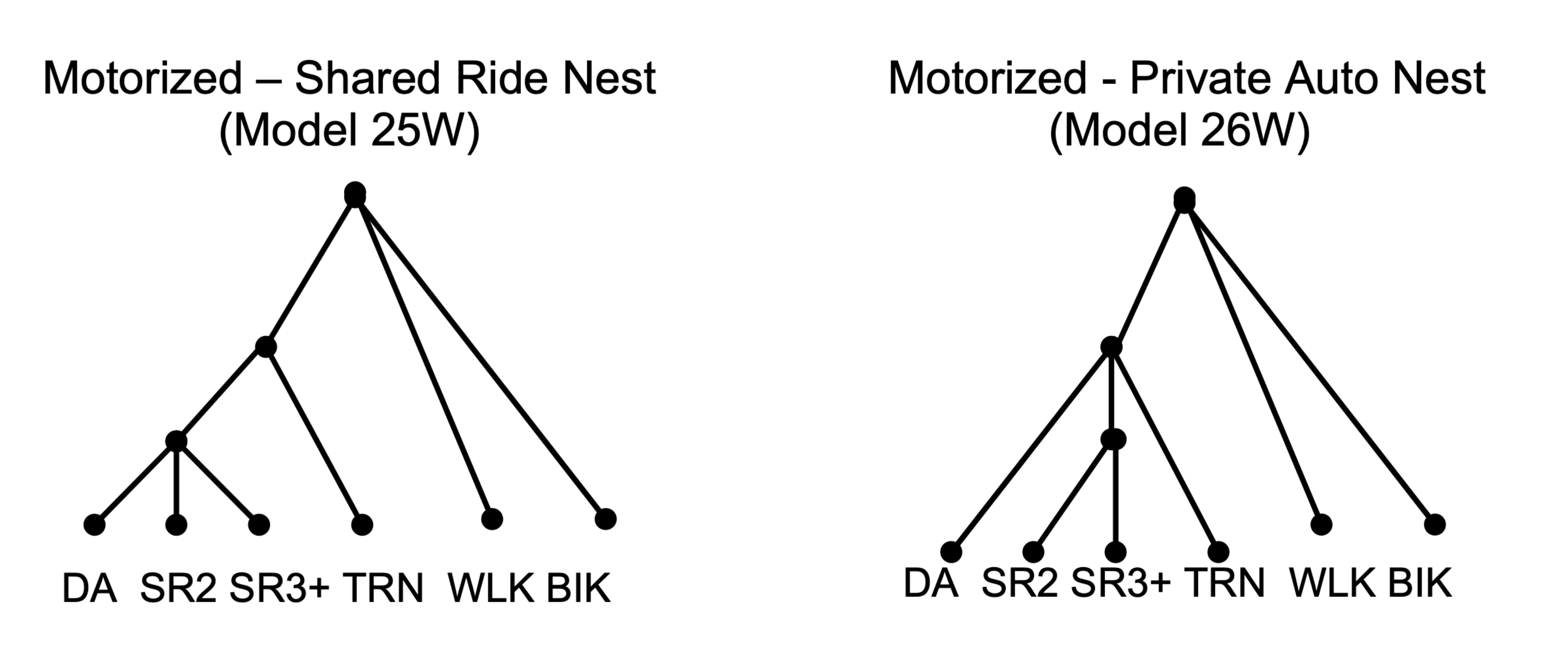
SFWork & SFShop

(Koppelman & Bhat, 2006)

San Francisco transportation data

Used to test context effect models:

Koppelman & Bhat ('06)
Benson et al. (WWW '16)
Ragain & Ugander (NeurIPS '16)
Seshadri et al. (ICML '19)



(Koppelman & Bhat, 2006)

Has regularity violations!

SF-WORK

Choice set (C)	Pr(DA | C)	N
{DA, SR 2, SR 3+, Transit}	0.72	1661
{DA, SR 2, SR 3+, Transit, Bike}	0.83	829

Context effects
or
choice set confounding?

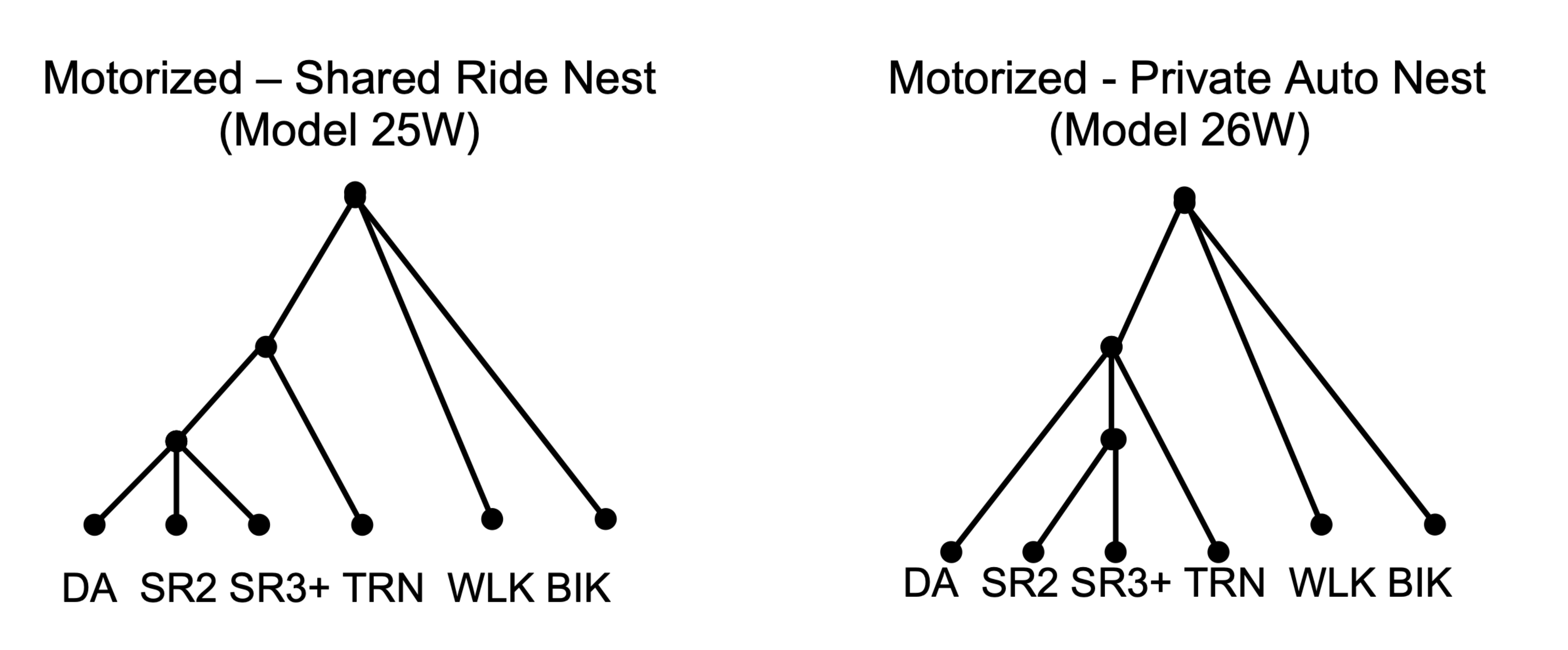
Choice set confounding in real data

SFWork & SFShop
(Koppelman & Bhat, 2006)

San Francisco transportation data

Used to test context effect models:

Koppelman & Bhat ('06)
Benson et al. (WWW '16)
Ragain & Ugander (NeurIPS '16)
Seshadri et al. (ICML '19)



(Koppelman & Bhat, 2006)

Has regularity violations!

SF-WORK Choice set (C)	Pr(DA | C)	N
{DA, SR 2, SR 3+, Transit}	0.72	1661
{DA, SR 2, SR 3+, Transit, Bike}	0.83	829

Context effects
or
choice set confounding?

Comparison	Testing	Controlling	$\Delta\ell$	LRT p
SF-WORK				
Logit to MNL	covariates	—	883	$< 10^{-10}$
Logit to CDM	context	—	85	$< 10^{-10}$
CDM to MCDM	covariates	context	819	$< 10^{-10}$
MNL to MCDM	context	covariates	20	0.08
SF-SHOP				
Logit to MNL	covariates	—	343	$< 10^{-10}$
Logit to CDM	context	—	96	$< 10^{-10}$
CDM to MCDM	covariates	context	276	$< 10^{-10}$
MNL to MCDM	context	covariates	29	0.36

CDM: context effect model (Seshadri et al, 2019)

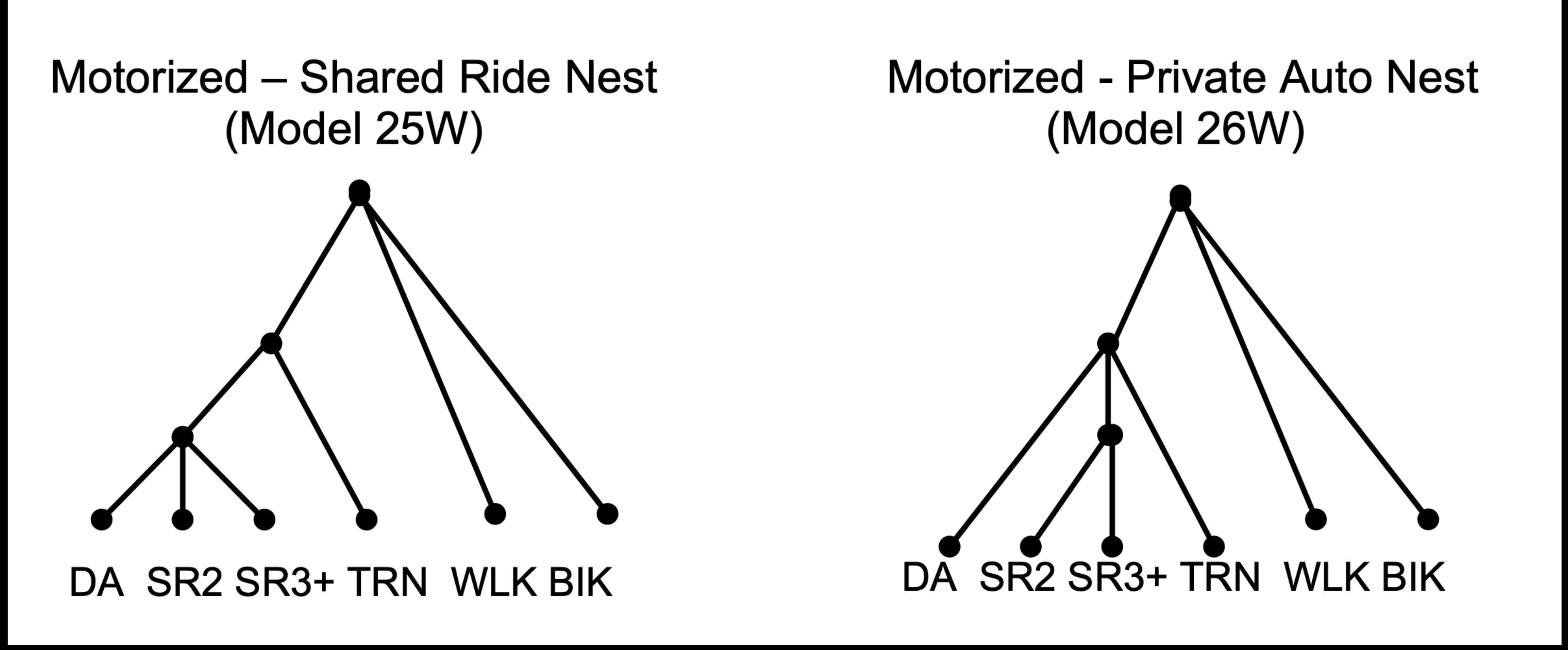
Choice set confounding in real data

SFWork & SFShop
(Koppelman & Bhat, 2006)

San Francisco transportation data

Used to test context effect models:

Koppelman & Bhat ('06)
Benson et al. (WWW '16)
Ragain & Ugander (NeurIPS '16)
Seshadri et al. (ICML '19)



(Koppelman & Bhat, 2006)

Has regularity violations!

SF-WORK

Choice set (C)	Pr(DA | C)	N
{DA, SR 2, SR 3+, Transit}	0.72	1661
{DA, SR 2, SR 3+, Transit, Bike}	0.83	829

Context effects
or
★ **choice set confounding**?

Comparison	Testing	Controlling	$\Delta\ell$	LRT p
SF-WORK				
Logit to MNL	covariates	—	883	$< 10^{-10}$
Logit to CDM	context	—	85	$< 10^{-10}$
CDM to MCDM	covariates	context	819	$< 10^{-10}$
MNL to MCDM	context	covariates	20	0.08
SF-SHOP				
Logit to MNL	covariates	—	343	$< 10^{-10}$
Logit to CDM	context	—	96	$< 10^{-10}$
CDM to MCDM	covariates	context	276	$< 10^{-10}$
MNL to MCDM	context	covariates	29	0.36

CDM: context effect model (Seshadri et al, 2019)

This is a *causal inference* problem

Causal inference methods

Option 1: *inverse probability weighting (IPW)*

Option 1: *inverse probability weighting (IPW)*

Idea: rebalance data so that we have
chooser-independent choice sets: $\Pr(C) = \Pr(C \mid a)$

Option 1: *inverse probability weighting (IPW)*

Idea: rebalance data so that we have *chooser-independent choice sets*: $\Pr(C) = \Pr(C \mid a)$

learn model from reweighed log-likelihood:

$$\ell(\theta; \mathscr{D}) = \sum_{(i,C,a)\in\mathscr{D}} \log \Pr_\theta(i \mid C)$$

Option 1: *inverse probability weighting (IPW)*

Idea: rebalance data so that we have
chooser-independent choice sets: $\Pr(C) = \Pr(C \mid a)$

learn model from reweighed log-likelihood:

$$\ell(\theta; \mathcal{D}) = \sum_{(i,C,a)\in\mathcal{D}} \log \Pr_\theta(i \mid C) \longrightarrow \ell(\theta; \tilde{\mathcal{D}}) = \sum_{(i,C,a)\in\mathcal{D}} \frac{\log \Pr_\theta(i \mid C)}{\Pr\left(C \mid x_a\right)}$$

Option 1: *inverse probability weighting (IPW)*

Idea: rebalance data so that we have
chooser-independent choice sets: $\Pr(C) = \Pr(C \mid a)$

learn model from reweighed log-likelihood:

$$\ell(\theta; \mathcal{D}) = \sum_{(i,C,a)\in\mathcal{D}} \log \Pr_\theta(i \mid C) \longrightarrow \ell(\theta; \tilde{\mathcal{D}}) = \sum_{(i,C,a)\in\mathcal{D}} \frac{\log \Pr_\theta(i \mid C)}{\Pr(C \mid x_a)}$$

choice set propensities

Need to learn $\Pr(C \mid x_a)$

(x_a: *covariates* for chooser a)

Option 1: *inverse probability weighting (IPW)*

Idea: rebalance data so that we have
chooser-independent choice sets: $\Pr(C) = \Pr(C \mid a)$

learn model from reweighed log-likelihood:

$$\ell(\theta; \mathcal{D}) = \sum_{(i,C,a)\in\mathcal{D}} \log \Pr_\theta(i \mid C) \longrightarrow \ell(\theta; \tilde{\mathcal{D}}) = \sum_{(i,C,a)\in\mathcal{D}} \frac{\log \Pr_\theta(i \mid C)}{\Pr(C \mid x_a)}$$

choice set propensities

requires assumption on x_a: *choice set ignorability*

Need to learn $\Pr(C \mid x_a)$

(x_a: *covariates* for chooser a)

Option 1: *inverse probability weighting (IPW)*

Idea: rebalance data so that we have
chooser-independent choice sets: $\Pr(C) = \Pr(C \mid a)$

learn model from reweighed log-likelihood:

$$\ell(\theta; \mathscr{D}) = \sum_{(i,C,a)\in\mathscr{D}} \log \Pr_\theta(i \mid C) \longrightarrow \ell(\theta; \tilde{\mathscr{D}}) = \sum_{(i,C,a)\in\mathscr{D}} \frac{\log \Pr_\theta(i \mid C)}{\Pr\left(C \mid x_a\right)}$$

↖ *choice set propensities*

requires assumption on x_a: *choice set ignorability*

= "covariates fully capture **choice set assignment**"

Need to learn $\Pr(C \mid x_a)$

(x_a: *covariates* for chooser a)

Option 1: *inverse probability weighting (IPW)*

Idea: rebalance data so that we have
chooser-independent choice sets: $\Pr(C) = \Pr(C \mid a)$

learn model from reweighed log-likelihood:

$$\ell(\theta; \mathscr{D}) = \sum_{(i,C,a)\in\mathscr{D}} \log \Pr_\theta(i \mid C) \longrightarrow \ell(\theta; \tilde{\mathscr{D}}) = \sum_{(i,C,a)\in\mathscr{D}} \frac{\log \Pr_\theta(i \mid C)}{\Pr(C \mid x_a)}$$

↖ *choice set propensities*

requires assumption on x_a: *choice set ignorability*

= "covariates fully capture **choice set assignment**"

Guarantee

Can learn a model as if choice sets were uniformly random

Need to learn $\Pr(C \mid x_a)$

(x_a: *covariates* for chooser a)

Option 2: *regression*

Option 2: *regression*

Idea: model preference variation

Option 2: *regression*

Idea: model preference variation

Logit:

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

Option 2: *regression*

Idea: model preference variation

Logit:

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

→

Multinomial logit (MNL):

$$\Pr(i \mid a, C) = \frac{\exp(u_i + \beta_i x_a)}{\sum_{j \in C} \exp(u_j + \beta_i x_a)}$$

Option 2: *regression*

Idea: model preference variation

Logit:

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

→ *Multinomial logit (MNL)*:

$$\Pr(i \mid a, C) = \frac{\exp(u_i + \beta_i x_a)}{\sum_{j \in C} \exp(u_j + \beta_i x_a)}$$

CDM → *Multinomial CDM (MCDM)*

Option 2: *regression*

Idea: model preference variation

Logit:

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

→ *Multinomial logit (MNL)*:

$$\Pr(i \mid a, C) = \frac{\exp(u_i + \beta_i x_a)}{\sum_{j \in C} \exp(u_j + \beta_i x_a)}$$

CDM → *Multinomial CDM (MCDM)*

requires assumption on x_a: *preference ignorability*

Option 2: *regression*

Idea: model preference variation

Logit:

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

→ *Multinomial logit (MNL)*:

$$\Pr(i \mid a, C) = \frac{\exp(u_i + \beta_i x_a)}{\sum_{j \in C} \exp(u_j + \beta_i x_a)}$$

CDM → *Multinomial CDM (MCDM)*

requires assumption on x_a: *preference ignorability*

= "covariates fully capture **preferences**"

Option 2: *regression*

Idea: model preference variation

Logit:

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

→

Multinomial logit (MNL):

$$\Pr(i \mid a, C) = \frac{\exp(u_i + \beta_i x_a)}{\sum_{j \in C} \exp(u_j + \beta_i x_a)}$$

CDM → *Multinomial CDM (MCDM)*

requires assumption on x_a: *preference ignorability*

= "covariates fully capture **preferences**"

Guarantee

If the model is correctly specified, can learn true choice probabilities

Option 2: *regression*

Idea: model preference variation

Logit:

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

→

Multinomial logit (MNL):

$$\Pr(i \mid a, C) = \frac{\exp(u_i + \beta_i x_a)}{\sum_{j \in C} \exp(u_j + \beta_i x_a)}$$

CDM → *Multinomial CDM (MCDM)*

requires assumption on x_a: *preference ignorability*

= "covariates fully capture **preferences**"

Guarantee

If the model is correctly specified, can learn true choice probabilities

Combine IPW and regression
→ *doubly robust*

Causal inference results

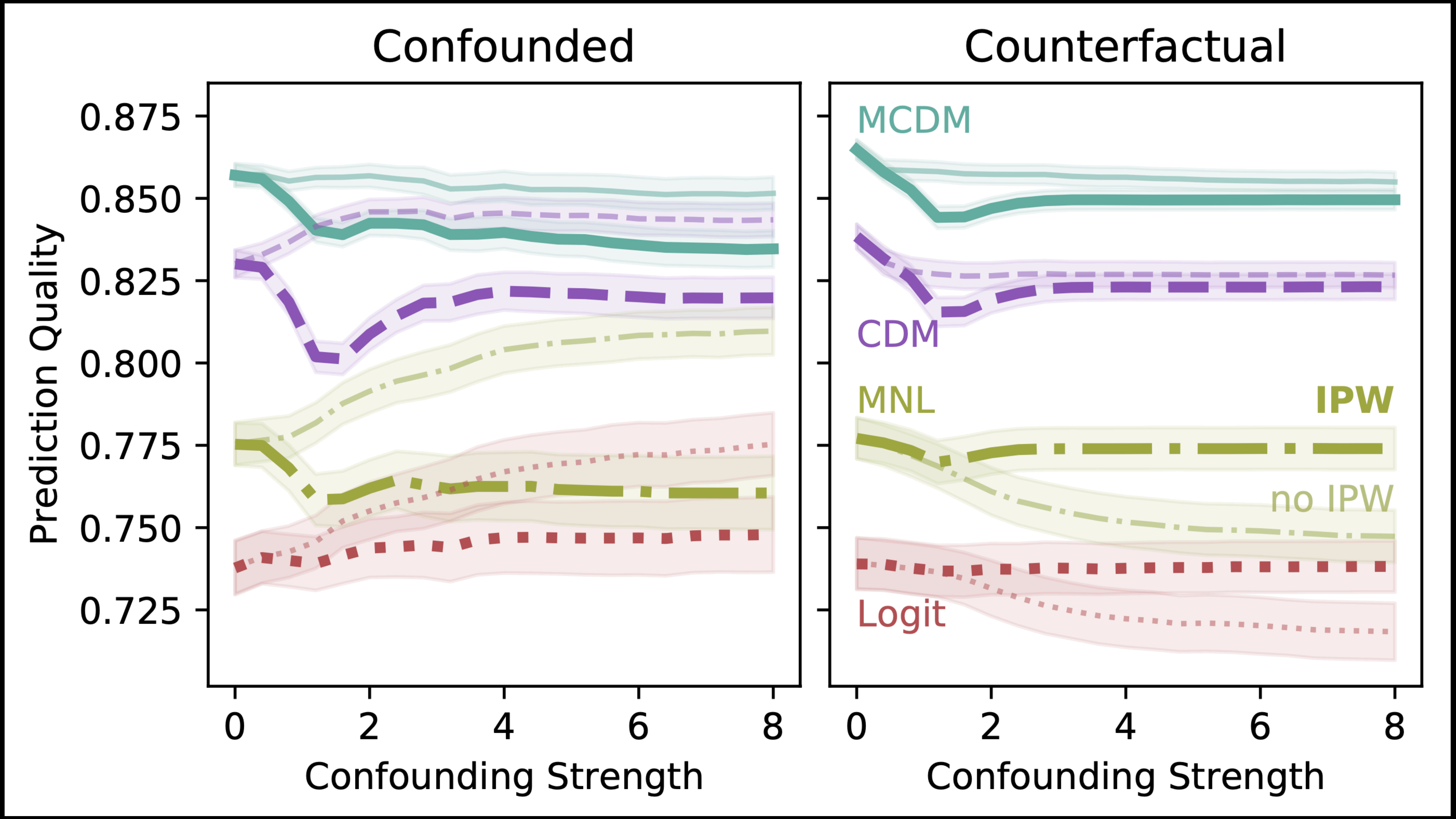
Synthetic CDM data

Synthetic CDM data

counterfactuals: new instances not drawn from data distribution

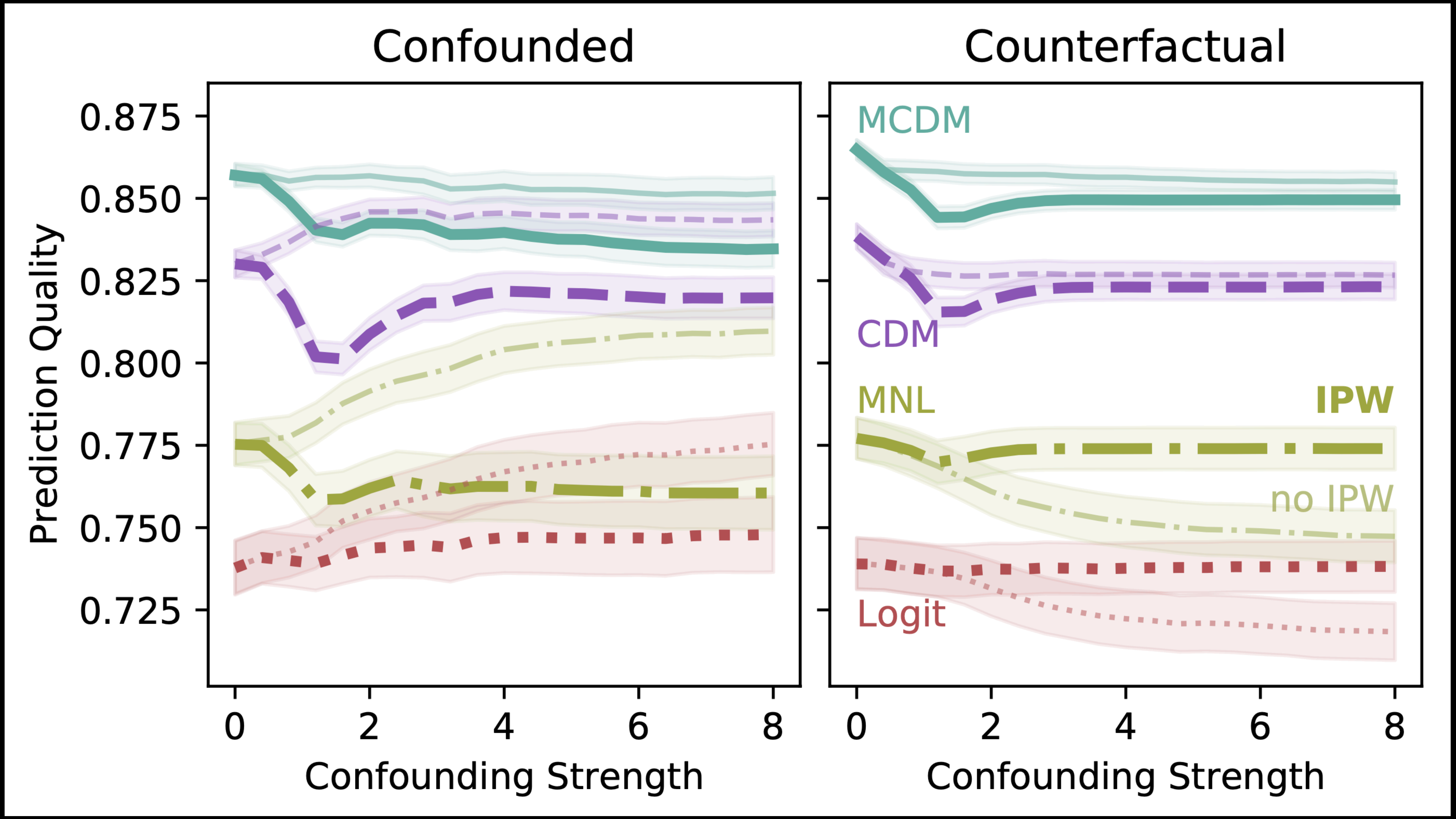
Synthetic CDM data

counterfactuals: new instances not drawn from data distribution



Synthetic CDM data

counterfactuals: new instances not drawn from data distribution



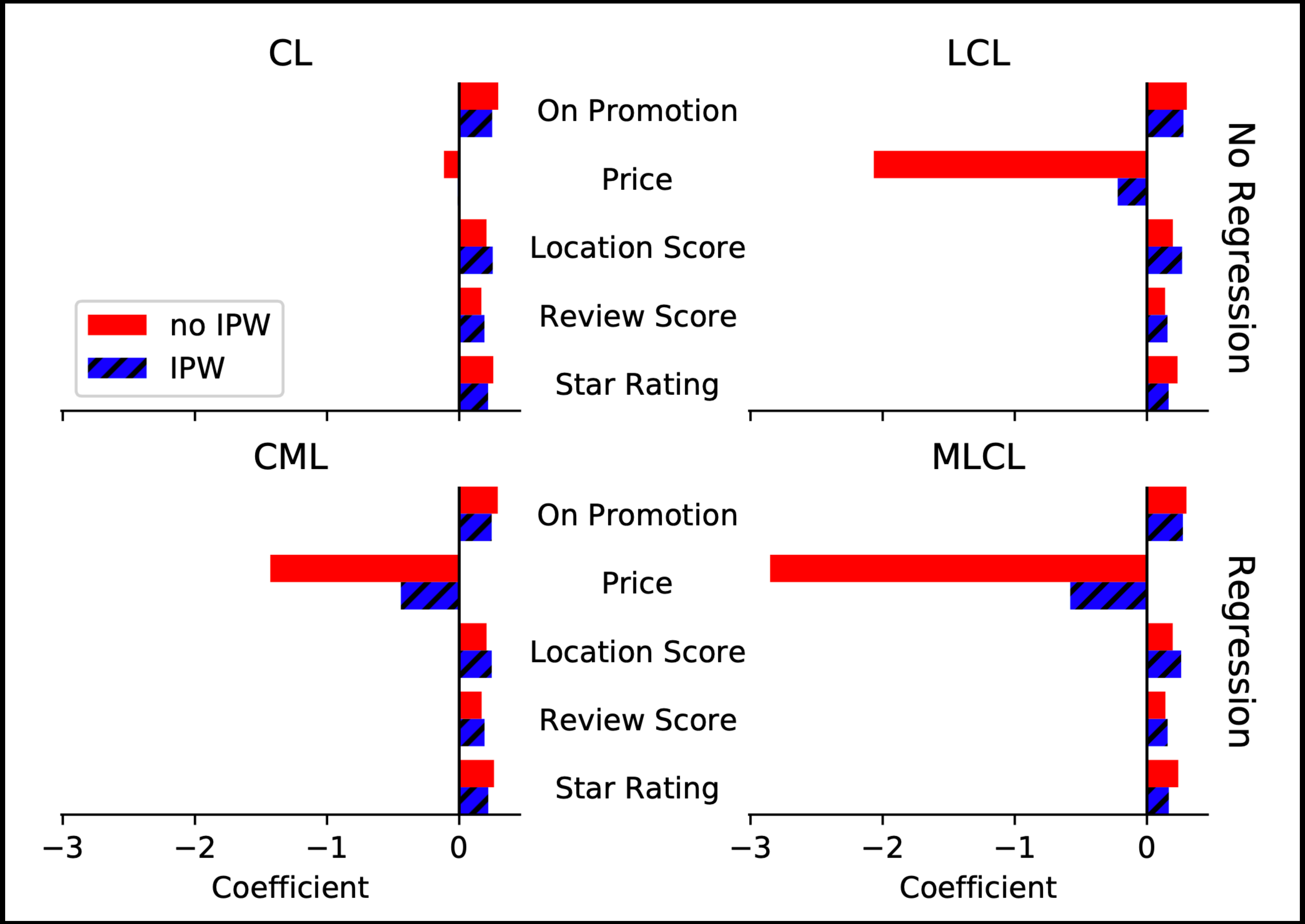
IPW & regression

(a) improve counterfactual prediction

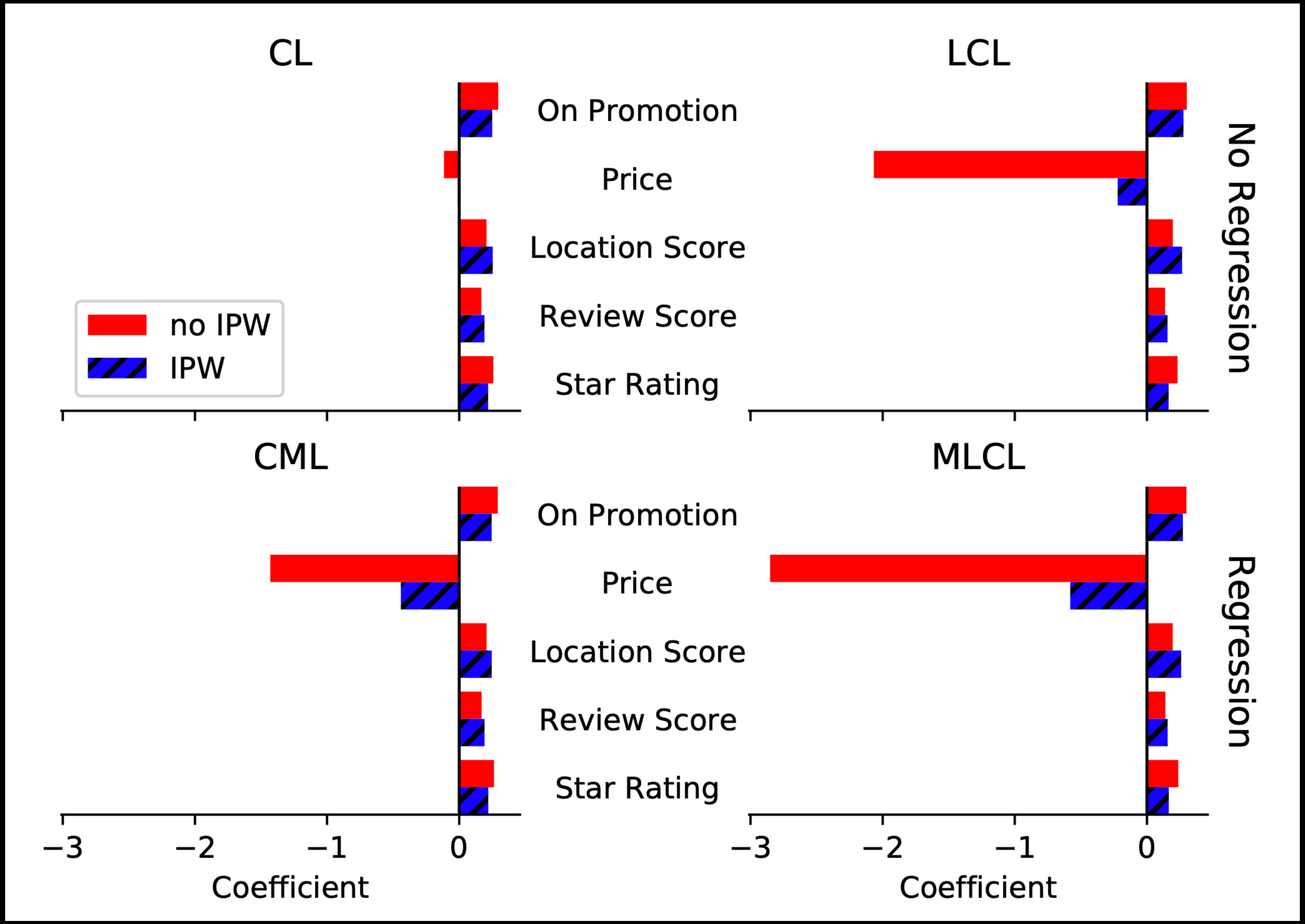
(b) prevent overconfidence on confounded data

Expedia hotel booking data

Expedia hotel booking data

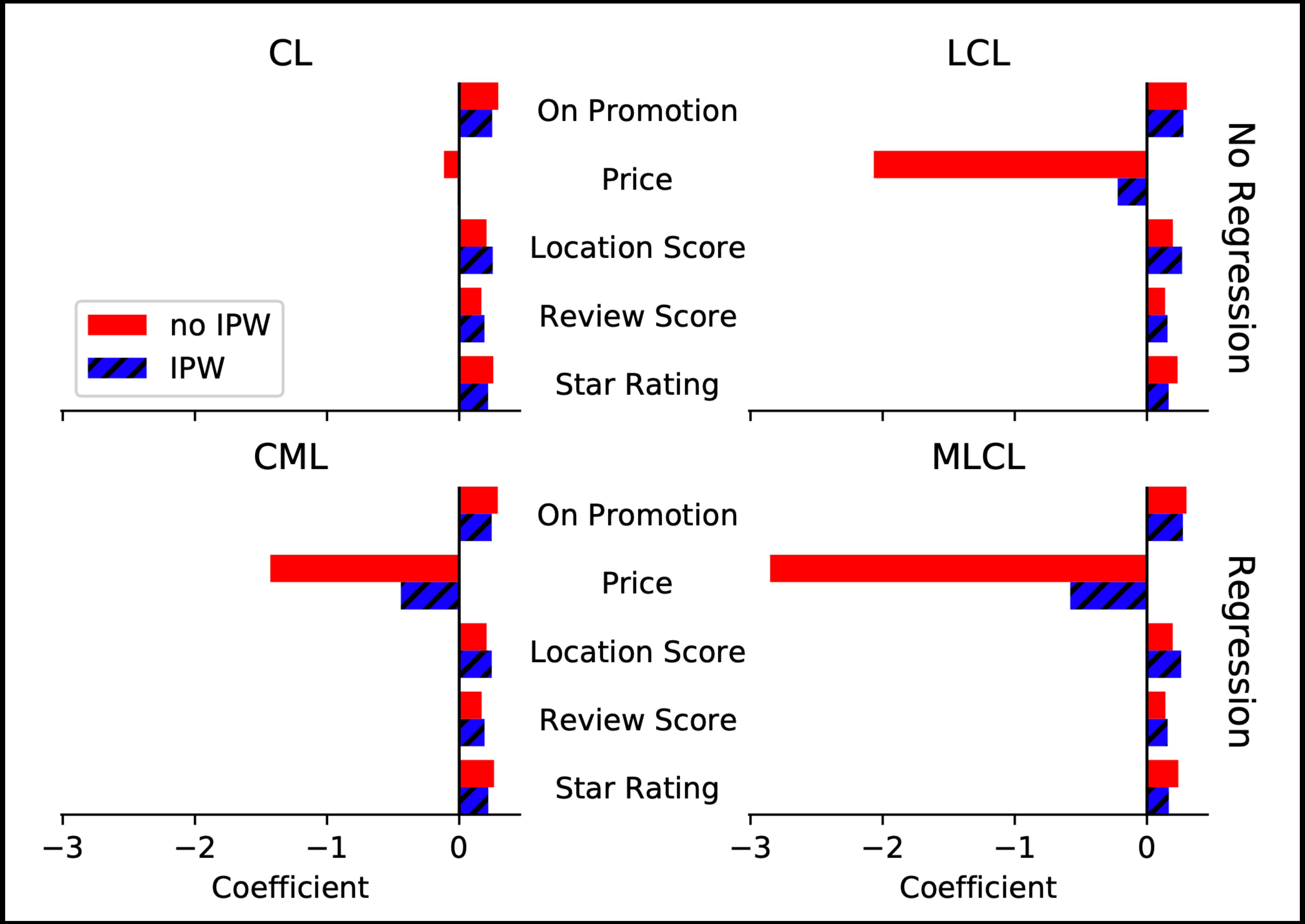


Expedia hotel booking data



Without IPW, importance of price is exaggerated

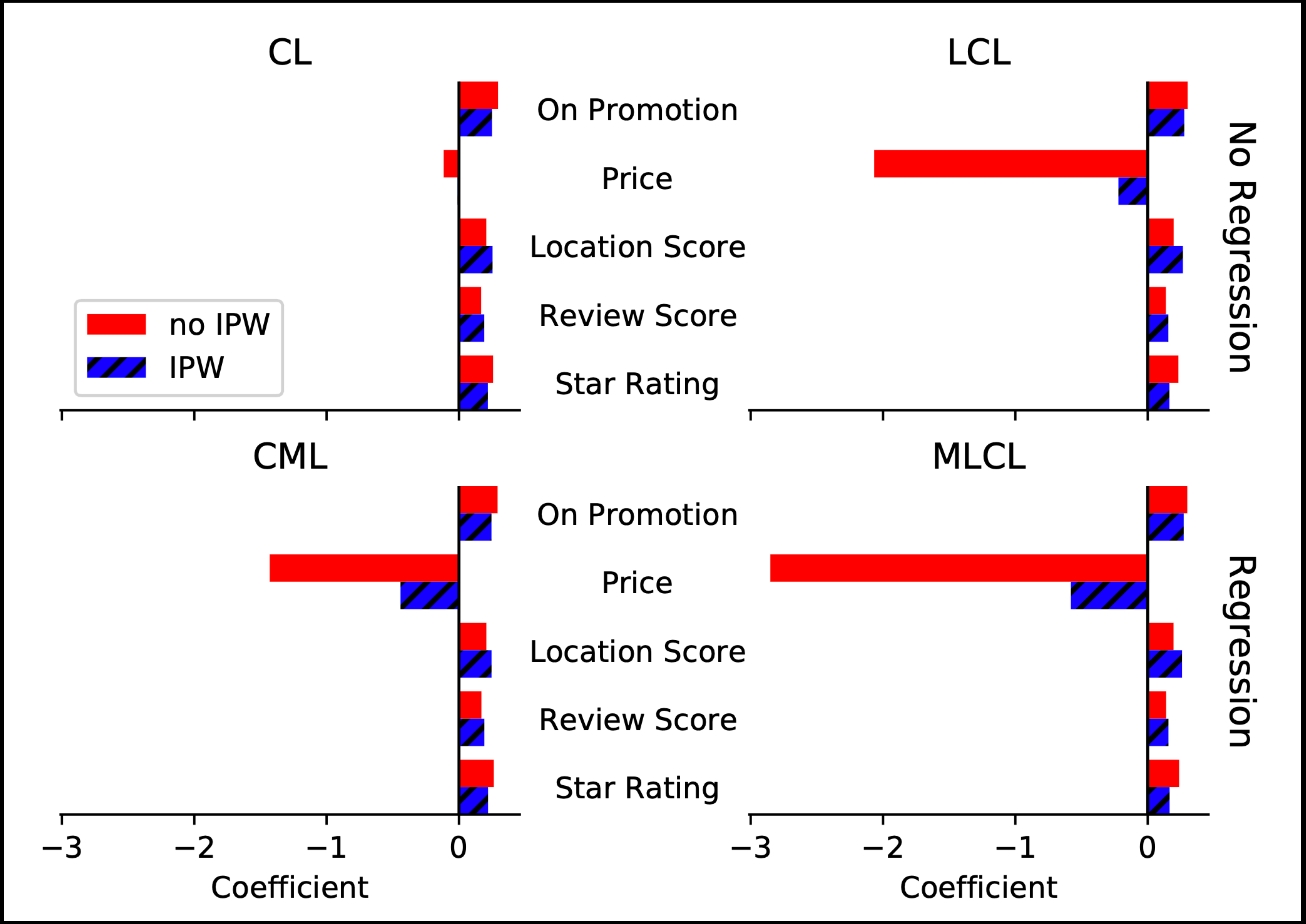
Expedia hotel booking data



Without IPW, importance of price is exaggerated

Expedia covariates more informative about choice sets than preferences
→ IPW > regression

Expedia hotel booking data



Without IPW, importance of price is exaggerated

Expedia covariates more informative about choice sets than preferences
→ IPW > regression

Dataset log-likelihood:

Model	Confounded	IPW-adjusted
CL	−839499	−786653
CML	−838281	−785753
LCL	−837154	−784770
MLCL	−835986	−783928

Managing without covariates

Clustering based on choice sets

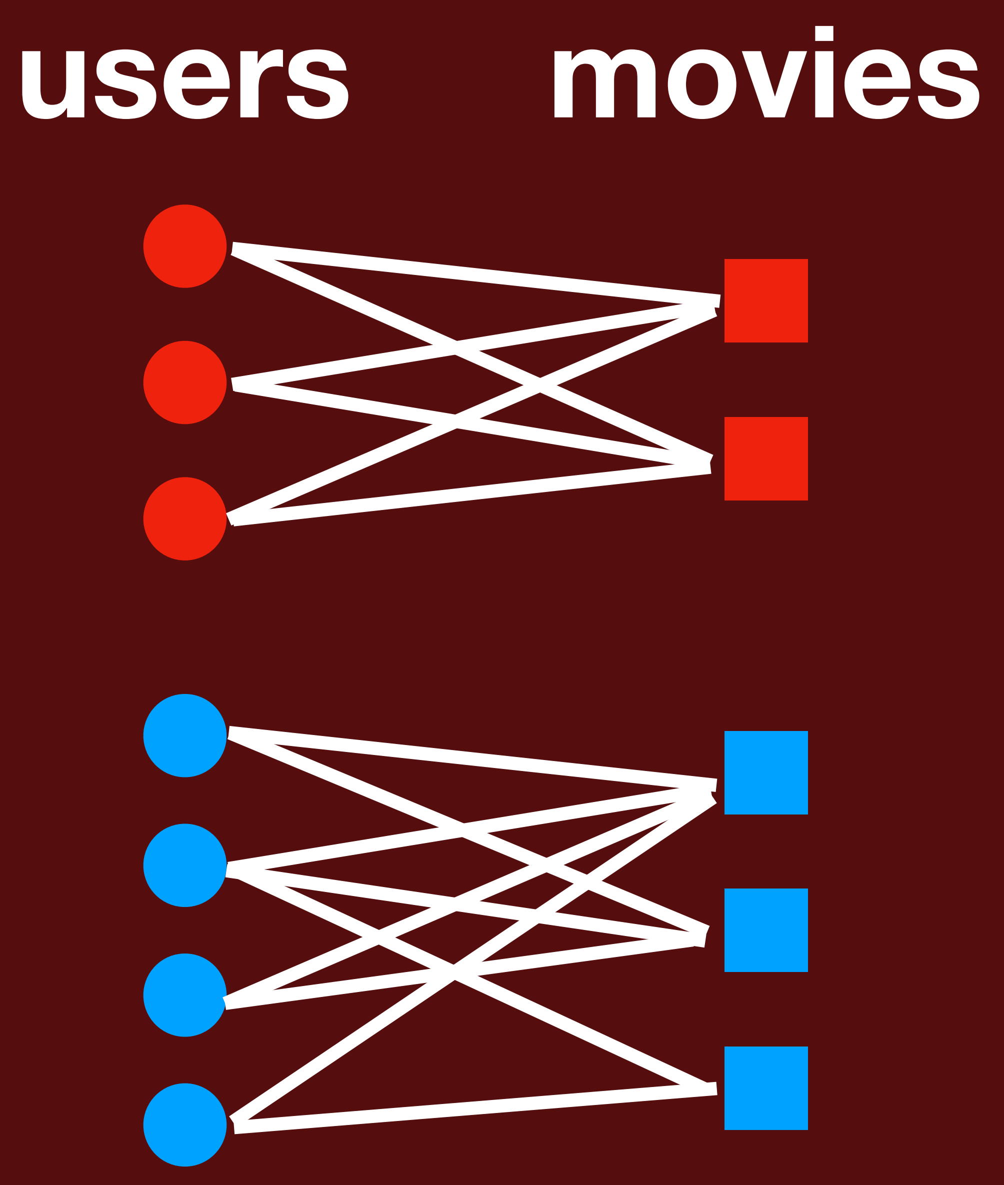
Clustering based on choice sets

Idea: take advantage of the correlation between choice sets and preferences

Clustering based on choice sets

Idea: take advantage of the correlation between choice sets and preferences

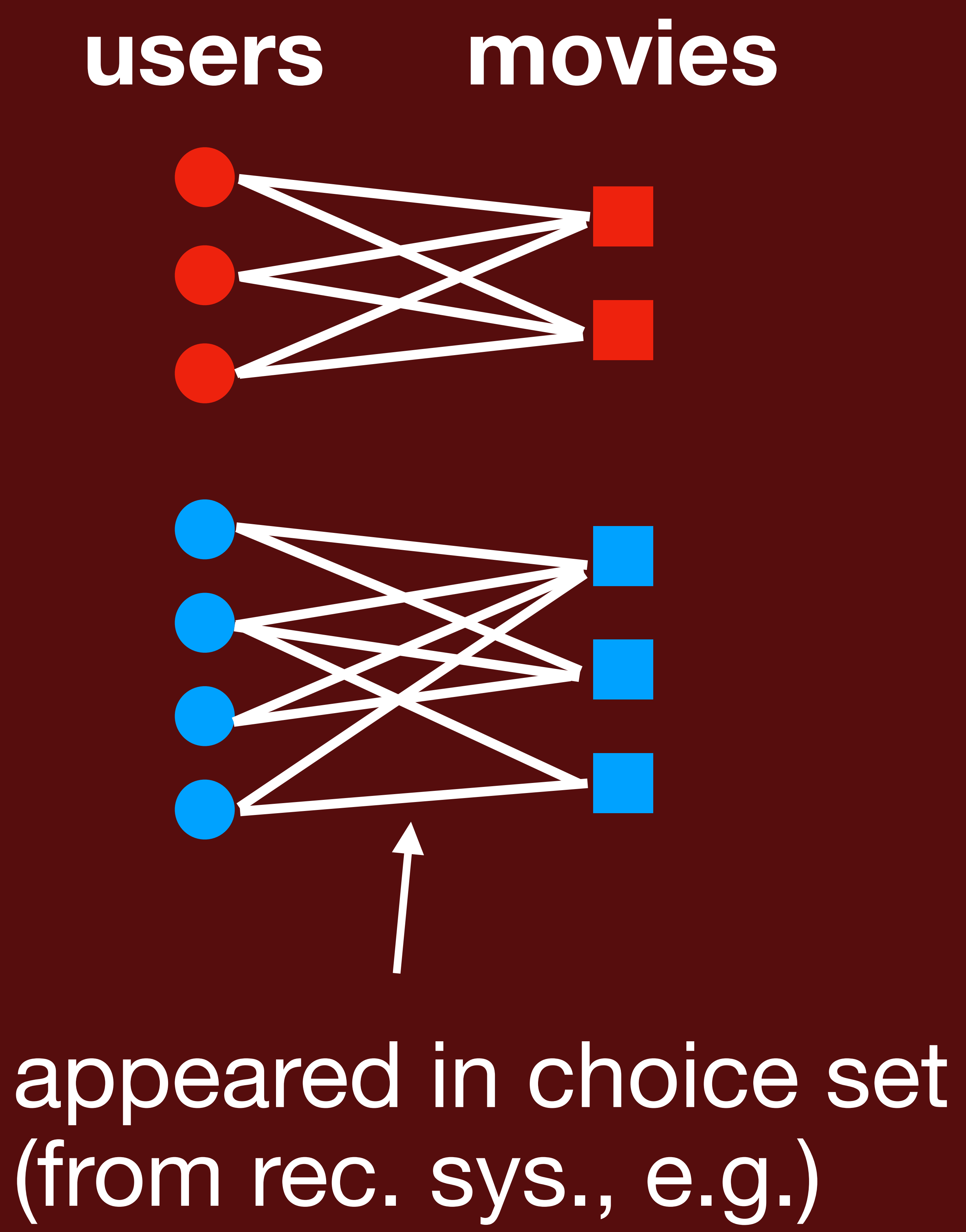
Example: movie recommendations



Clustering based on choice sets

Idea: take advantage of the correlation between choice sets and preferences

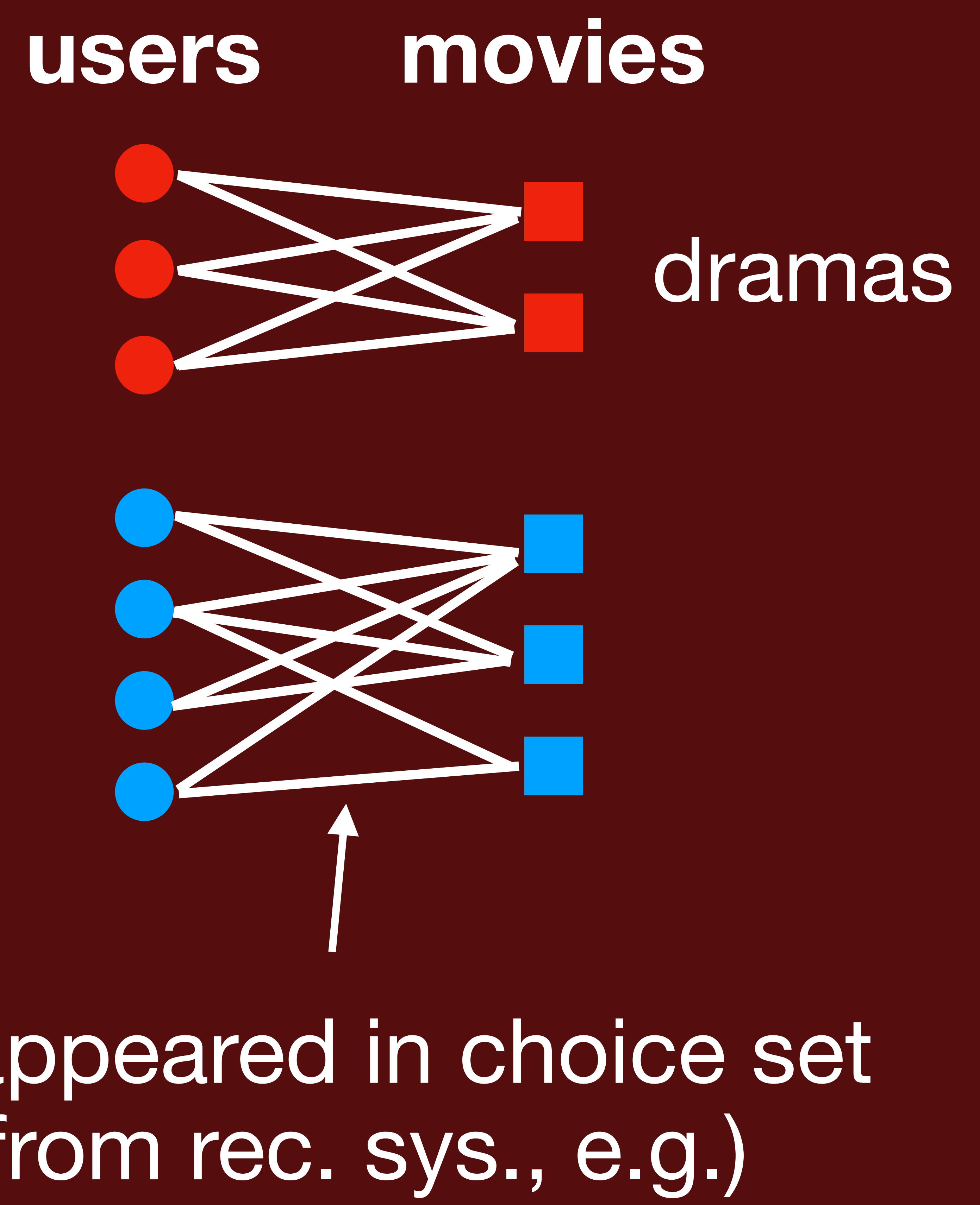
Example: movie recommendations



Clustering based on choice sets

Idea: take advantage of the correlation between choice sets and preferences

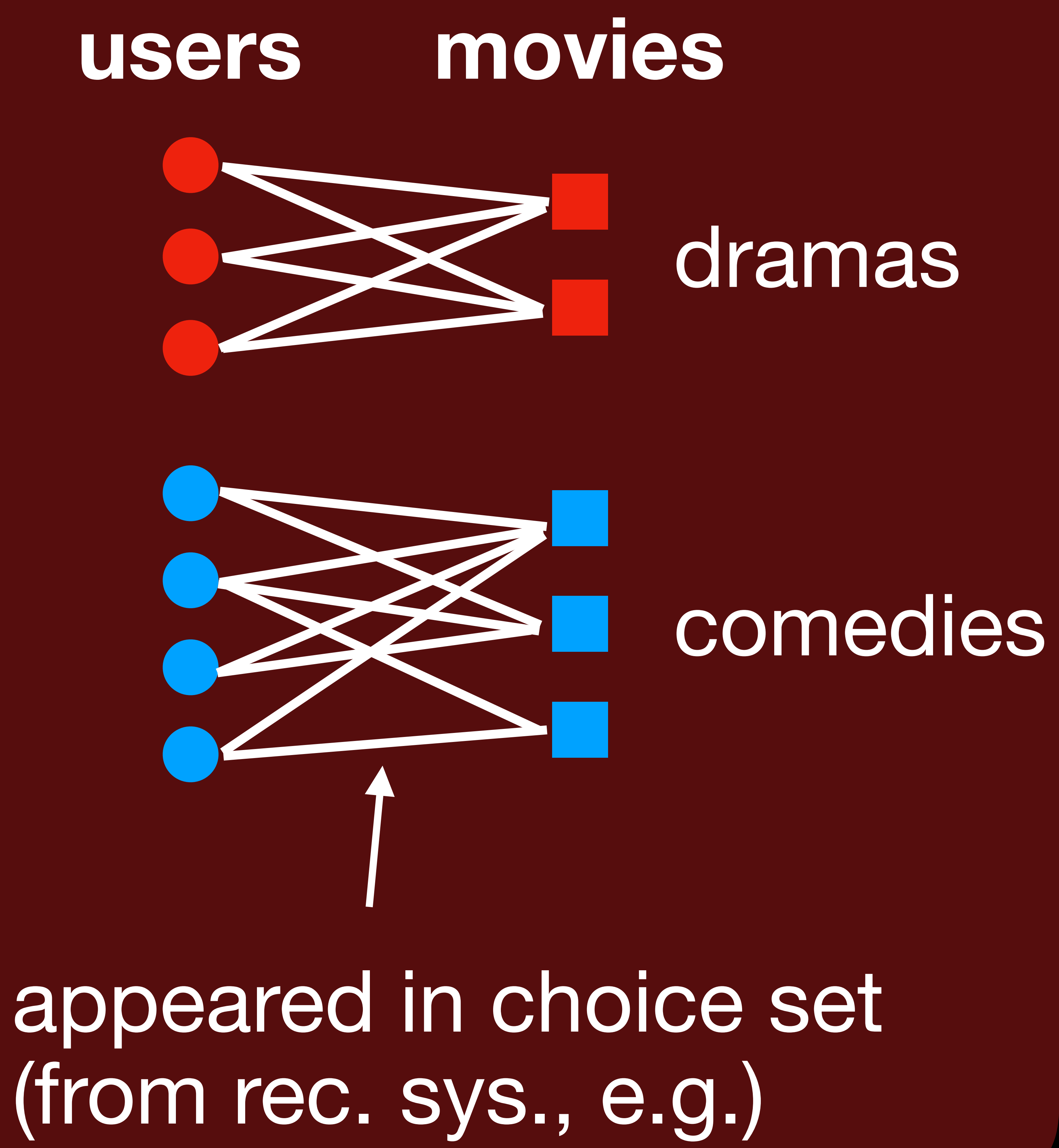
Example: movie recommendations



Clustering based on choice sets

Idea: take advantage of the correlation between choice sets and preferences

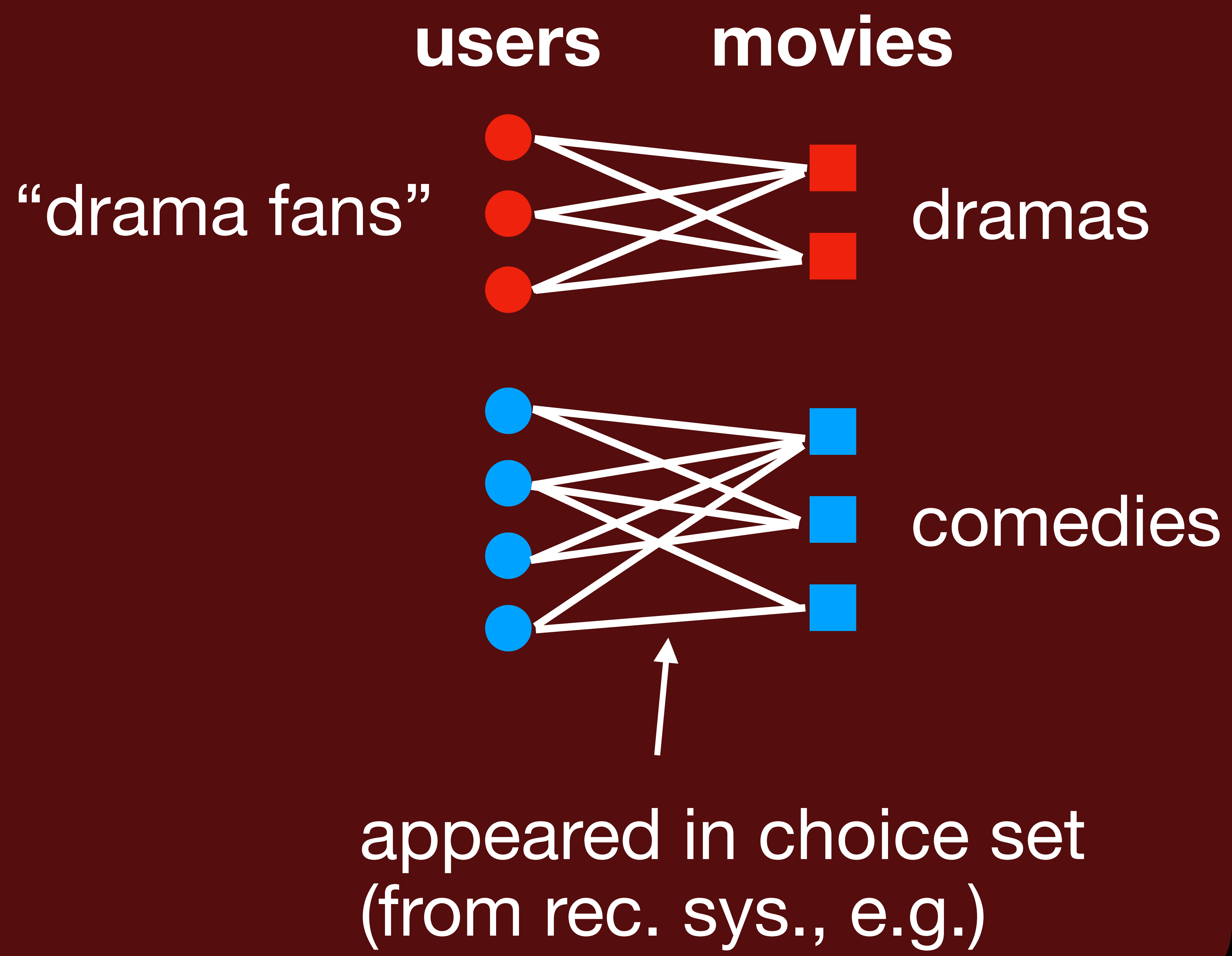
Example: movie recommendations



Clustering based on choice sets

Idea: take advantage of the correlation between choice sets and preferences

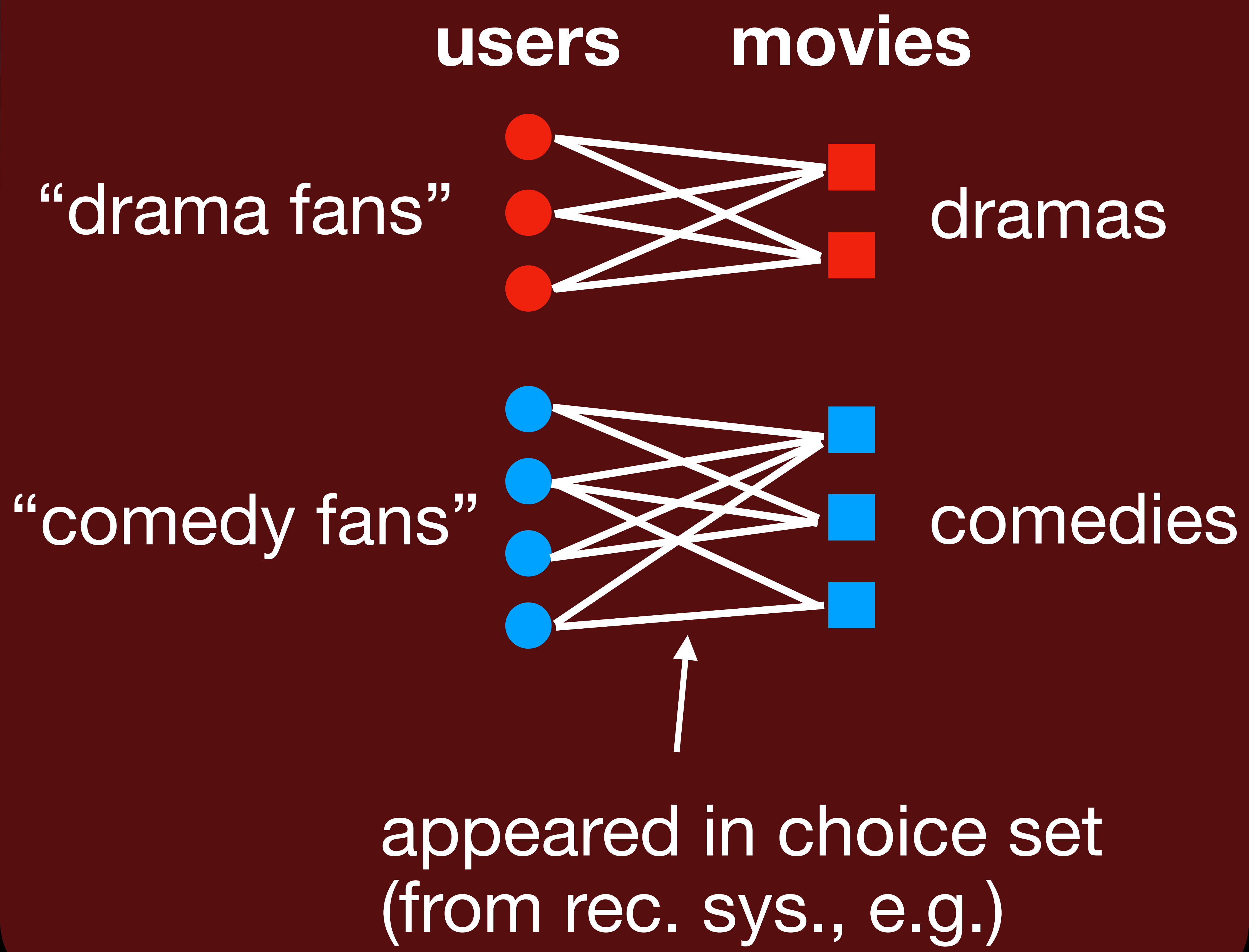
Example: movie recommendations



Clustering based on choice sets

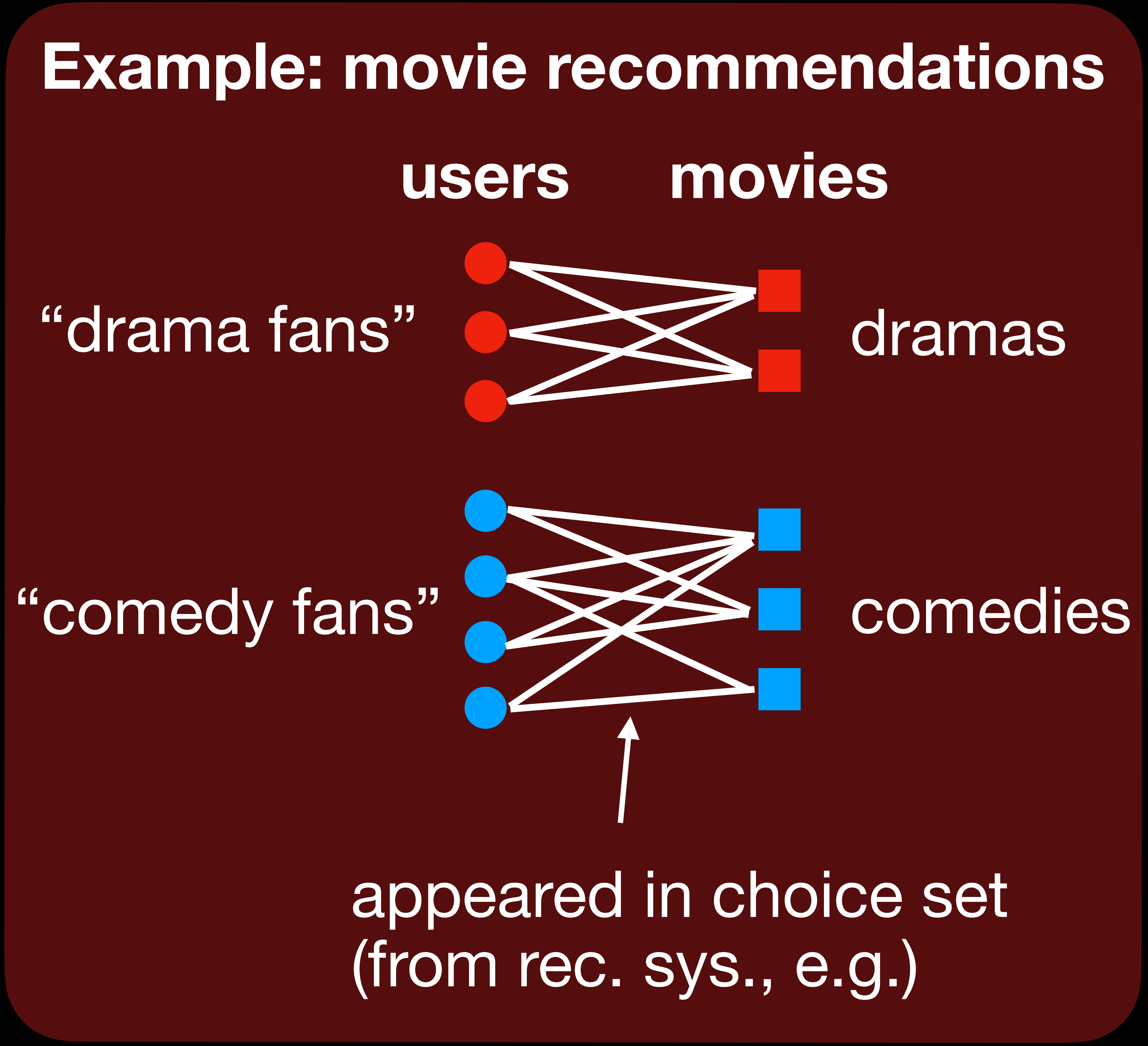
Idea: take advantage of the correlation between choice sets and preferences

Example: movie recommendations



Clustering based on choice sets

Idea: take advantage of the correlation between choice sets and preferences

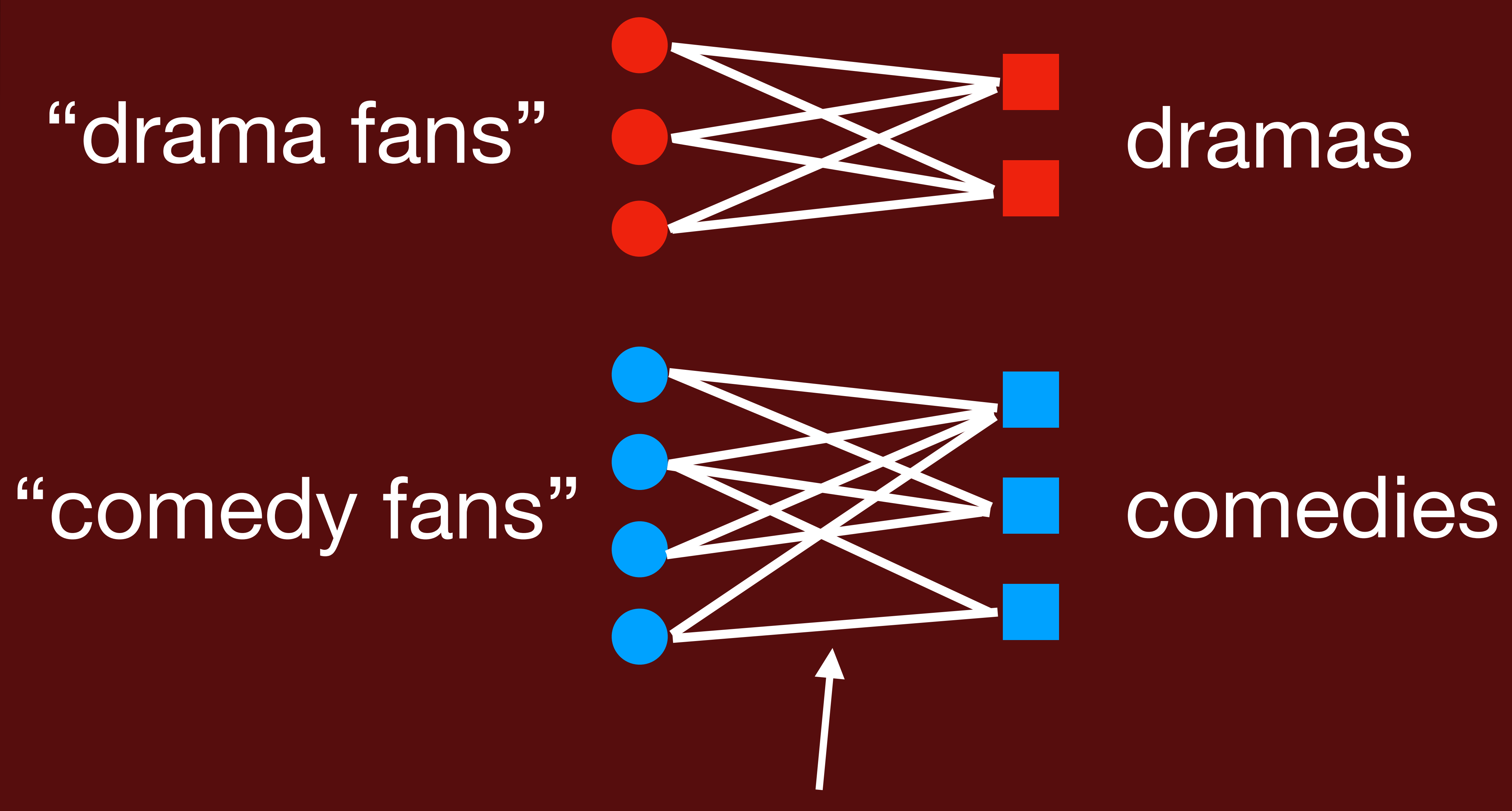


Cluster users (e.g., spectral co-clustering), learn choice model per-cluster (Dhillon, 2001)

Clustering based on choice sets

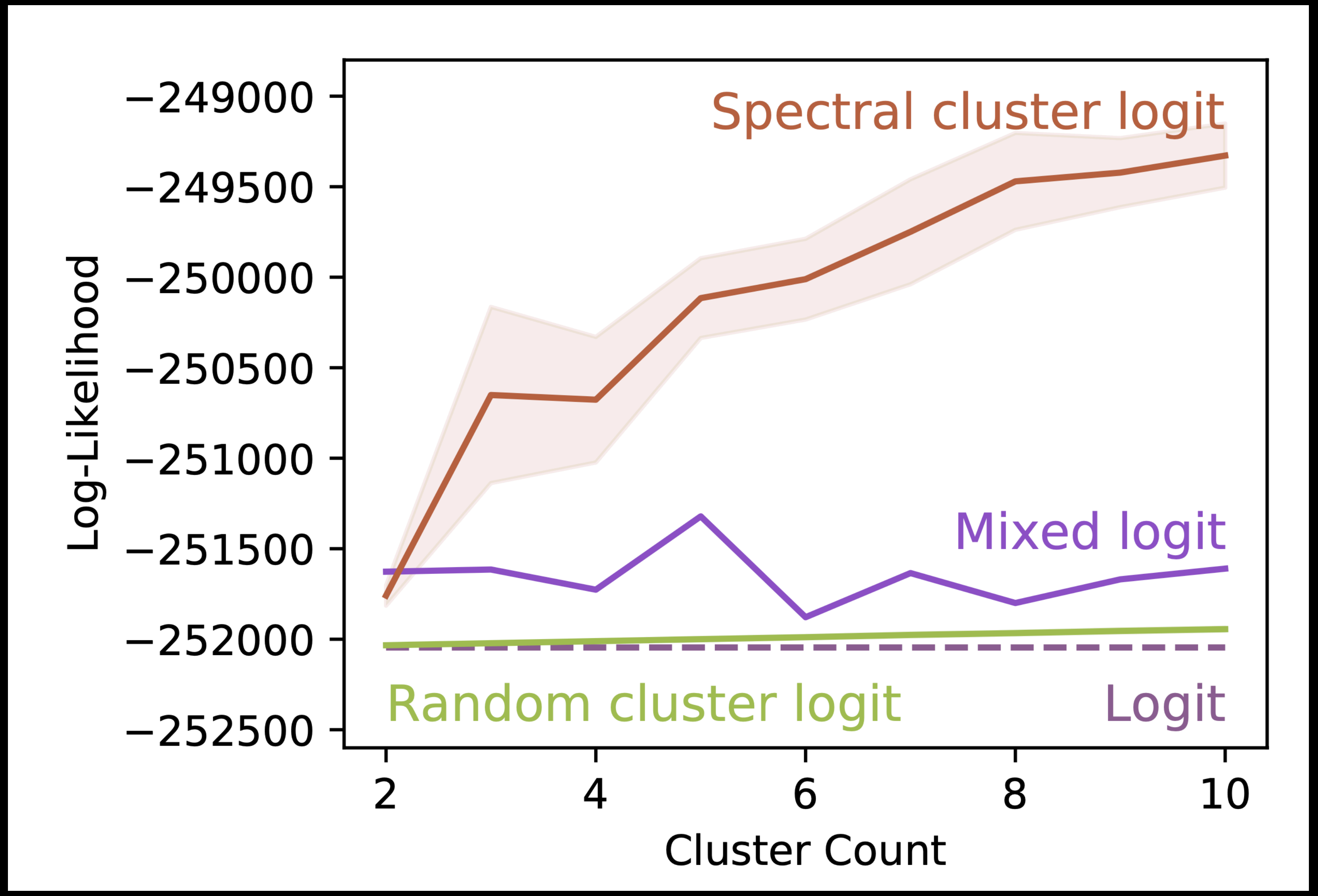
Idea: take advantage of the correlation between choice sets and preferences

Example: movie recommendations



Cluster users (e.g., spectral co-clustering), learn choice model per-cluster (Dhillon, 2001)

Much better than mixed logit!
(YOOCHOOSE online shopping data)



(RecSys, 2015)

More things in the paper

More things in the paper

The power of choice set confounding

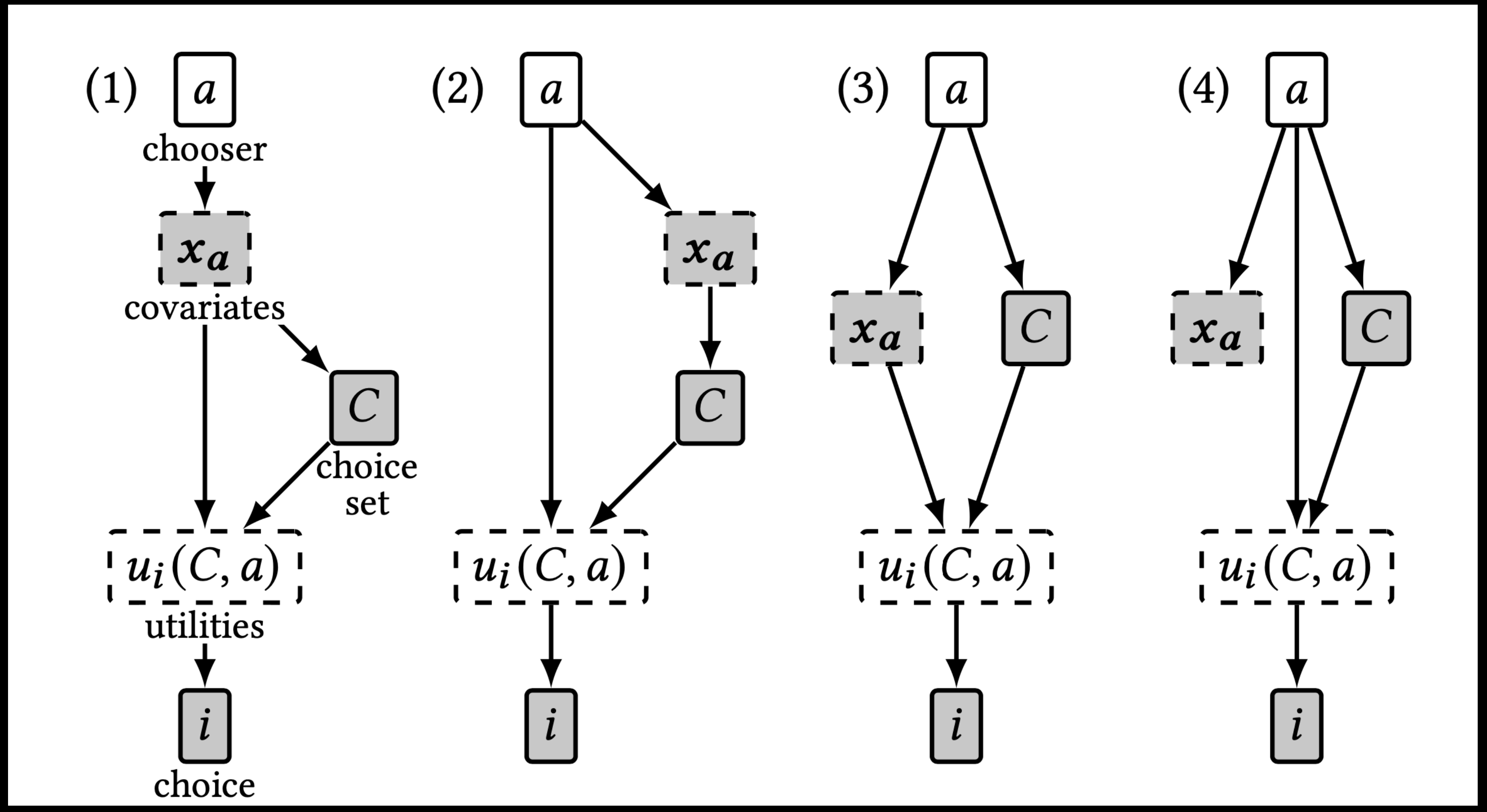
Theorem 2. *Mixed logit with chooser-dependent choice sets is powerful enough to express any system of choice probabilities.*

More things in the paper

The power of choice set confounding

> THEOREM 2. *Mixed logit with chooser-dependent choice sets is powerful enough to express any system of choice probabilities.*

Graphical intuition about ignorability assumptions

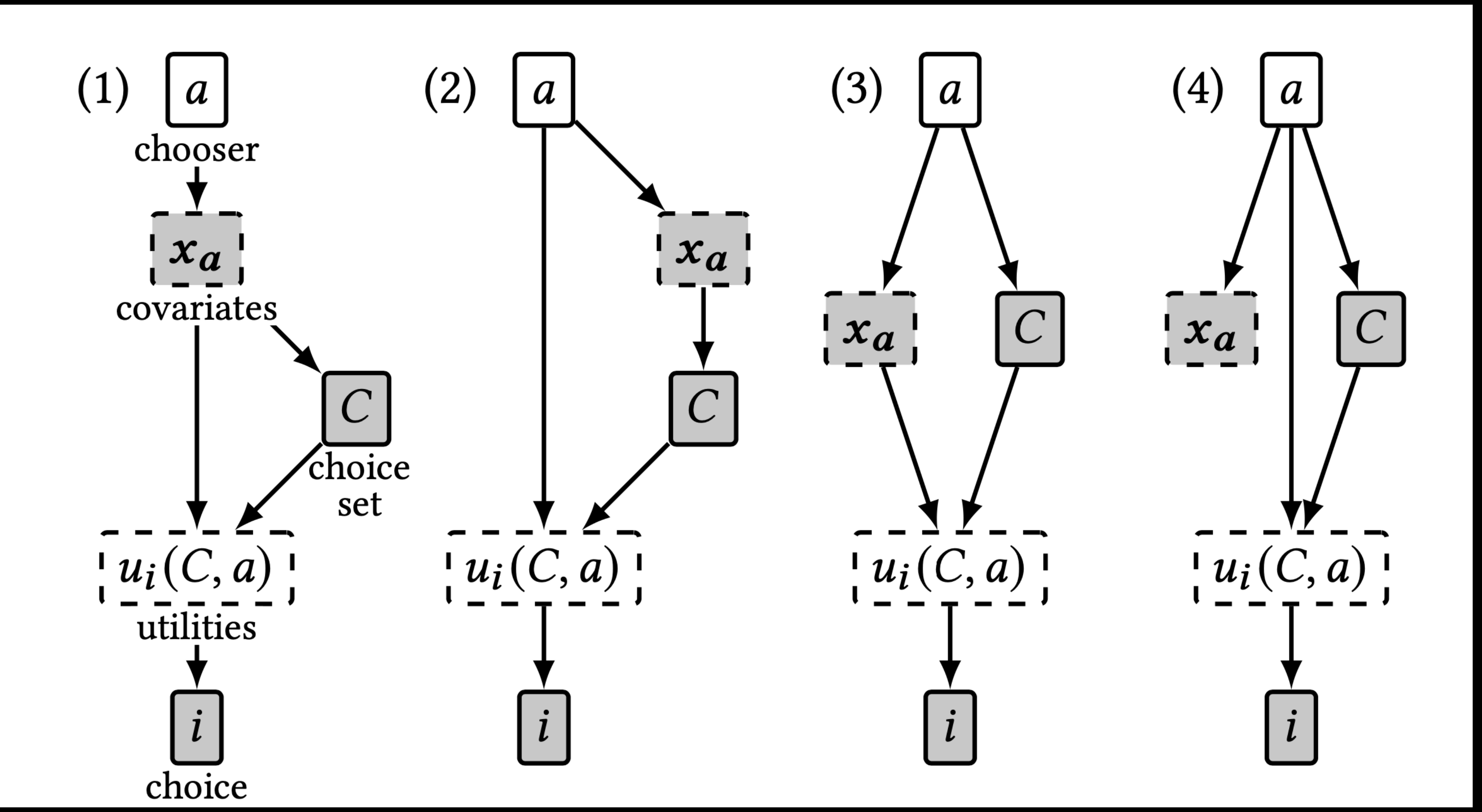


More things in the paper

The power of choice set confounding

THEOREM 2. *Mixed logit with chooser-dependent choice sets is powerful enough to express any system of choice probabilities.*

Graphical intuition about ignorability assumptions



Duality between context effect models and models of choice set confounding

Code: bit.ly/csc-kdd-code
Slides: bit.ly/csc-kdd-slides

Concluding thoughts

Key takeaways
Choice set confounding can mislead choice models
We can adjust for it using chooser covariates

Future work
Learning choice set propensities
Other causal inference methods:
- instrumental variables?
- matching?

Interested in context effect models?
See our other KDD ’21 paper:
“Learning Interpretable Feature Context Effects in Discrete Choice”

Submit to our NeurIPS ’21 workshop!
bit.ly/WHMD2021

Thank you!

More questions or ideas?
Email me: kt@cs.cornell.edu
@kiran_tomlinson

Acknowledgments

Funding from NSF, ARO, Koret Foundation, JP Morgan Chase