The Limit of Rationality in Choice Modeling

Srikanth Jagabathula
NYU Stern School of Business, New York

Joint with:
Paat Rusmevichientong
USC Marshall School of Business

Random Utility Max. (RUM) class: most popular discrete-choice model class used for demand predictions

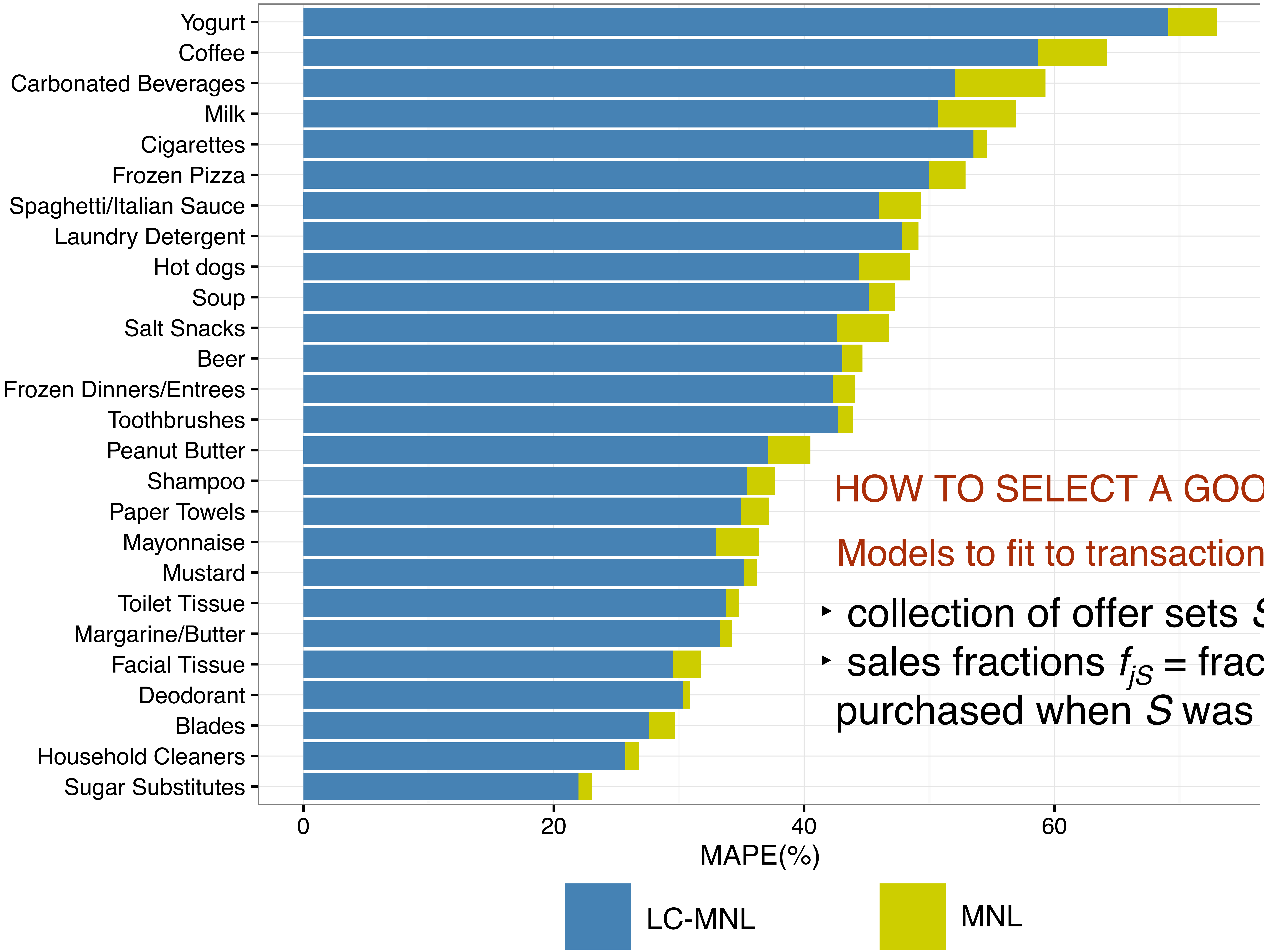
Goal: predict customer's choice from a menu of items

Model: discrete-choice model to capture substitution

Random Utility Maximization (RUM)

Assumes
- randomly sample product utilities
- choose max utility product

Models multinomial logit (MNL), nested logit (NL), mixed logit (ML), rank-based,…



HOW TO SELECT A GOOD MODEL?

Models to fit to transaction-level data

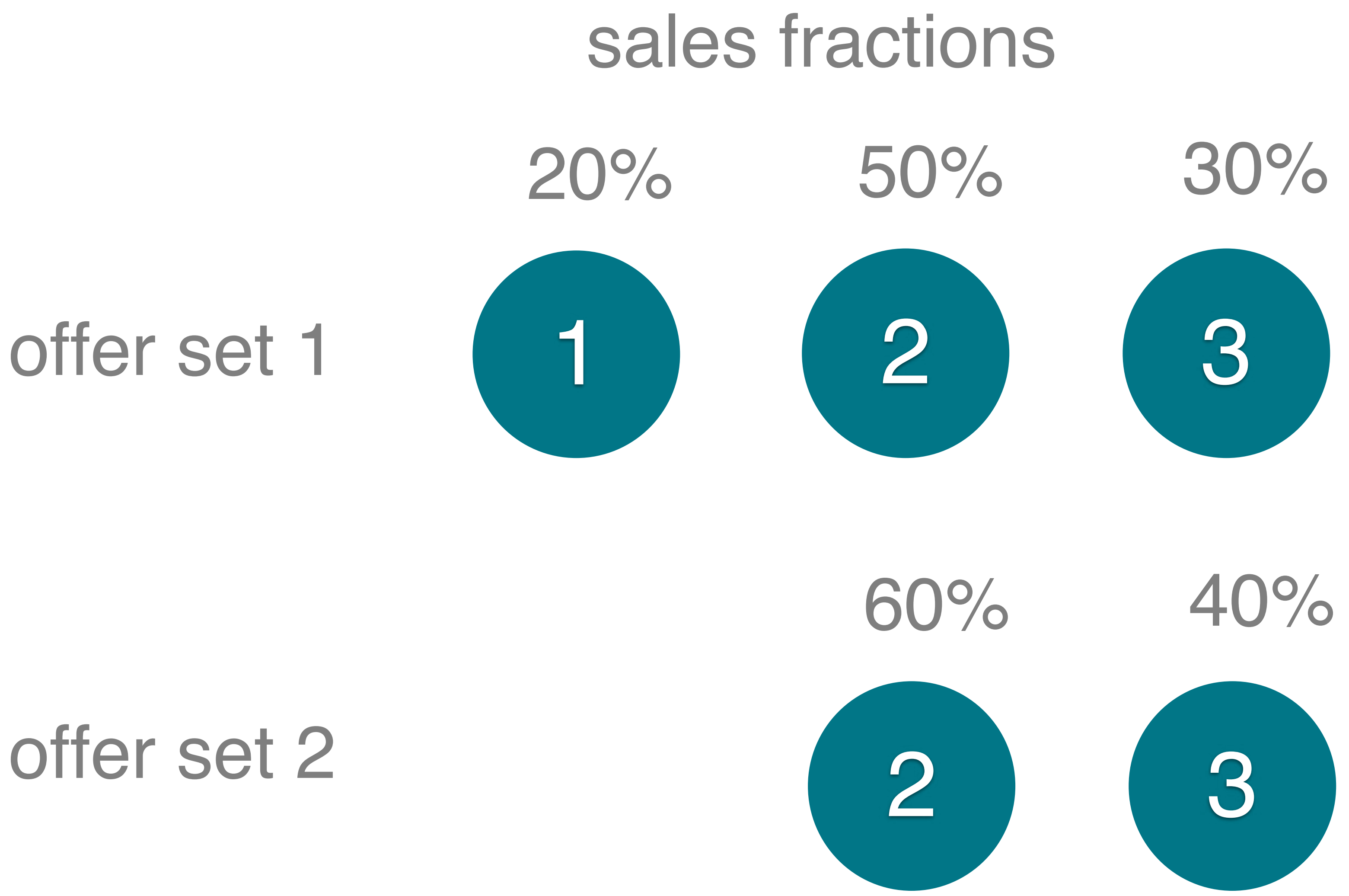
- collection of offer sets $S_1, S_2, .., S_m$
- sales fractions f_{jS} = fraction of times j purchased when S was offered

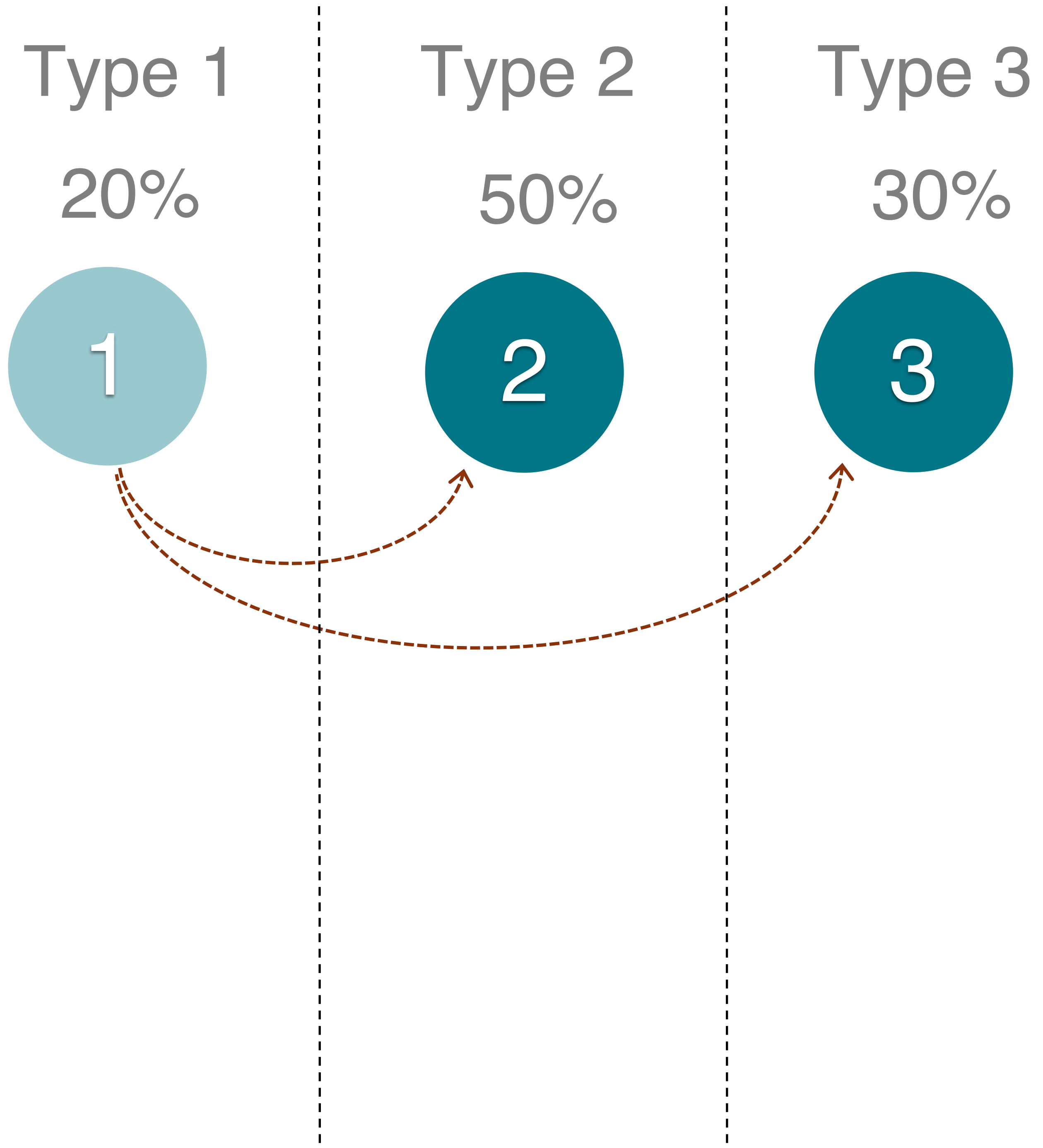
RUM model fit: are revealed customer preferences transitive?

Question: are the choice observations (stochastically) rationalizable?

RUM model fit: are revealed customer preferences transitive?

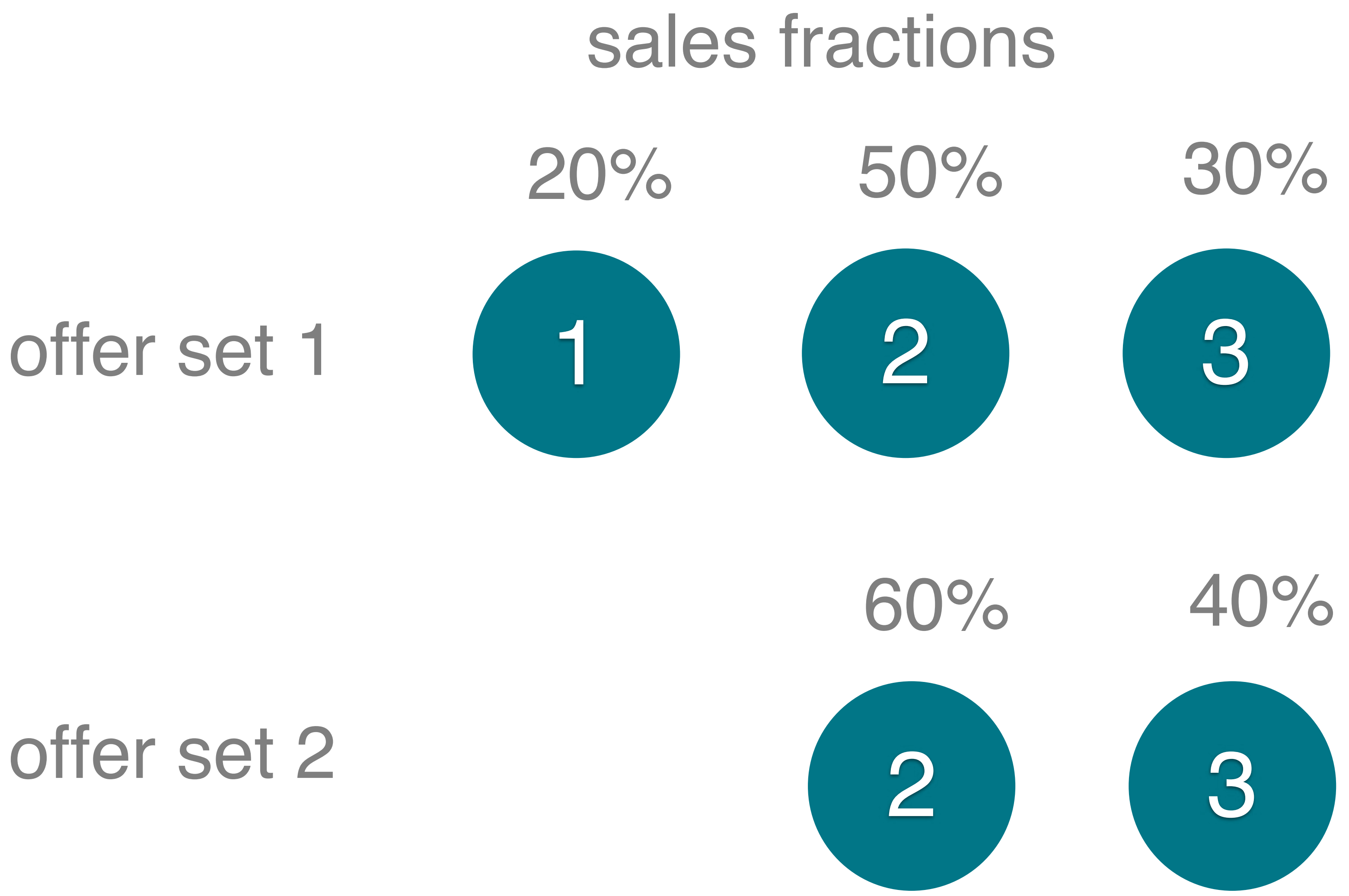
Question: are the choice observations (stochastically) rationalizable?

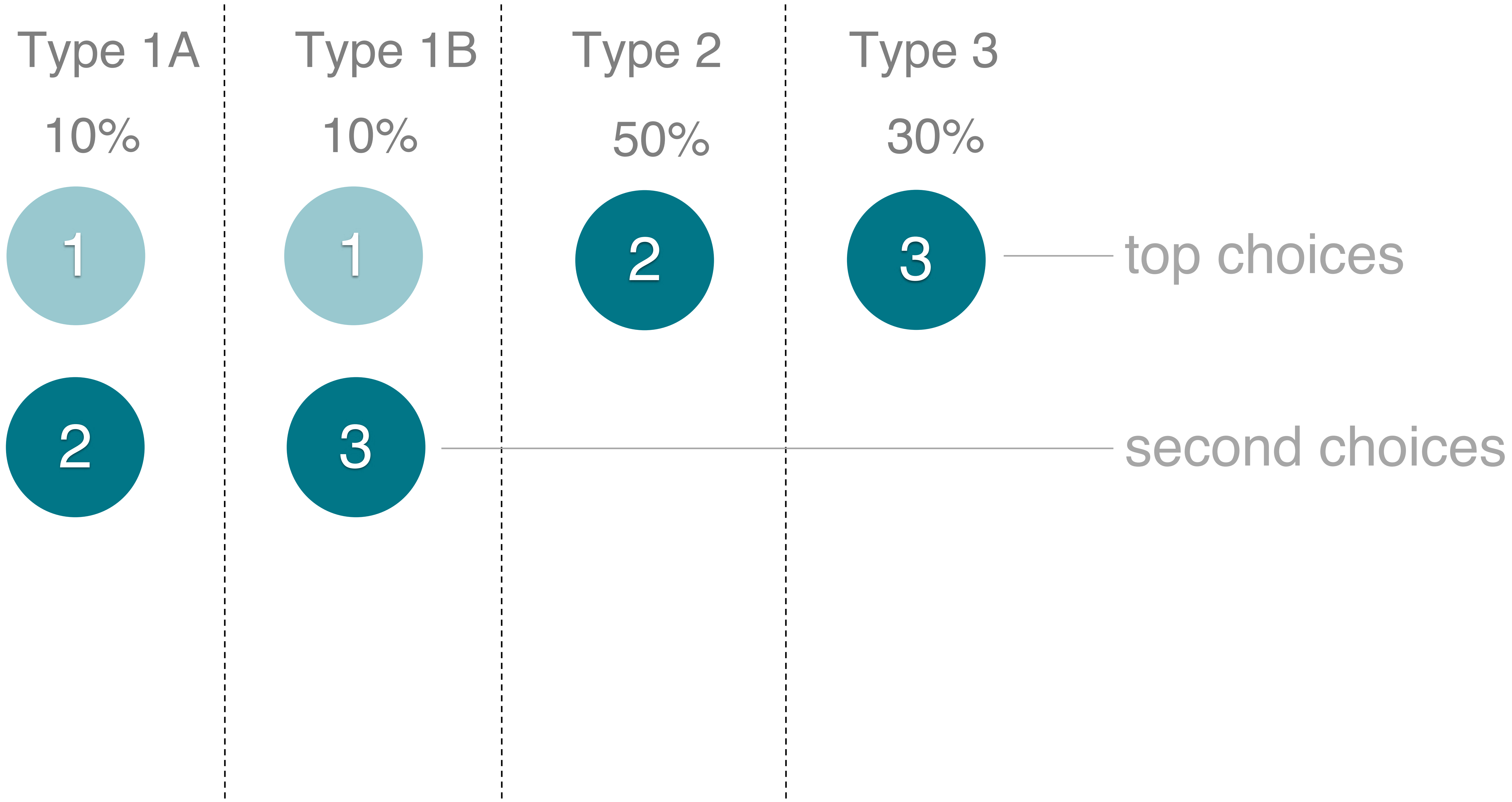




RUM model fit: are revealed customer preferences transitive?

Question: are the choice observations (stochastically) rationalizable?





$$\mathbb{P}\left[2 \,\middle|\, 2 \;\; 3 \right] = 10\% + 50\% = 60\%$$

RUM model fit: are revealed customer preferences transitive?

Question: are the choice observations (stochastically) rationalizable?

sales fractions

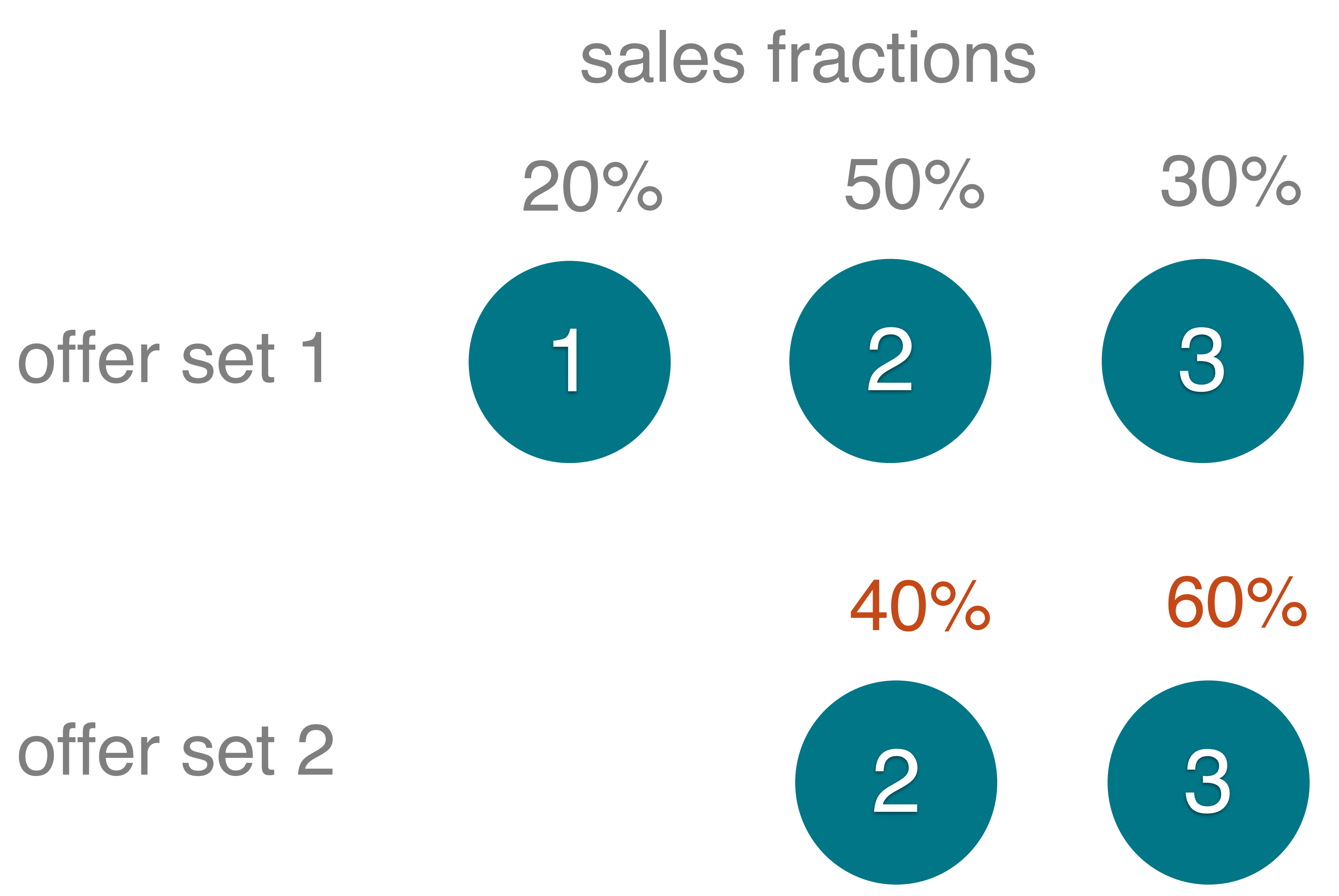
	20%	50%	30%
offer set 1	1	2	3
		60%	40%
offer set 2		2	3

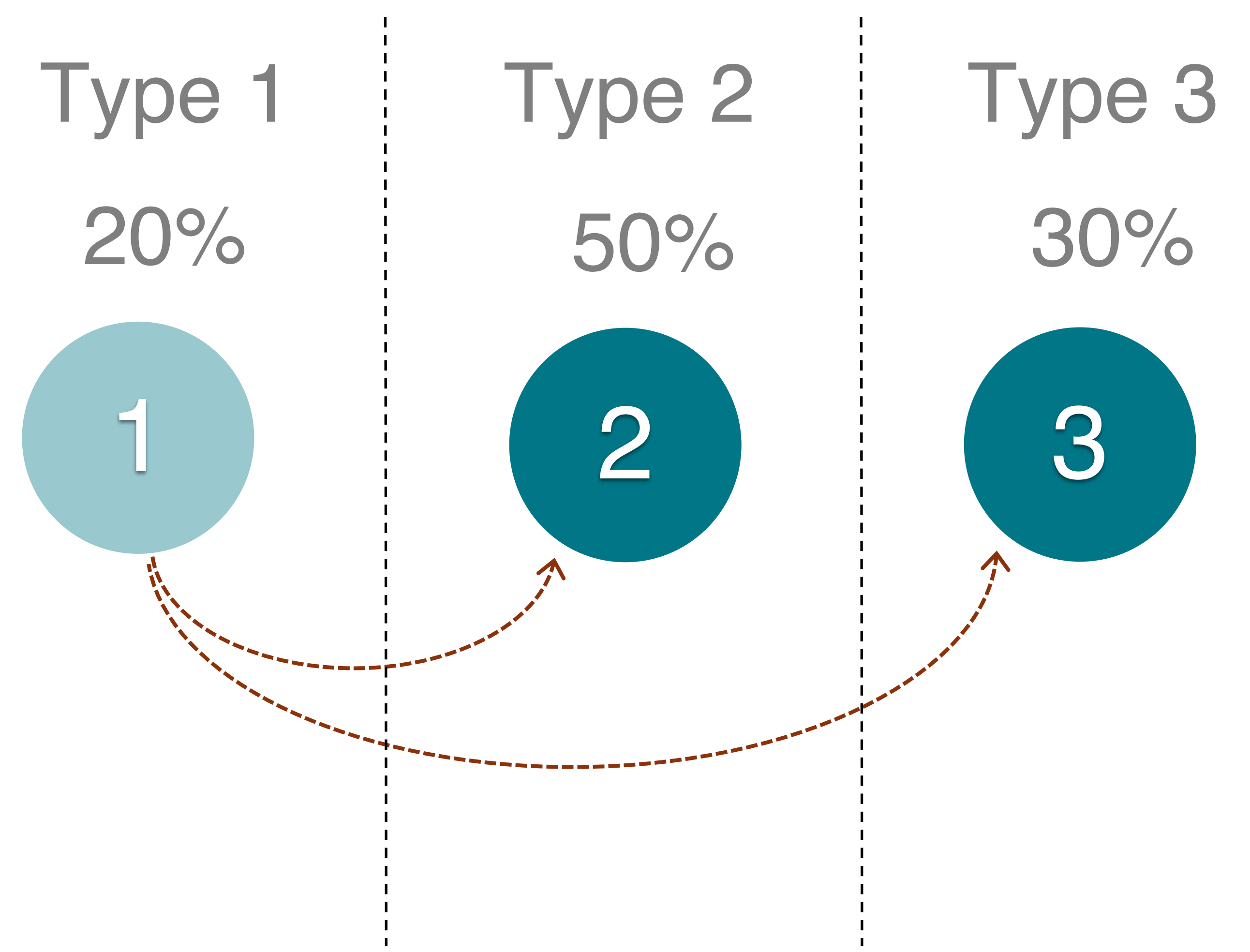
	Type 1A	Type 1B	Type 2	Type 3
	10%	10%	50%	30%
top choices	1	1	2	3
second choices	2	3	3	2
third choices	3	2	1	1

YES, CHOICES CONSISTENT
W/ DISTRIBUTION OVER RANKINGS

RUM model fit: are revealed customer preferences transitive?

Question: are the choice observations (stochastically) rationalizable?





sales of 2 cannot decrease if preferences are rational

Choices not consistent with ANY distribution over rankings

CHOICE DATA NOT STOCHASTICALLY RATIONALIZABLE

The LIMIT OF RATIONALITY: the smallest possible loss from assuming rationality

n mutually substitutable products

observations: choice/sales fractions for sets S_1, S_2,…, S_m

$f_{j,S}$ = fraction of times *j* purchased when *S* offered

Limit of Rationality
LoR (S_1,..,S_m)

$$\min_{\boldsymbol{x}} \sum_S \underbrace{\text{loss}}_{\text{convex loss fn.}}\Big(\underbrace{(x_{j,S} : j \in S)}_{\text{rational choice fractions consistent with a distribution over rankings}}, \underbrace{(f_{j,S} : j \in S)}_{\text{obs. choice fractions}}\Big)$$

convex loss fn.

obs. choice fractions

rational choice fractions consistent with a distribution over rankings

The LIMIT OF RATIONALITY: the smallest possible loss from assuming rationality

n mutually substitutable products

observations: choice/sales fractions for sets $S_1, S_2, \ldots, S_m$

$f_{j,S}$ = fraction of times *j* purchased when *S* offered

Limit of Rationality
LoR ($S_1,..,S_m$)

$$\min_{\boldsymbol{x}} \quad -\frac{1}{m}\sum_S \sum_{j \in S} f_{j,S} \log\left(x_{j,S}/f_{j,S}\right)$$

KL-divergence:
between obs. and pred. sales fractions

s.t. $$x_{j,S} = \sum_\sigma \lambda_\sigma \cdot \mathbb{I}[j \text{ top ranked among items in } S \text{ under } \sigma]\ \forall\ j, S$$

$$\sum_\sigma \lambda_\sigma = 1, \lambda_\sigma \geq 0\ \forall\ \sigma$$

The LIMIT OF RATIONALITY: the smallest possible loss from assuming rationality

Limit of Rationality
LoR (S_1,..,S_m)

$$\min_{\boldsymbol{x}} \; -\frac{1}{m}\sum_{S}\sum_{j\in S} f_{j,S}\log\left(x_{j,S}/f_{j,S}\right)$$

KL-divergence:
between obs. and pred. sales fractions

s.t. $\cancel{x_{j,S} = \sum_{\sigma}\lambda_\sigma \cdot \mathbb{I}[j \text{ top ranked among items in } S \text{ under } \sigma]\ \forall\ j, S}$

$$\cancel{\sum_{\sigma}\lambda_\sigma = 1, \lambda_\sigma \geq 0\ \forall\ \sigma}$$

vectors of length L

$|S_1| + |S_2| + \cdots + |S_m| = L$

$$\boldsymbol{x} \in \operatorname{conv}\left(\{\boldsymbol{e}_\sigma : \sigma \text{ is a ranking}\}\right)$$

convex hull of permutation representations

$$(\boldsymbol{e}_\sigma)_{j,S} = \mathbb{I}[j \text{ top ranked among items in } S \text{ under } \sigma]$$

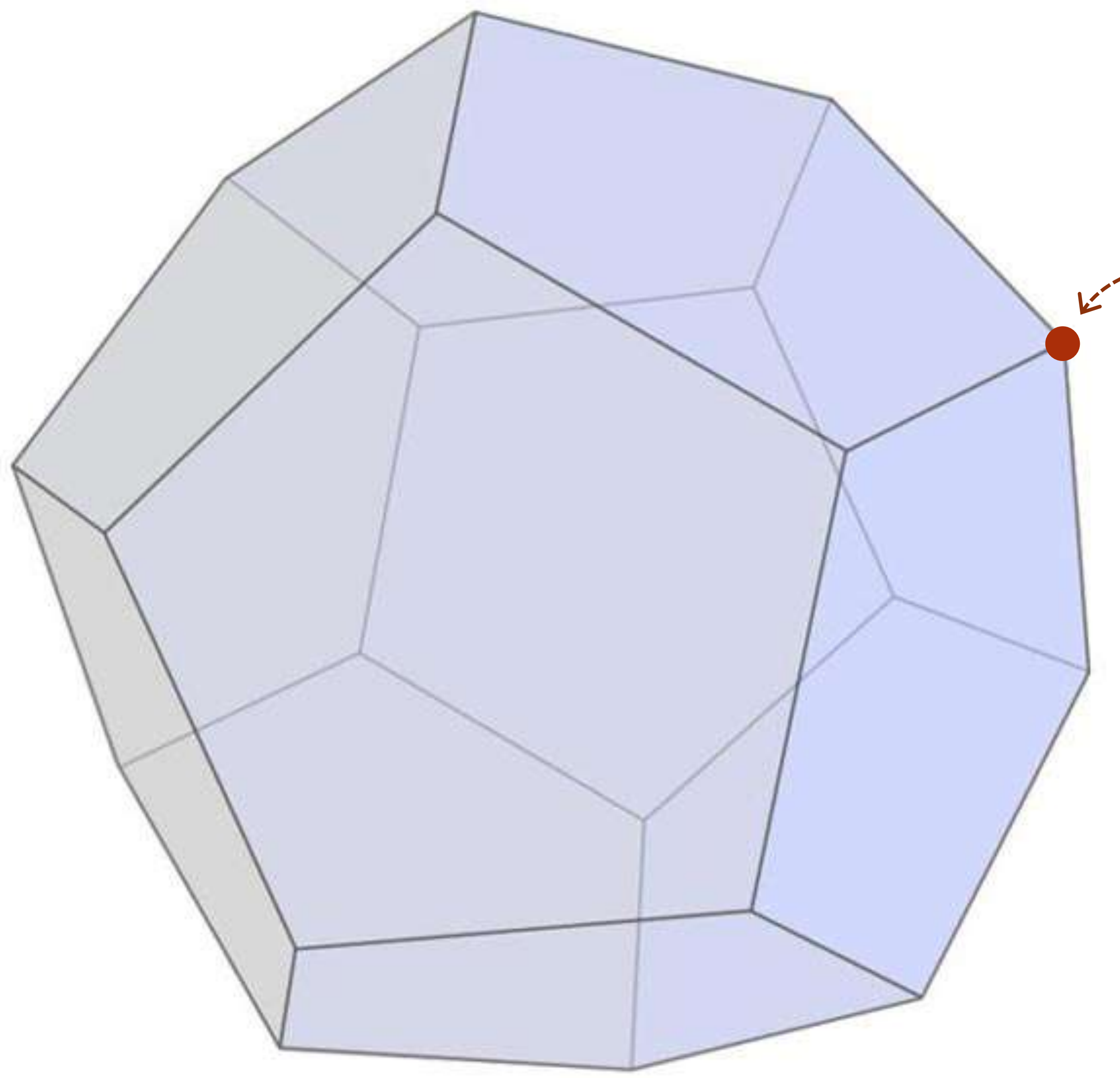
The LIMIT OF RATIONALITY: the smallest possible loss from assuming rationality

Limit of Rationality
LoR ($S_1,..,S_m$)

$$\min_{\boldsymbol{x}} \quad -\frac{1}{m}\sum_{S}\sum_{j\in S} f_{j,S}\log\left(x_{j,S}/f_{j,S}\right)$$

KL-divergence:
between obs. and pred. sales fractions

$$\textit{s.t.} \qquad \boldsymbol{x} \in \operatorname{conv}\left(\{\boldsymbol{e}_\sigma : \sigma \text{ is a ranking}\}\right)$$



$\boldsymbol{e}_\sigma$

The LIMIT OF RATIONALITY: the smallest possible loss from assuming rationality

Limit of Rationality
LoR ($S_1,..,S_m$)

$$\min_{\boldsymbol{x}} \quad -\frac{1}{m}\sum_{S}\sum_{j\in S} f_{j,S}\log\left(x_{j,S}/f_{j,S}\right)$$

KL-divergence:
between obs. and pred. sales fractions

s.t. $$\boldsymbol{x} \in \mathrm{conv}\left(\{\boldsymbol{e}_\sigma : \sigma \text{ is a ranking}\}\right)$$

Key Contributions:

1. characterization of the complexity of the above polytope in terms of structure of the set collection $S_1, .., S_m$
2. model-selection diagnostic tool for practitioners

Existing work doesn't address computation

Economics

Focus is on the binary question:
are the observed sales fractions stochastically rationalizable?

Key result: necessary & sufficient conditions for LoR(S_1,..,S_m) = 0
[Falmage '78][Barbera, Pattanaik '86][McFadden, Richter '90][McFadden '05]
Does not address computability

Computer Science

Focus on rank aggregation for PAIRS = $\{ \{i, j\}\ i \neq j \}$
Key results: NP-hardness [Dwork 01] and approx. algorithms
[Dwork, Kumar, Naor, Sivakumar '01][Kenyon-Mathieu, Schudy '07] [Ali, Melia '12]
Does not consider general set collections

Computational hardness depends on the
structure of collection of subsets

The Limit of Rationality in Choice Modeling

1. Case study with IRI Academic Dataset
   what is the ‘right’ choice model?

2. Complexity of computing LoR
   rational separation, complexity in terms of graph properties

3. Summary/Conclusions
   takeaway messages

The Limit of Rationality in Choice Modeling

1. Case study with IRI Academic Dataset
what is the 'right' choice model?

2. Complexity of computing LoR
rational separation, complexity in terms of graph properties

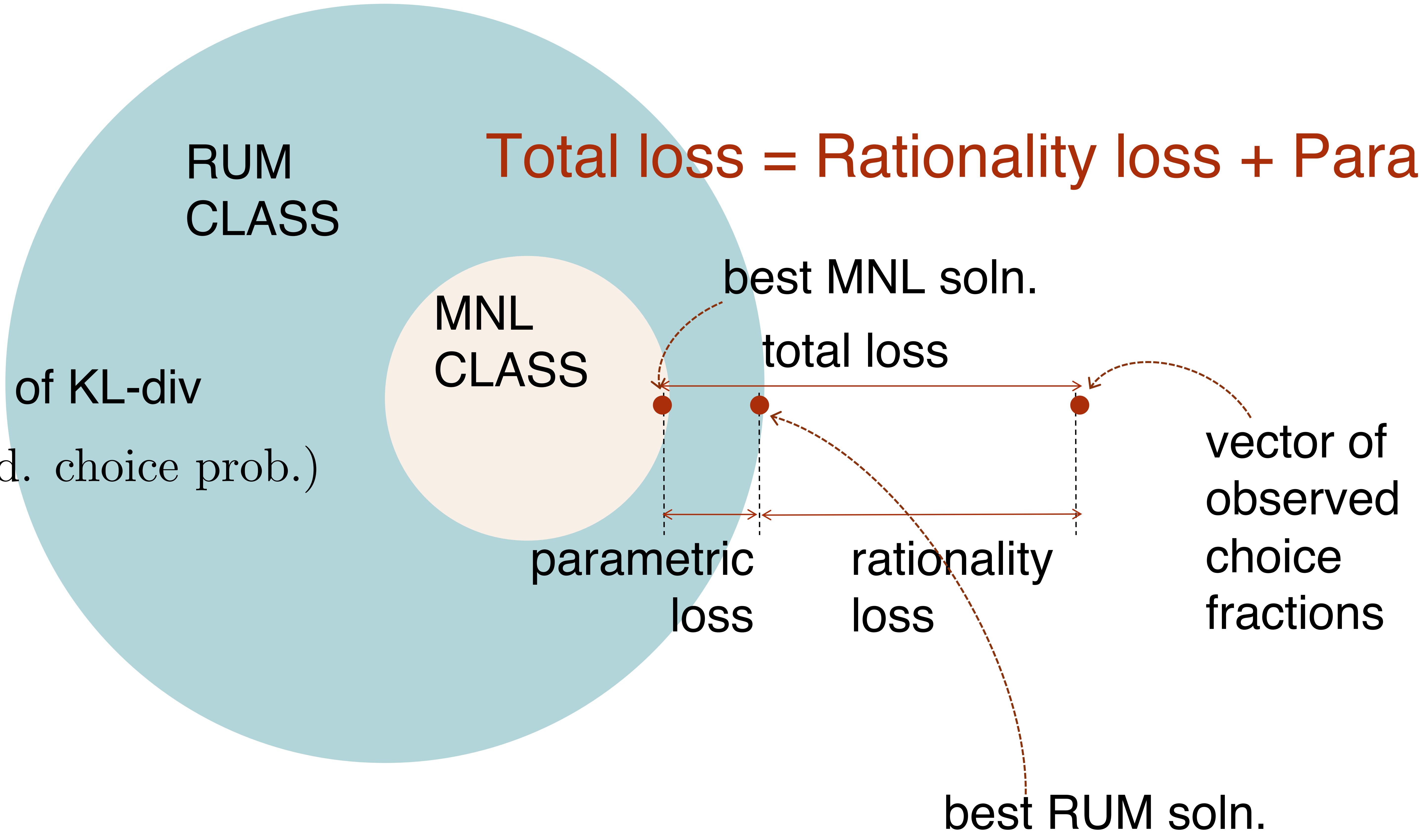
3. Summary/Conclusions
takeaway messages

LoR provides guidance on which model to fit: more complex RUM or outside RUM?

Total loss = Rationality loss + Parametric loss

Distance measured in terms of KL-div

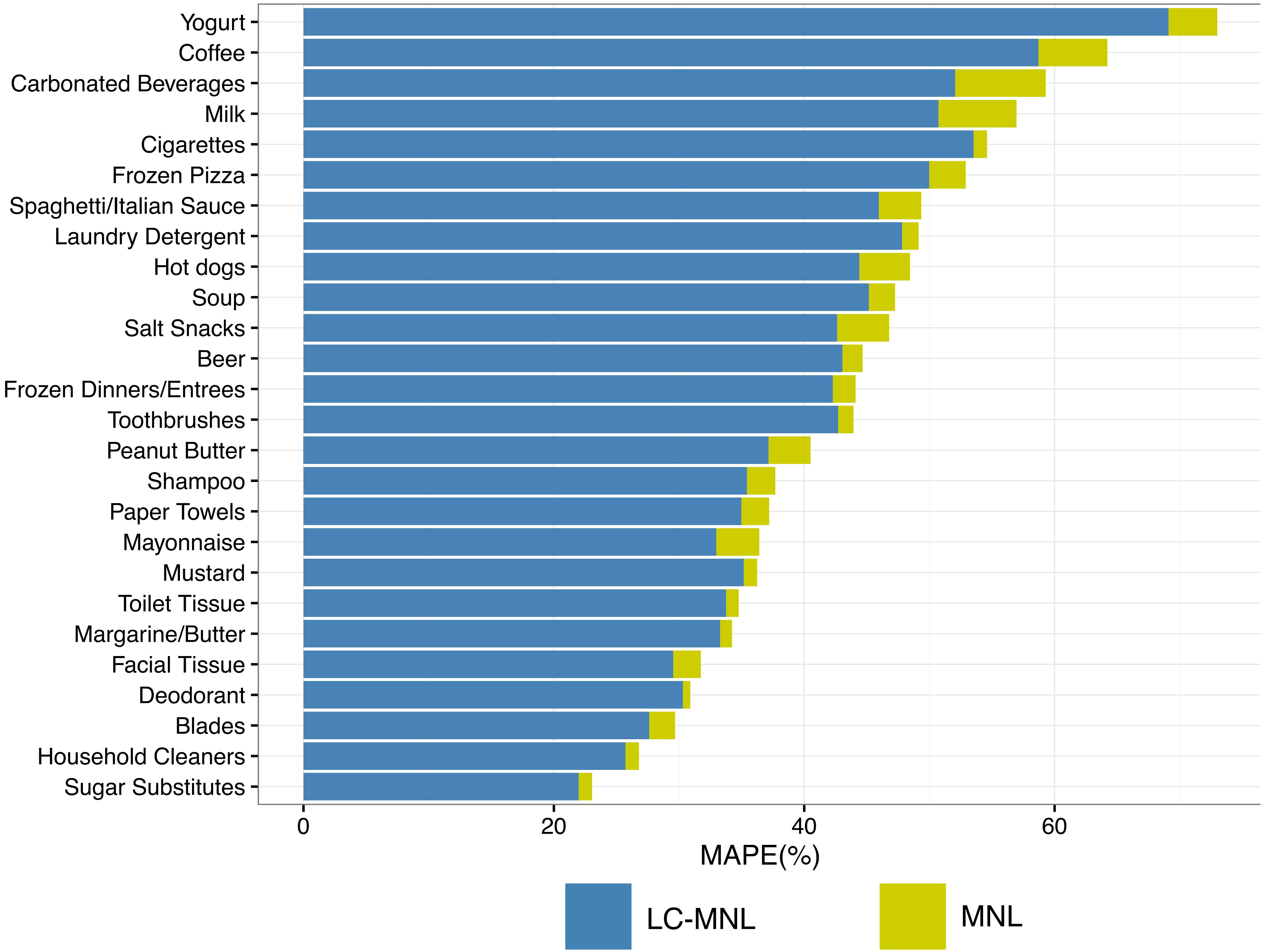
$$D_{\mathrm{KL}}\left(\text{obs. choice prob.} \,\|\, \text{pred. choice prob.}\right)$$

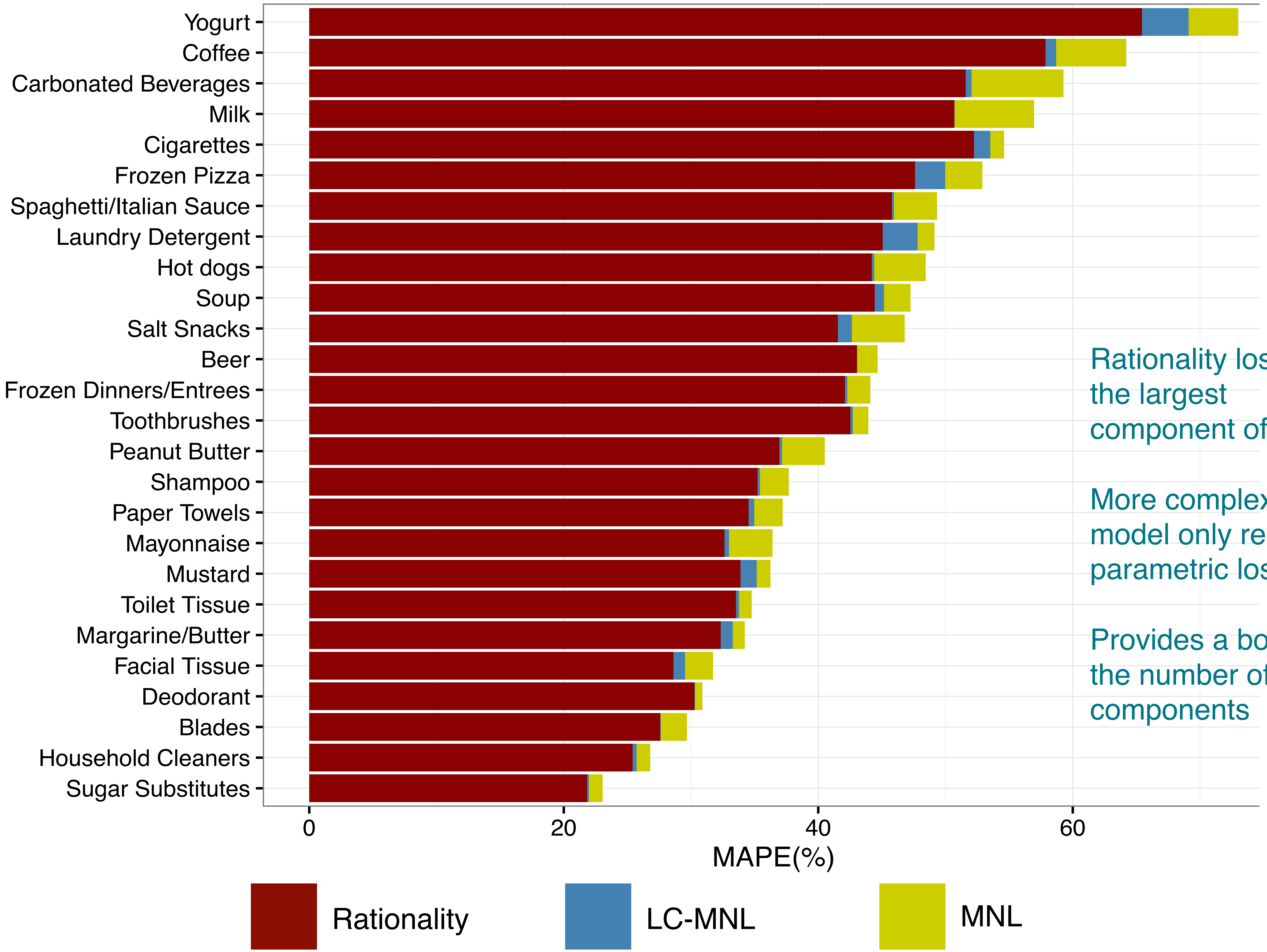


To reduce:

Parametric loss fit more complex RUM models

Rationality loss go outside RUM class

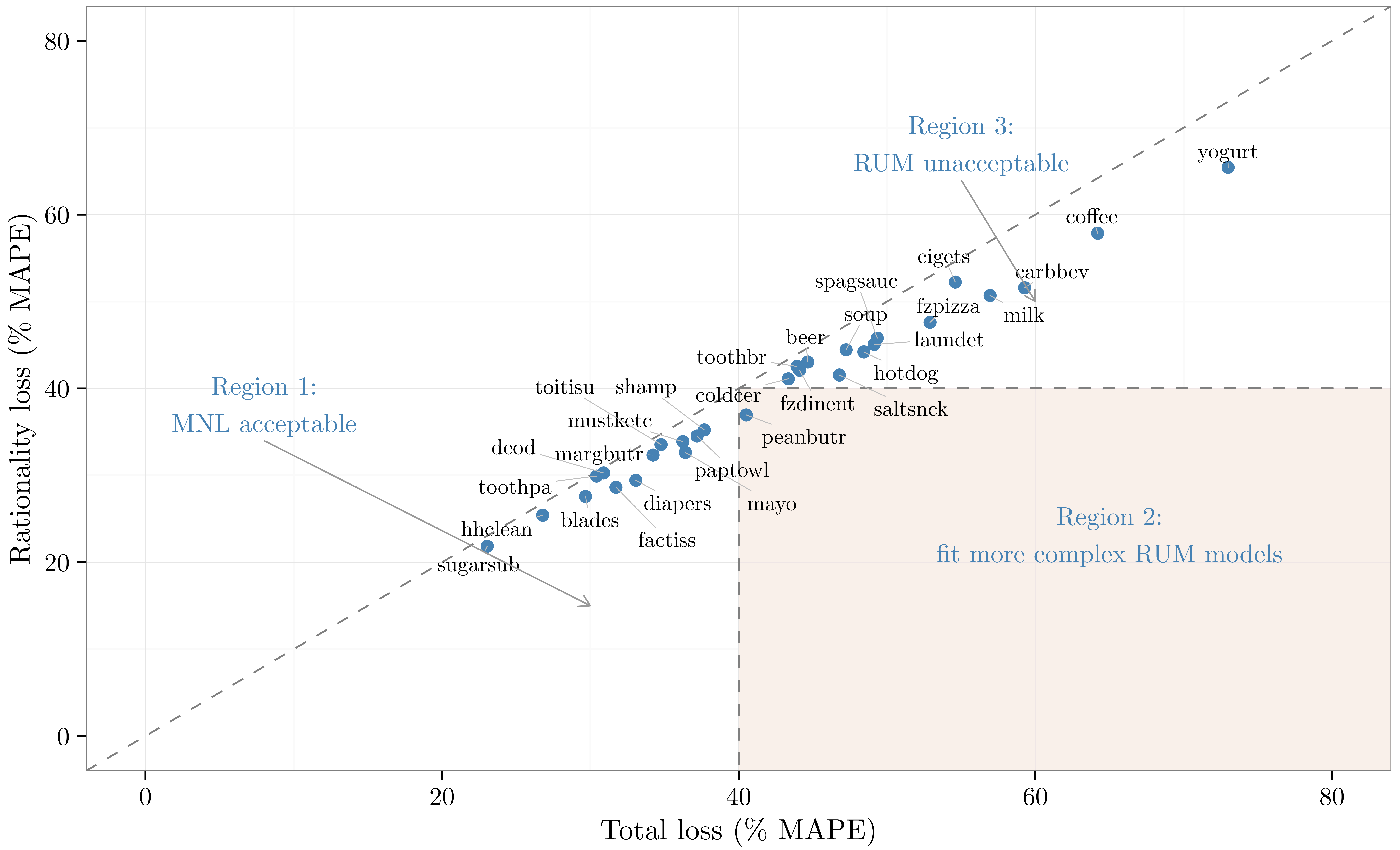

Yogurt
Coffee
Carbonated Beverages
Milk
Cigarettes
Frozen Pizza
Spaghetti/Italian Sauce
Laundry Detergent
Hot dogs
Soup
Salt Snacks
Beer
Frozen Dinners/Entrees
Toothbrushes
Peanut Butter
Shampoo
Paper Towels
Mayonnaise
Mustard
Toilet Tissue
Margarine/Butter
Facial Tissue
Deodorant
Blades
Household Cleaners
Sugar Substitutes
0
20
40
60
MAPE(%)
LC-MNL
MNL



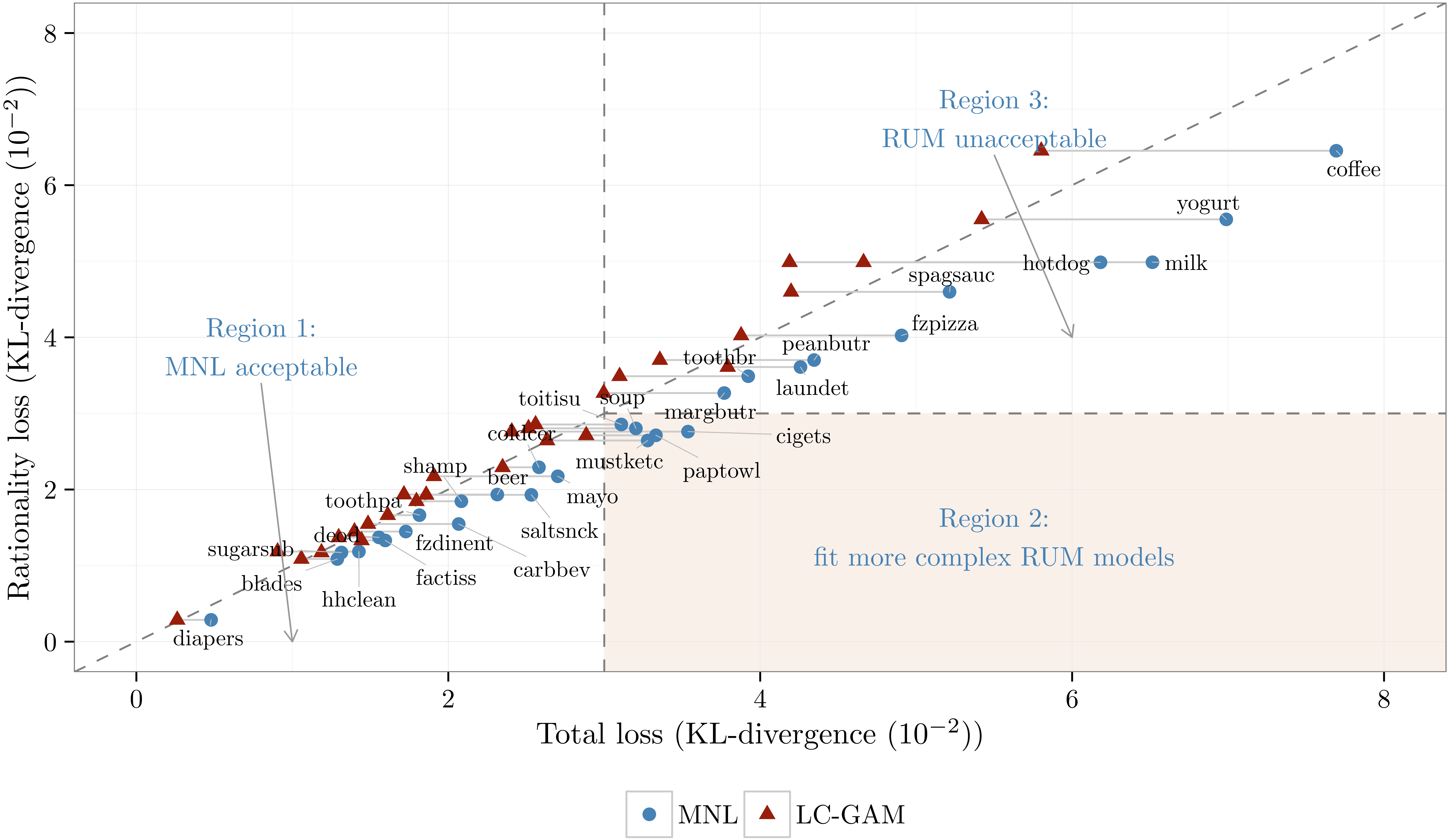
Rationality loss is the largest component of loss

More complex RUM model only reduces parametric loss

Provides a bound on the number of mixture components



Region 1:
MNL acceptable
Region 2:
fit more complex RUM models
Region 3:
RUM unacceptable
yogurt
coffee
cigets
carbbev
milk
spagsauc
fzpizza
soup
laundet
beer
toothbr
hotdog
coldcer
fzdinent
saltsnck
toitisu
shamp
peanbutr
mustketc
deod
margbutr
paptowl
mayo
toothpa
diapers
blades
hhclean
factiss
sugarsub
80
60
40
20
0
Rationality loss (% MAPE)
Total loss (% MAPE)


Region 3:
RUM unacceptable
Region 1:
MNL acceptable
Region 2:
fit more complex RUM models
coffee
yogurt
hotdog
milk
spagsauc
fzpizza
peanubtr
toothbr
laundet
toitisu
soup
margbutr
cigets
coldcer
mustketc
paptowl
shamp
beer
mayo
toothpa
saltsnck
deod
fzdinent
sugarsub
carbbev
factiss
blades
hhclean
diapers
Rationality loss (KL-divergence (10^{-2}))
Total loss (KL-divergence (10^{-2}))
0
2
4
6
8
MNL
LC-GAM

Factors influencing rationality loss: variety-seeking and hedonic purchasing

$$\text{loss}_j = \beta_0 + \beta_1 \cdot (\text{market concentration})_j + \beta_2 \cdot (\text{hedonic indicator})_j + \varepsilon_j$$

negative of market
share entropy

food/snack
(excl. condiments)
cigarettes
drinks

$\beta_1 < 0$

$\beta_2 > 0$

variety seeking $\Longrightarrow$
low concentration
(e.g., yogurt w/ new product intros.)

hedonic $\Longrightarrow$
complex choice behavior

non-transitive preferences

non-transitive preferences

The Limit of Rationality in Choice Modeling

1. Case study with IRI Academic Dataset
what is the 'right' choice model?

2. Complexity of computing LoR
rational separation, complexity in terms of graph properties

3. Summary/Conclusions
takeaway messages

Computing the Limit of Rationality: solve a large-dimensional constrained convex program

Limit of Rationality
LoR ($S_1,..,S_m$)

$$\min_{\boldsymbol{x}} \quad -\frac{1}{m}\sum_{S}\sum_{j\in S} f_{j,S}\log\left(x_{j,S}/f_{j,S}\right)$$

$$\textit{s.t.} \qquad \boldsymbol{x} \in \mathrm{conv}\left(\left\{\boldsymbol{e}_\sigma : \sigma \text{ is a ranking}\right\}\right)$$

Constrained convex program in a large dimensional space

Frank-Wolfe algorithm

solution end of iteration k

$$\min_{\boldsymbol{x}\in\mathcal{D}} h(\boldsymbol{x}) \xrightarrow[\text{optimize linear approx.}]{\text{in iteration } k+1,} \min_{\boldsymbol{y}\in\mathcal{D}} \langle \boldsymbol{y}, \nabla h(\boldsymbol{x}^{(k)})\rangle$$

$\boldsymbol{x}^{(k+1)} =$ convex combination of $\boldsymbol{x}^{(k)}$ and $\arg\min_{\boldsymbol{y}\in\mathcal{D}} \langle \boldsymbol{y}, \nabla h(\boldsymbol{x}^{(k)})\rangle$

Computing the Limit of Rationality: solve a large-dimensional constrained convex program

Limit of Rationality
LoR (S_1,..,S_m)

$$\min_{\boldsymbol{x}} \; -\frac{1}{m}\sum_{S}\sum_{j\in S} f_{j,S}\log\left(x_{j,S}/f_{j,S}\right)$$

s.t. $\quad \boldsymbol{x} \in \operatorname{conv}\left(\left\{\boldsymbol{e}_\sigma : \sigma \text{ is a ranking}\right\}\right)$

$$\min_{\boldsymbol{y}\in\mathcal{D}} \langle \boldsymbol{y}, \nabla h(\boldsymbol{x}^{(k)})\rangle \xrightarrow{\text{in our case}} \min_{\boldsymbol{y}} \; -\frac{1}{m}\sum_{S}\sum_{j\in S}(f_{j,S}/x_{j,S}^{(k)})\cdot y_{j,S}$$

s.t. $\boldsymbol{y} \in \operatorname{conv}\left(\left\{\boldsymbol{e}_\sigma : \sigma \text{ is a ranking}\right\}\right)$

because LP,
optimal at extreme point

$$\min_{\sigma} \; -\frac{1}{m}\sum_{S}\sum_{j\in S}(f_{j,S}/x_{j,S}^{(k)})\cdot (e_\sigma)_{j,S}$$

Computing the Limit of Rationality: solve a rank-aggregation problem

FW-subproblem

optimize linear approx.

$$\min_{\sigma} \; -\frac{1}{m} \sum_{S} \sum_{j \in S} (f_{j,S}/x_{j,S}^{(k)}) \cdot (e_\sigma)_{j,S}$$

$$\min_{\sigma} \;\; \sum_{S} \sum_{j \in S} c_{j,S} \cdot \mathbb{I}[j \text{ is top-ranked among items in } S \text{ under } \sigma]$$

cost coef. $-f_{j,S}/(m \cdot x_{j,S}^{(k)})$

Also called the RANK AGGREGATION problem

LoR

choosing best ranking

computational complexity

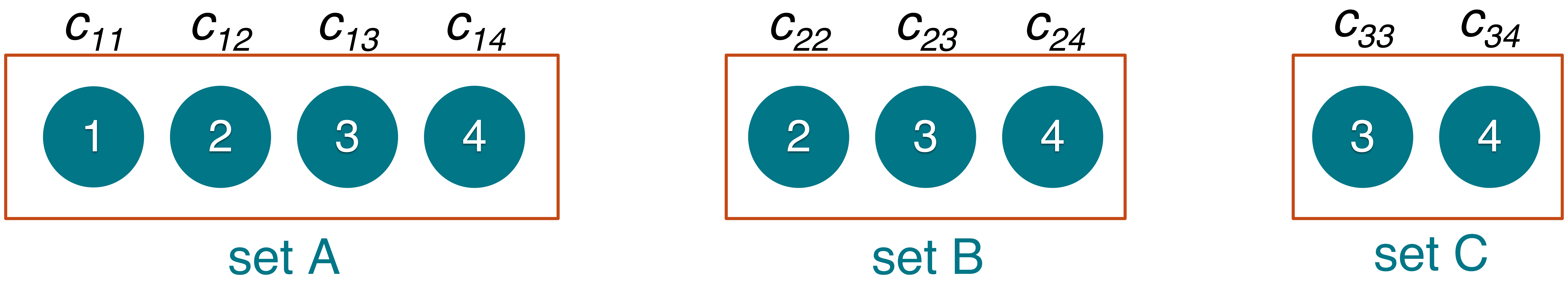
≡

rank-aggregation

choosing best item from each subset

computational complexity

The challenge in solving rank-aggregation: choices from different sets are not independent



$$\min_{x,y,z} \quad c_{1x} + c_{2y} + c_{3z}$$

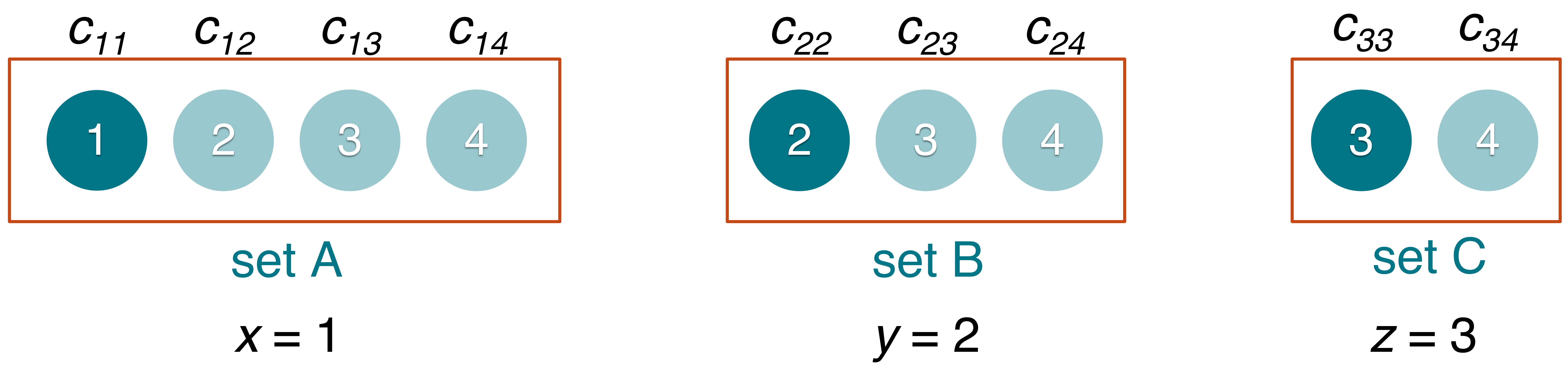
x chosen from set A

y chosen from set B

z chosen from set C

such that the choices are transitive

The challenge in solving rank-aggregation: choices from different sets are not independent



$$\min_{x,y,z} \quad c_{1x} + c_{2y} + c_{3z}$$

x chosen from set A

y chosen from set B

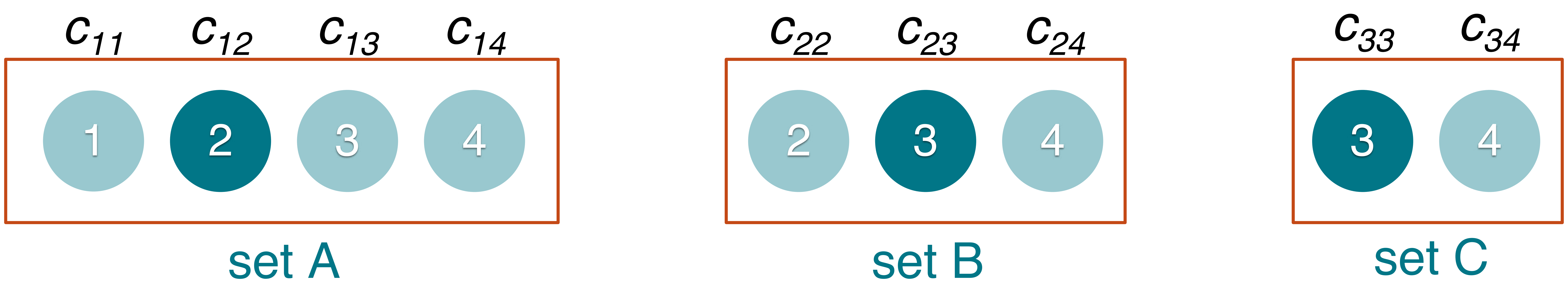
z chosen from set C

such that the choices are transitive

choices consistent with

$1 \succ 2 \succ 3$

The challenge in solving rank-aggregation: choices from different sets are not independent



$x = 2$ $y = 3$ $z = 3$

$$\min_{x,y,z} \quad c_{1x} + c_{2y} + c_{3z}$$

x chosen from set A

y chosen from set B

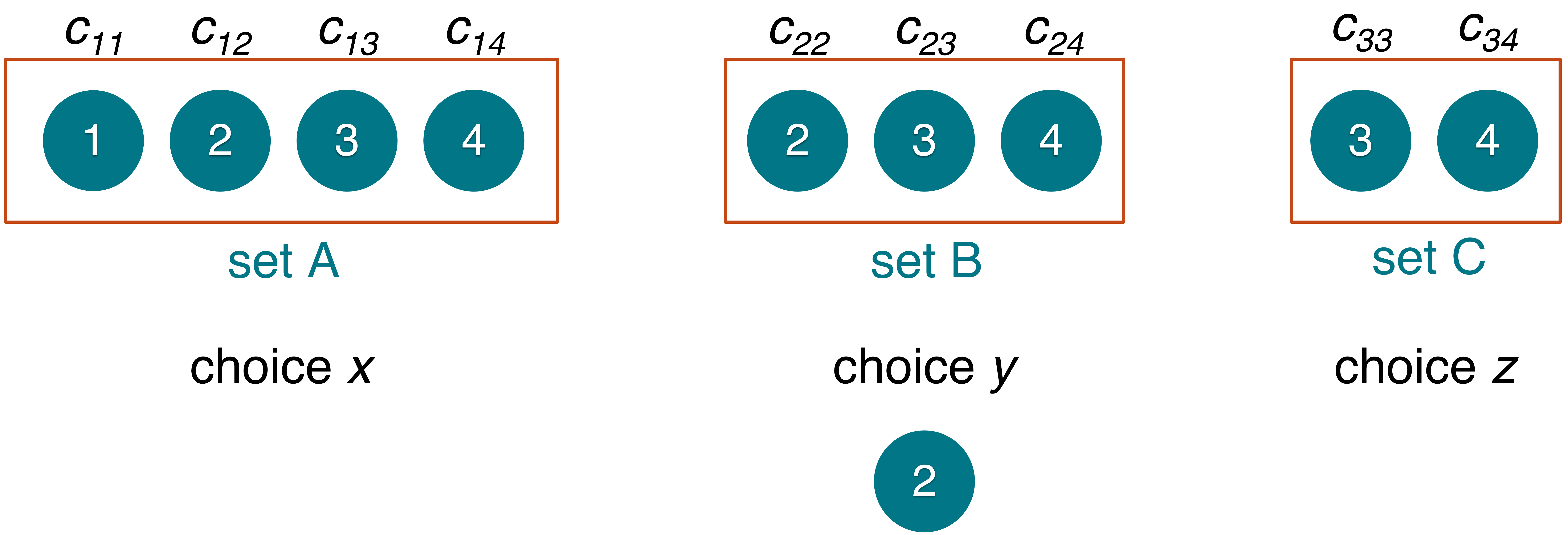
z chosen from set C

such that the choices are transitive

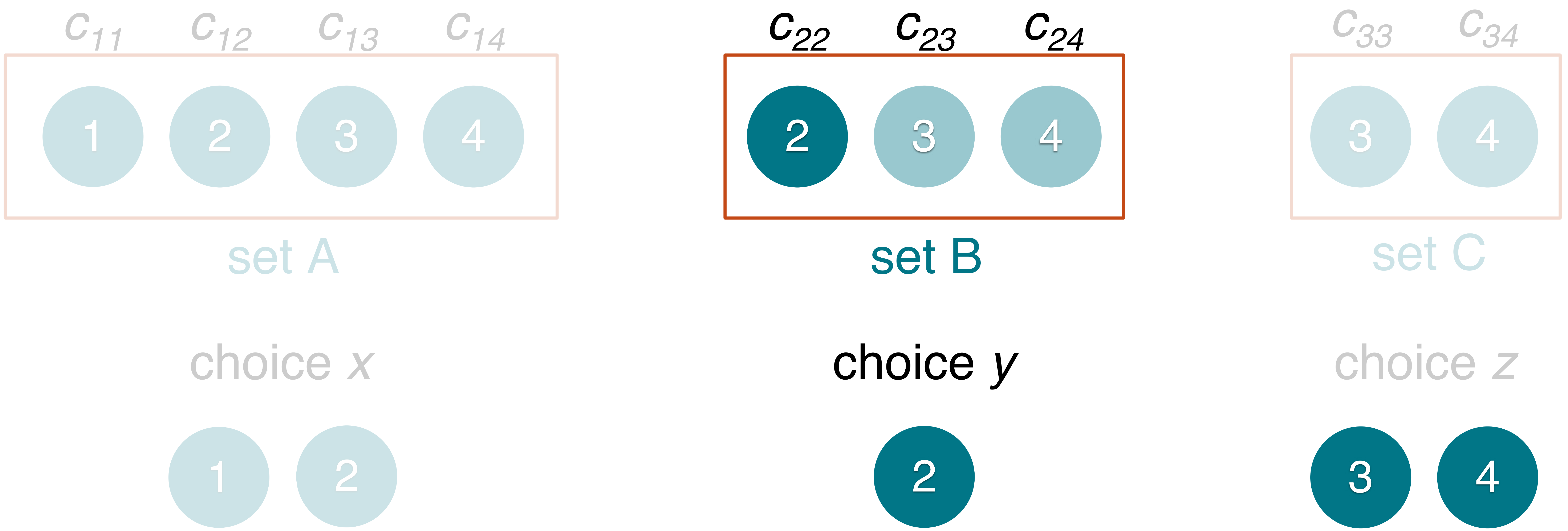
choices NOT transitive

$2 \succ 3$ contradicts $3 \succ 2$

Solution: “conditional” choices can be independent



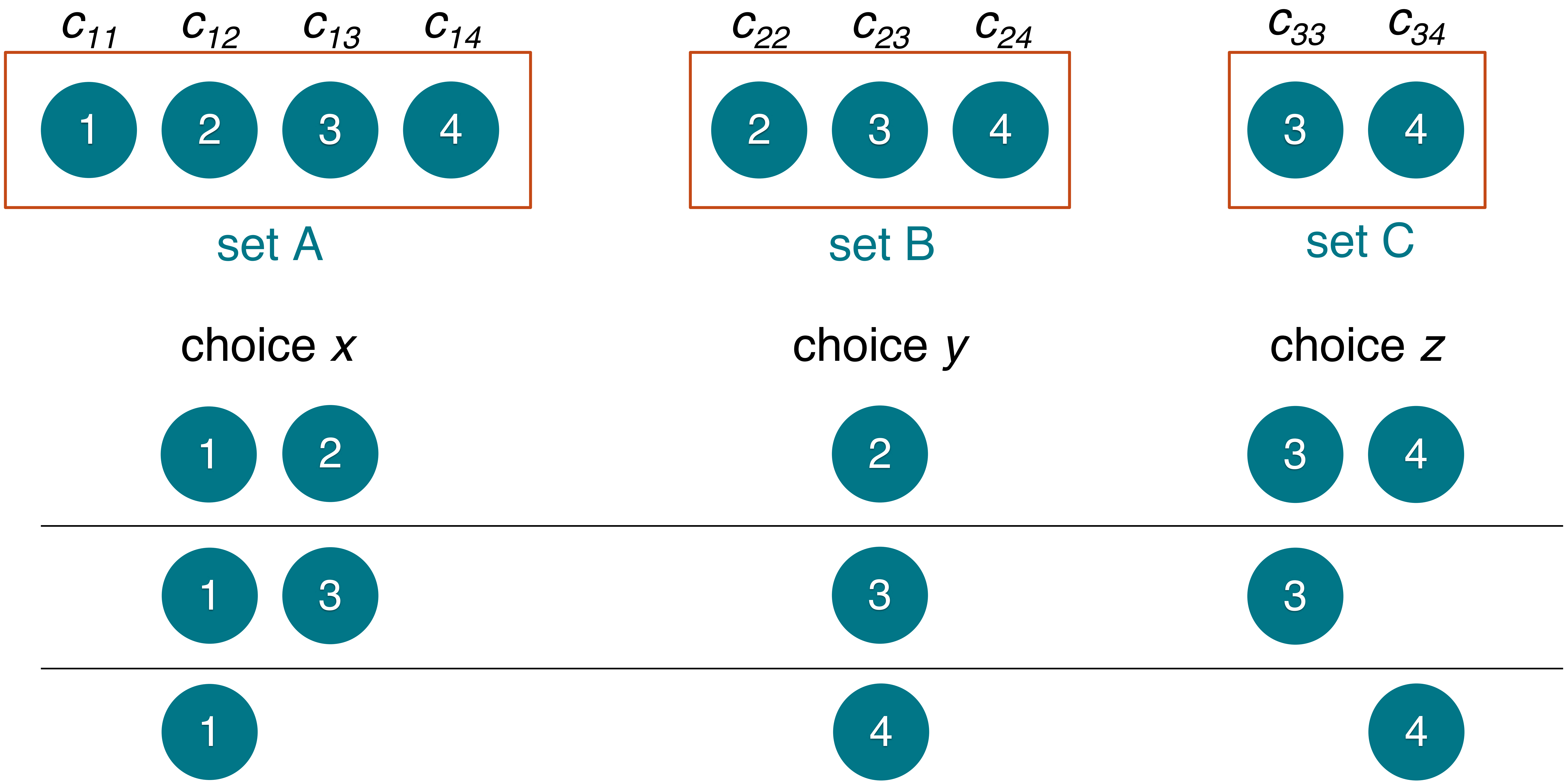
Solution: "conditional" choices can be independent



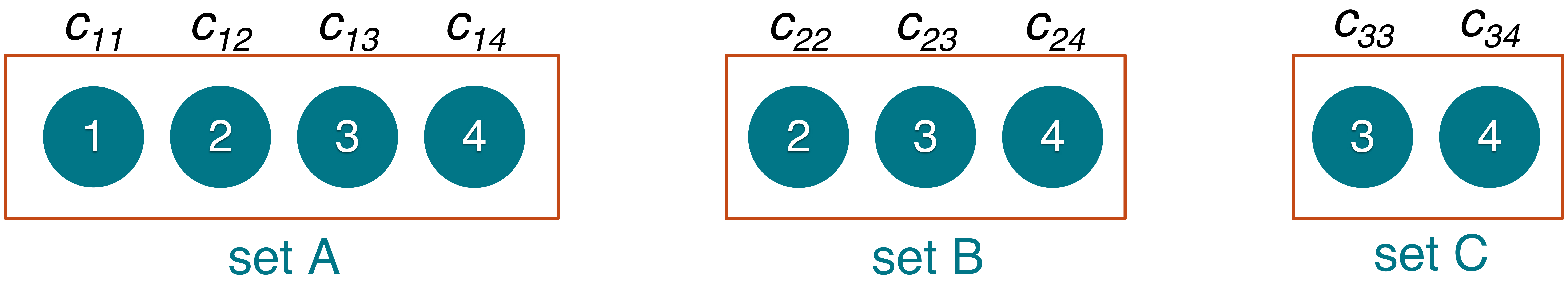
ALL POSSIBLE CHOICE COMBINATIONS ARE CONSISTENT

(x, y, z)	(1, 2, 3)	(2, 2, 3)	(1, 2, 4)	(2, 2, 4)
consistent with	$1 \succ 2 \succ 3 \succ 4$	$1 \succ 2 \succ 4 \succ 3$	$2 \succ 1 \succ 3 \succ 4$	$2 \succ 1 \succ 4 \succ 3$

Solution: “conditional” choices can be independent



Solution: “conditional” choices can be independent



choice x choice y choice z

Given choice y in set B,

choice x consistent w/ y and choice z consistent w/ y

$\Downarrow$

choices x,y,z mutually consistent

The concept of “rational separation”

Given choice *y* in set B,

choice *x* consistent w/ *y* and choice *z* consistent w/ *y*

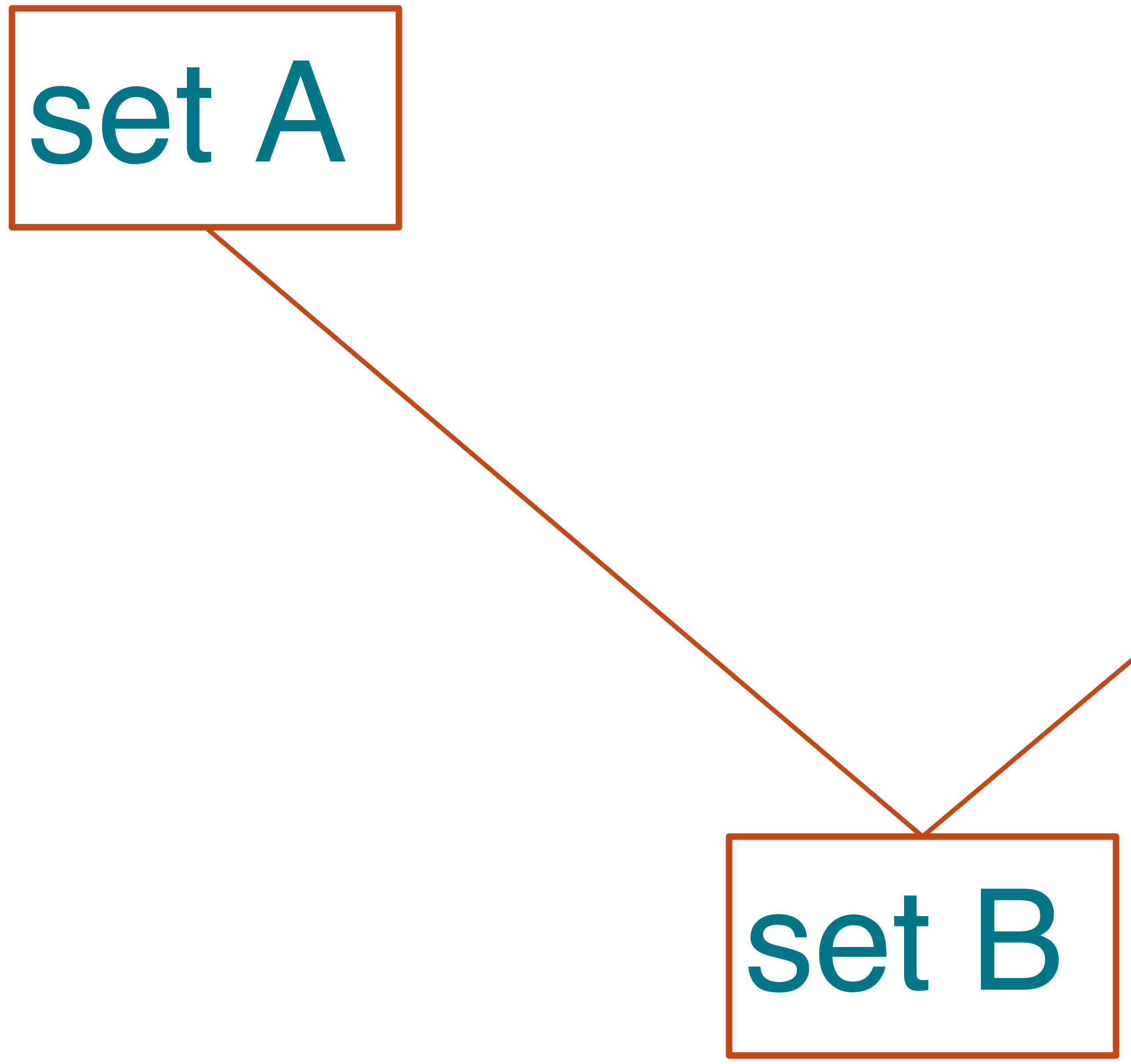
choices *x*,*y*,*z* mutually consistent for all choices *x*,*y*,*z*

B rationally separates *A* and *C* $A \perp\!\!\!\perp C \mid B$

Definition extends to collections of subsets: $\mathcal{A}$ $\mathcal{B}$ $\mathcal{C}$ $\mathcal{A} \perp\!\!\!\perp \mathcal{C} \mid \mathcal{B}$

Choice graph: graphical representation of rational separation

G is a choice-graph with subsets as nodes
if whenever $\mathcal{B}$ separates $\mathcal{A}$ and $\mathcal{C}$ in G,
$\mathcal{B}$ must rationally separate $\mathcal{A}$ and $\mathcal{C}$, i.e., $\mathcal{A} \perp\!\!\!\perp \mathcal{C} \mid \mathcal{B}$
for all collections of nodes $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$



B separates A and C in G:

if set B is removed,
then A and C are disconnected

or every path from set A to set C goes through set B

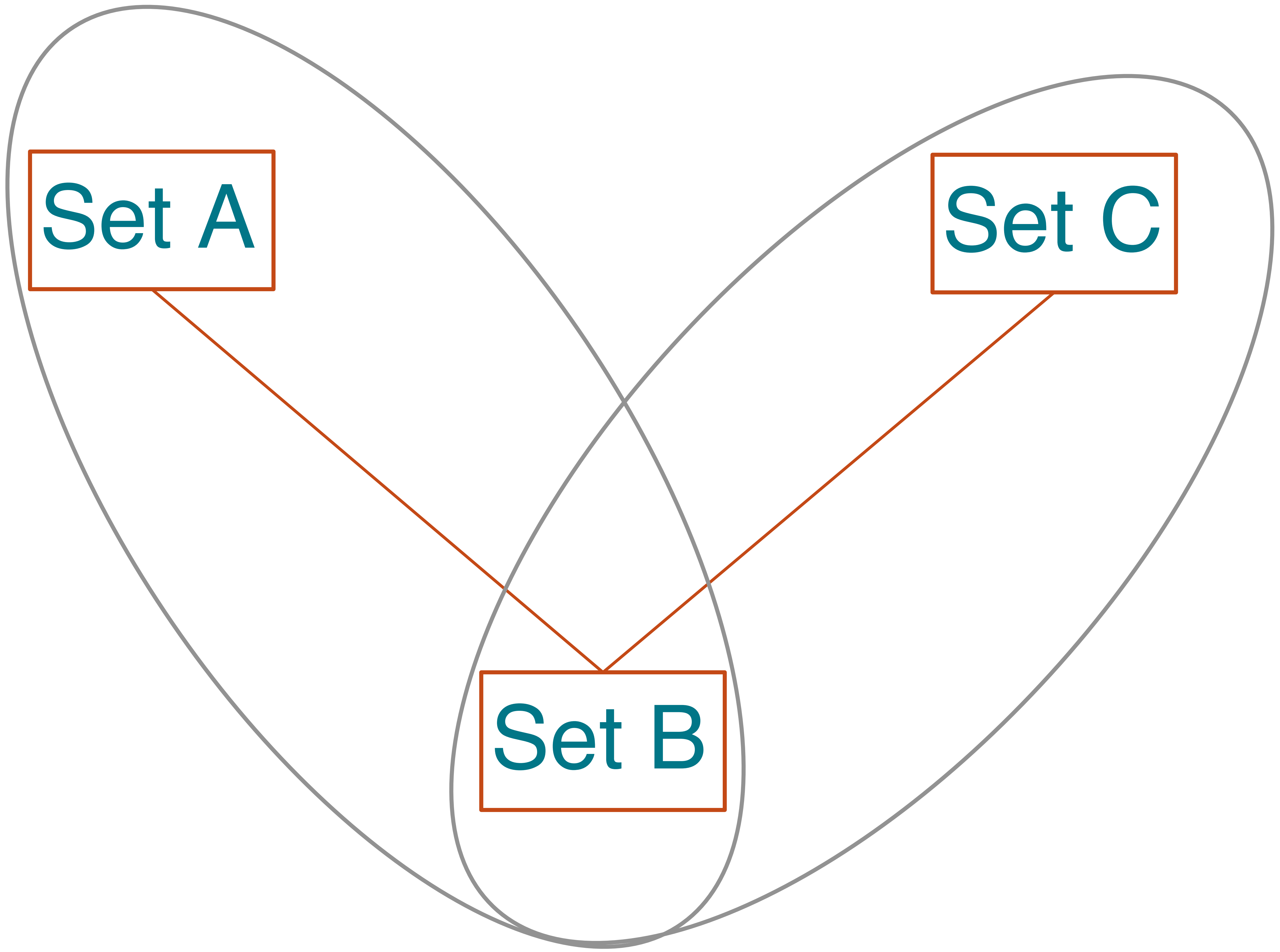
Rational separation reduces computational complexity of rank-aggregation

Reduces search space from $O(n^3)$ to $O(n^2)$

[for each choice from *C*, search over $O(n)$ choices]

for every element in *B*, find best element in *A*

$O(n^2)$

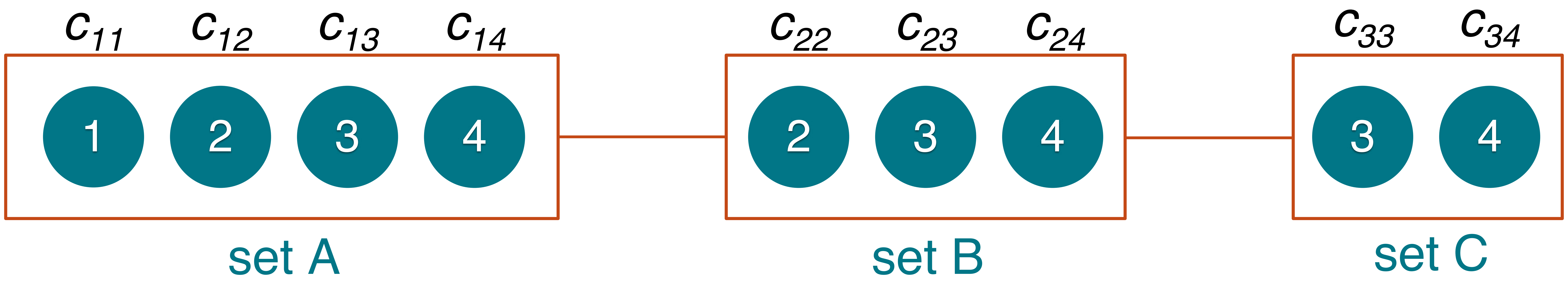


for every element in *B*, find best element in *C*

$O(n^2)$

Exploit rational separation to join the two solutions $O(n^2)$

Exploiting rational separation: DP for finding the best set of choices



$$V_A(x) = c_{1x} + \boxed{\min_y \left\{ V_B(y) \,\middle|\, x \;\&\; y \text{ consistent} \right\}}$$

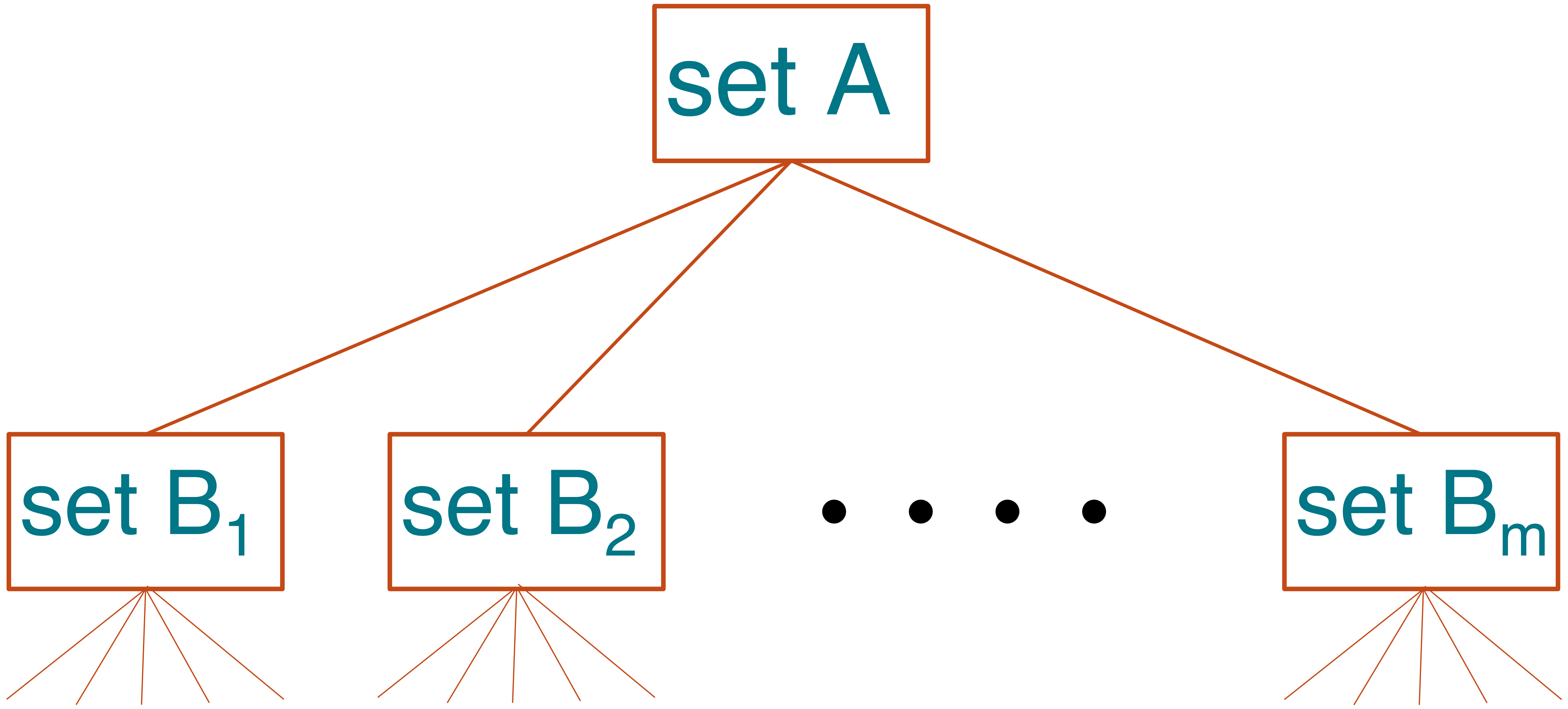
through pre-processing, computed in O(1)

Theorem:

If the underlying choice graph is a **line graph**, then rank aggregation LP can be solved via a DP in $O(n|\mathcal{M}|)$ **operations**

# of offer sets

Exploiting rational separation: DP for finding the best set of choices for a tree



$$V_A(x) = c_{1x} + \sum_{i=1}^{m} \min_{y_i} \left\{ V_{B_i}(y_i) \middle| \; x \;\&\; y_i \text{ consistent} \right\}$$

Theorem:

If the underlying choice graph is a **tree,** then rank aggregation LP can be solved via a DP in $O(n|\mathcal{M}|)$ **operations**

\# of offer sets

Exploiting rational separation: rank aggregation for a tree can also be solved with an LP

$$\min_{\boldsymbol{x}} \sum_{S \in \mathcal{M}} \sum_{i \in S} c_{i,S}\, x_{i,S}$$

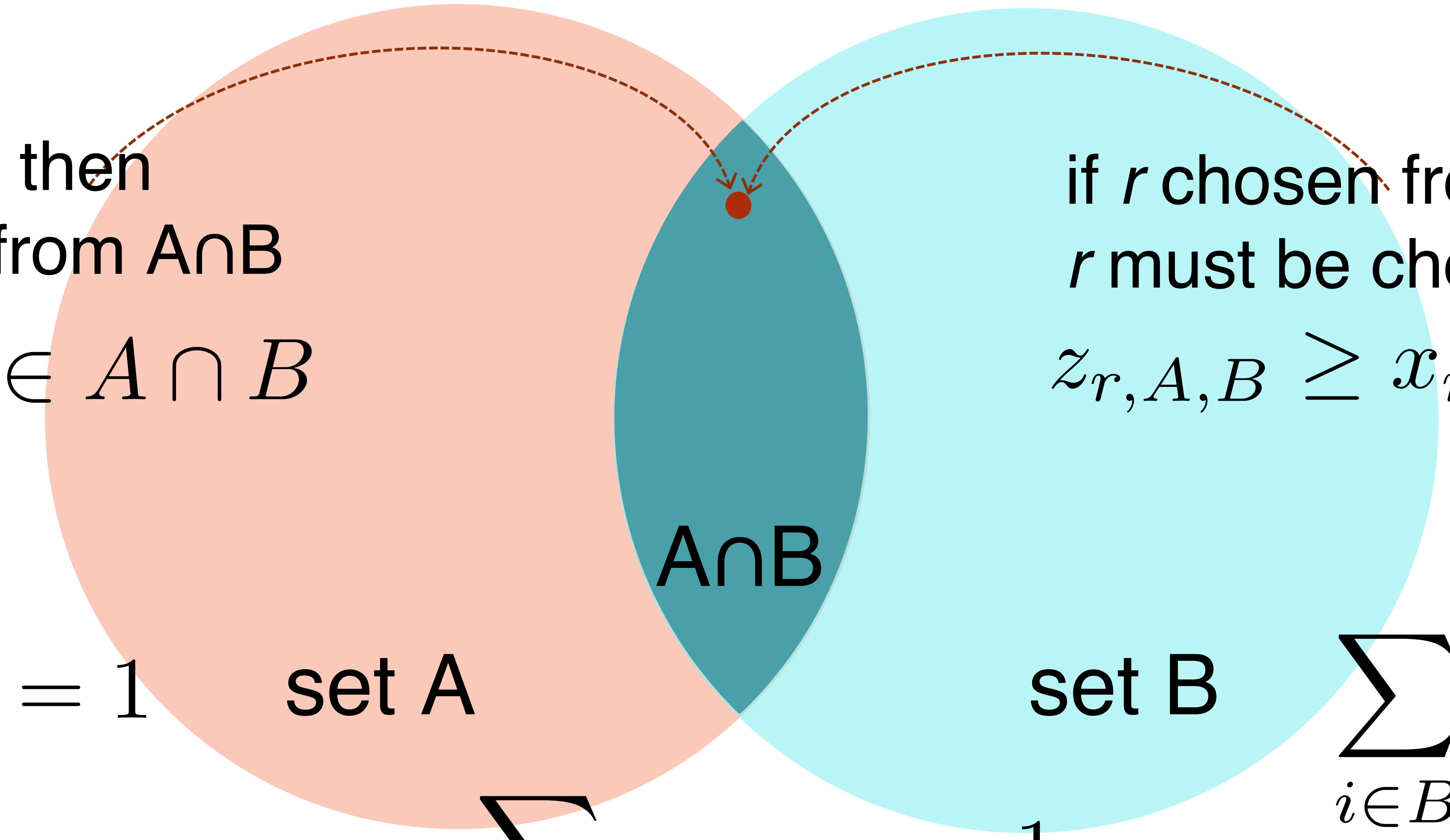
x_{iS} = 1 if i is chosen from S; 0 o.w.

if *r* chosen from *A*, then
r must be chosen from A∩B

$$z_{r,A,B} \geq x_{r,A} \;\forall\; r \in A \cap B$$

if *r* chosen from *B*, then
r must be chosen from A∩B

$$z_{r,A,B} \geq x_{r,B} \;\forall\; r \in A \cap B$$



$$\sum_{i \in A} x_{i,A} = 1$$

$$\sum_{i \in B} x_{i,B} = 1$$

$$\sum_{r \in A \cap B} z_{r,A,B} = 1$$

Exploiting rational separation: rank aggregation for a tree can also be solved with an LP

$$\min_{\boldsymbol{x}} \sum_{S \in \mathcal{M}} \sum_{i \in S} c_{i,S}\, x_{i,S}$$

x_{iS} = 1 if i is chosen from S; 0 o.w.

$$z_{r,A,B} \geq x_{r,A} \;\forall\; r \in A \cap B \quad \forall\; \{A, B\} \in \mathsf{E}$$

edge set of tree

$$z_{r,A,B} \geq x_{r,B} \;\forall\; r \in A \cap B \quad \forall\; \{A, B\} \in \mathsf{E}$$

$$\sum_{r \in A \cap B} z_{r,A,B} = 1 \quad \forall\; \{A, B\} \in \mathsf{E}$$

$$\sum_{i \in A} x_{i,A} = 1 \quad \forall\; A$$

$$\boldsymbol{x}, \boldsymbol{z} \geq 0$$

Normalization and non-negativity

\# variables $O\left(n|\mathsf{E}|\right)$

\# constraints $O\left(n|\mathsf{E}|\right)$

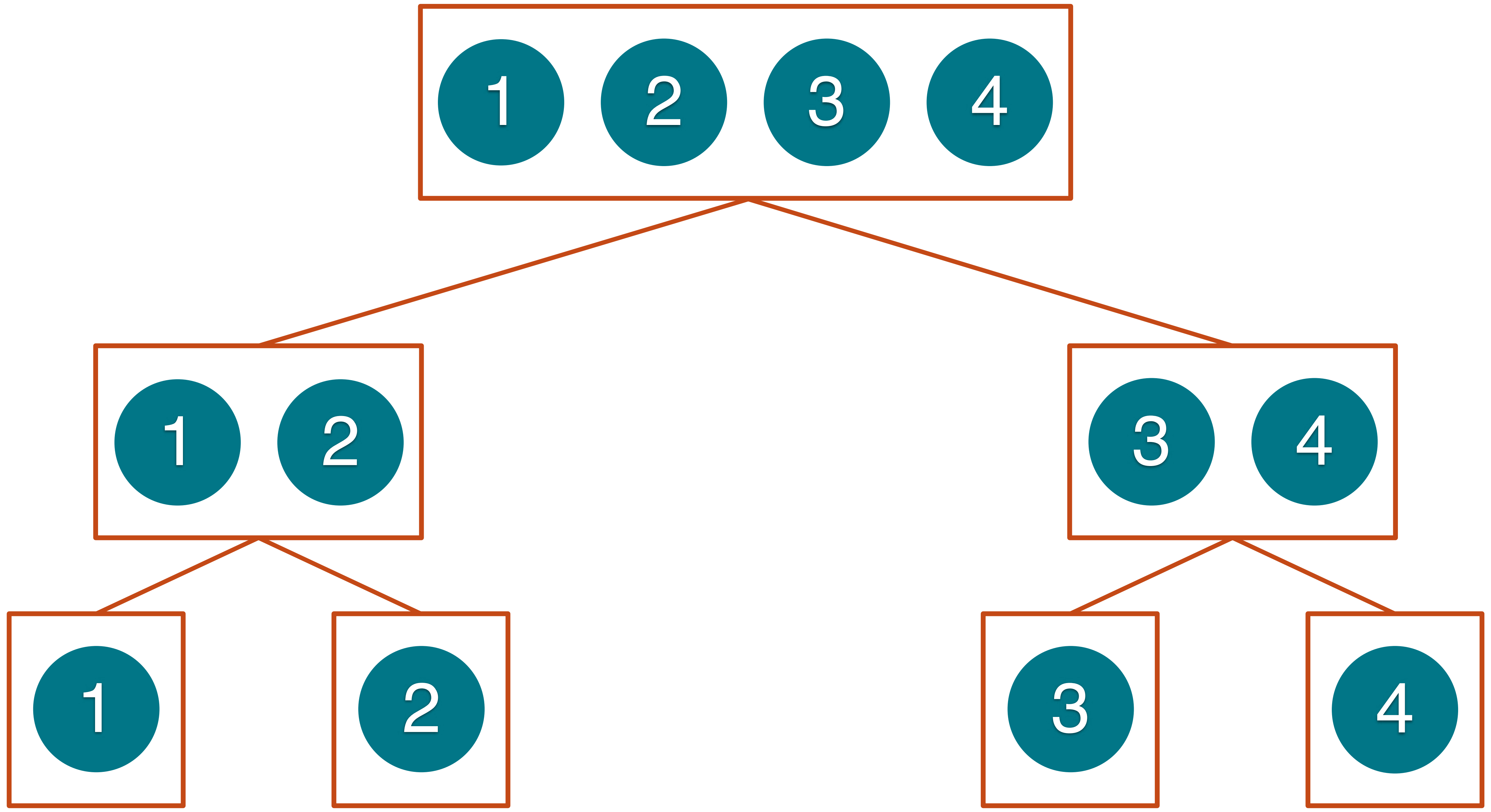
Theorem:

If the underlying choice graph is a **tree**, then the above LP is **integral**

Set collections with tree choice graph: nested, laminar, and differentiated collections

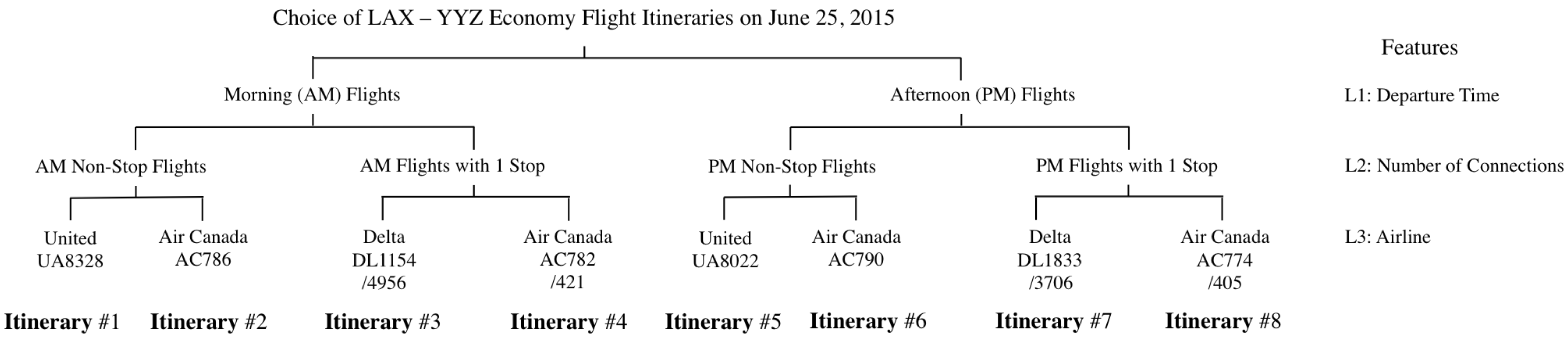
Nested collection: arises in revenue management   choice graph: line graph

Set collections with tree choice graph: nested, laminar, and differentiated collections

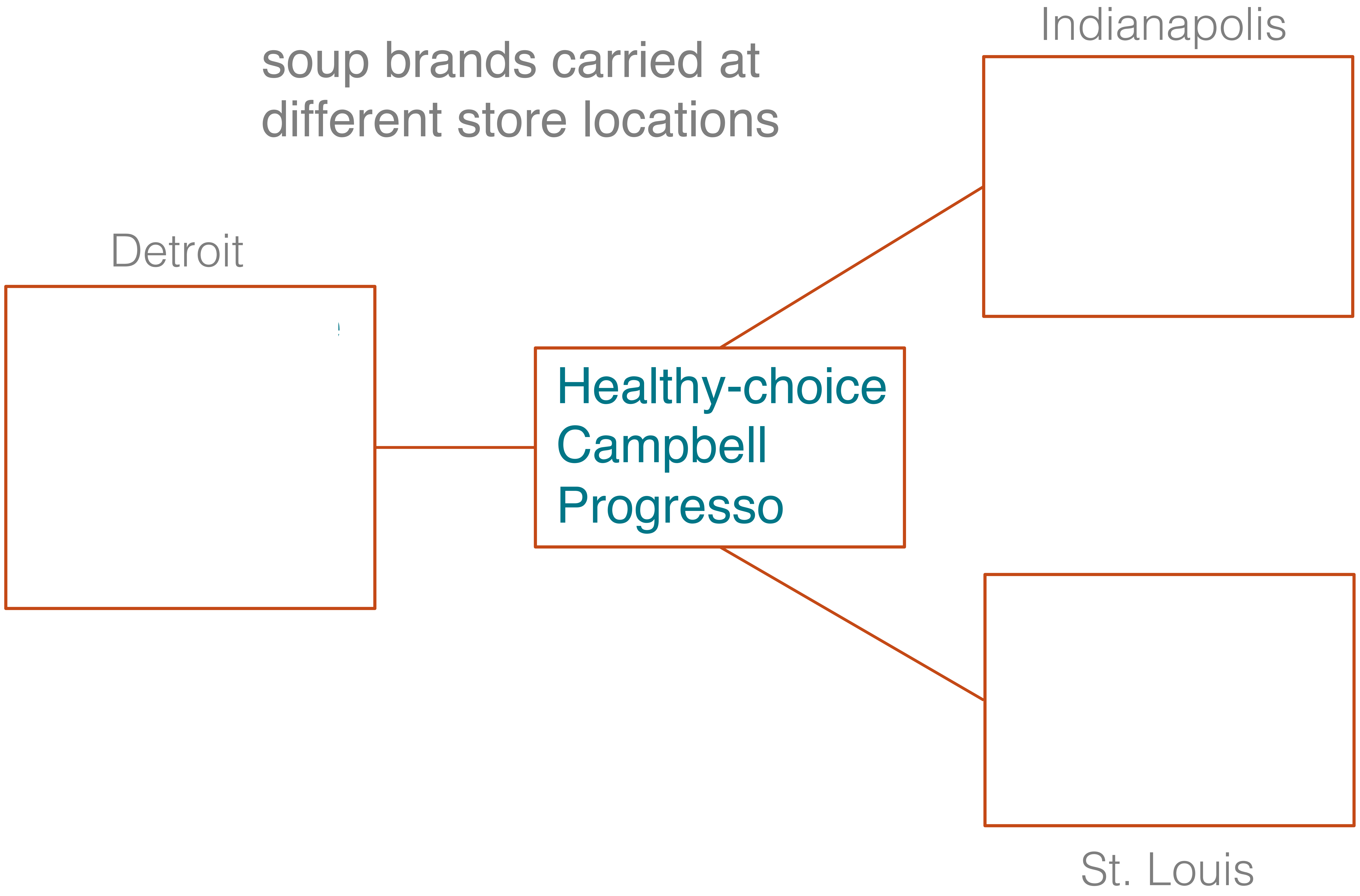


Laminar collection:
choice graph: tree

arises through elimination-by-aspects (EBA)



Set collections with tree choice graph: nested, laminar, and differentiated collections



Differentiated collection:

- core brands offered in all stores
- unique brands catering to local tastes

Star-shaped choice graph

Computational complexity for general choice graphs: scales exponentially in min(tree-width, choice-depth)

tree-width(tw) [equivalent definitions]

- size of largest vertex set in tree-decomposition of graph
- size of largest clique in the chordal completion of graph

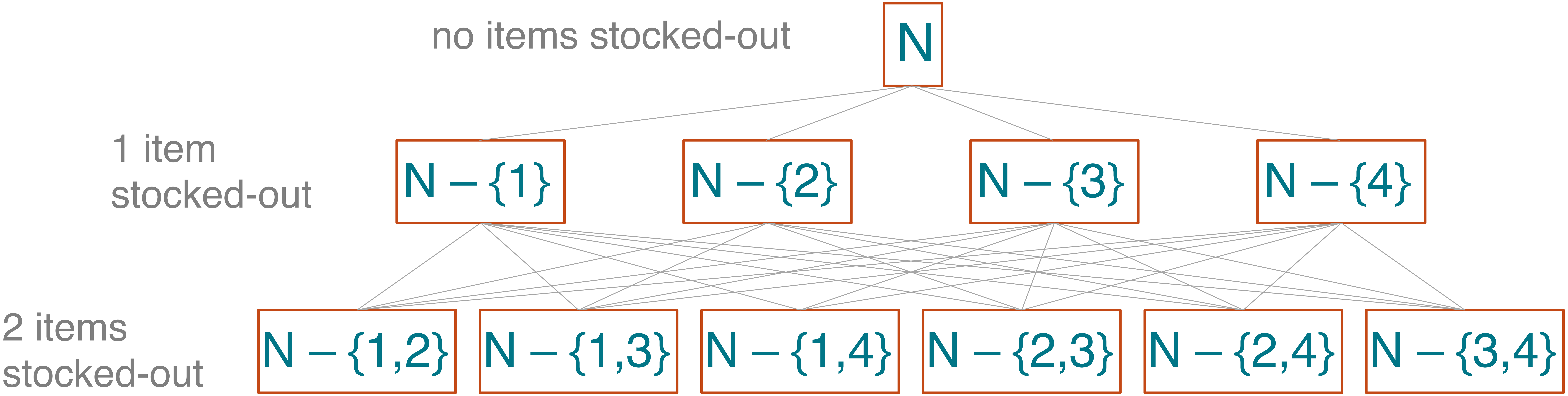
choice-depth (CD): "spread" of subsets in the same bag of tree-decomposition

$$\max_{b} \left\{ \left| \bigcup_{S \in \mathcal{X}_b} S \right| - \min_{A \in \mathcal{X}_b} |A| \right\}$$

Theorem:

Rank aggregation solved (using DP and LP) in $O(n^{\min(CD,\, tw)})$

k-deletion collection: at most k items are stocked out



Theorem:

choice-depth ≤ k ≤ n ≤ tree-width for k-del choice graph. Complexity = $O(n^k)$

The Limit of Rationality in Choice Modeling

1. Case study with IRI Academic Dataset
what is the ‘right’ choice model?

2. Complexity of computing LoR
rational separation, complexity in terms of graph properties

3. Summary/Conclusions
takeaway messages

Summary and Key Findings

KEY CONTRIBUTIONS

- characterization of complexity of computing LoR
- approx. techniques to compute LoR efficiently

MAIN TAKEAWAYS

Computational complexity of rank-aggregation

exponential in tree-width/choice-depth of the choice graph

Rationality loss

- larger than parametric loss
- suggests the need to go beyond rationality

FUTURE DIRECTIONS

- relate complexity lower bounds to tree-width/choice-depth
- general constructions of choice graphs
- general model classes that go beyond rationality