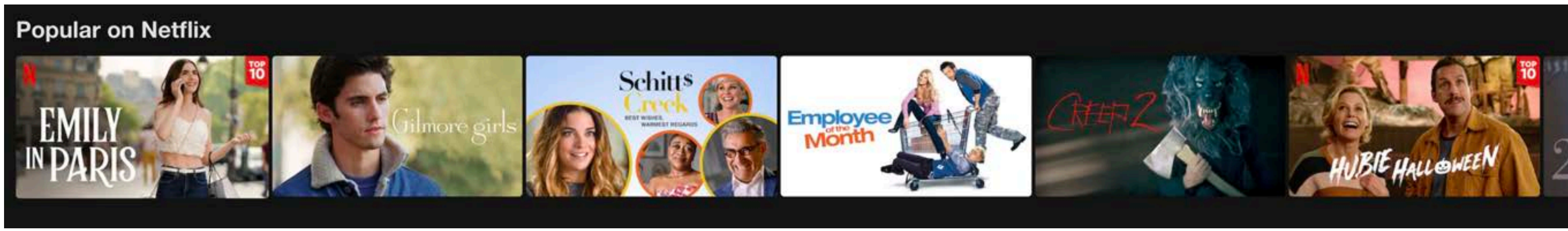
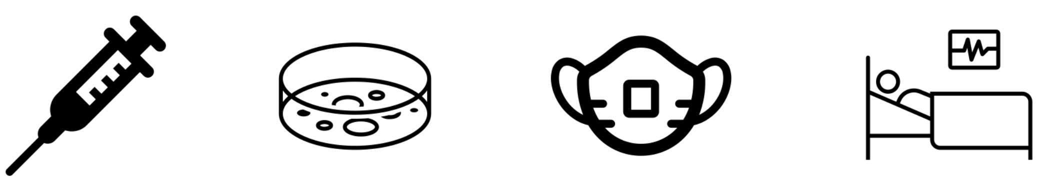
Thrust III: Data-driven modeling and experimental evaluations. Choice-set effects and habit formation

Austin Benson, Ali Jadbabaie, and Jon Kleinberg
Cornell University and MIT
MURI Meeting · October 22, 2020

Joint work with Kiran Tomlinson, Amir Tohidi, Anuran Makur, Katie Von Koevering, Devavrat Shah, and Dean Eckles.

Discrete choice models provide a framework for reasoning about decision making and habits.





Universe U of items.
Given subset C of alternatives.
Choose item x from C.

We have made progress in several directions in understanding behavior with choice models and data.

1. Can we learn biases (context effects) directly from data?
   New tractable and interpretable choice models reveal many context effects.
   Learning Interpretable Feature Context Effects in Discrete Choice. Tomlinson & Benson, arXiv, 2020.

2. Can we optimize choice sets to steer behavior?
   Making groups agree or disagree can be harder than promoting one decision.
   Choice Set Optimization Under Discrete Choice Models of Group Decisions. Tomlinson & Benson, ICML, 2020.

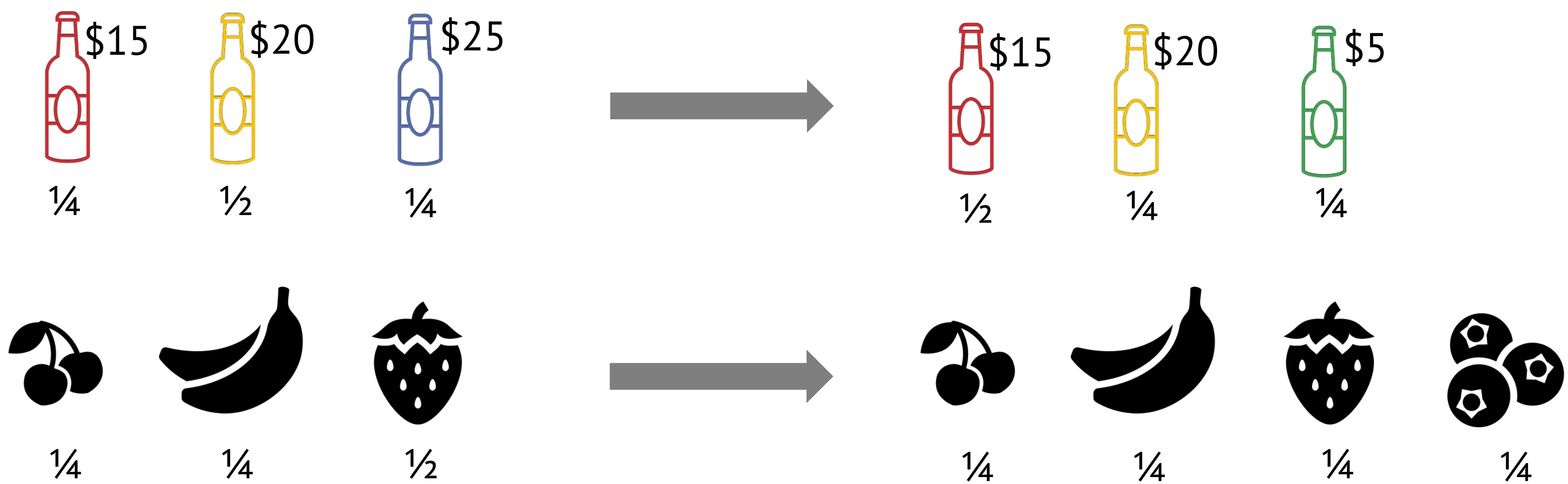
3. How do changes in choice sets let us understand habits?
   Store closures as a "shock" where choice sets change and habits are tested.
   Amir Tohidi, Ali Jadbabaie, and Dean Eckles, In Preparation.

4. How do choices and comparisons reveal latent habits or skills?
   Informal text online reveal habits; simple pairwise preferences reveal skill distributions.
   Frozen Binomials on the Web: Word Ordering and Language Conventions in Online Text. Van Koevering, Benson, & Kleinberg, WWW, 2020.
   Estimation of skill distribution from a tournament. Jadbabaie, Makur & Shah, NeurIPS, 2020.

Context effects in discrete choice are a well-known behavioral "bias".

[Huber+ 82; Simonson & Tversky 92; Shafir+ 1993; Trueblood+ 2013]

Changing the choice set *C* can affect relative preferences.



compromise effect

similarity effect

Key Question. Can we learn context effects from data?

We want to model context effects in an understandable and computationally tractable way.

$$\Pr(x \mid C) = \frac{s_x}{\sum_z s_z}$$

Multinomial logit / Bradley-Terry-Luce

Expressive and computationally reasonable

$$\Pr(x \mid C) = P_{x,C}$$

Universal logit / full specification

Behavioral economics & psychology. Stylized models for specific context effects, often theoretical, difficult to estimate from general data. [Tversky & Simonson 93; Rooderkerk+ 2011; Bruch+ 16]

Deep learning. Difficult to interpret, computationally questionable. [Pfannschmidt+ 19; Rosenfeld+ 20]

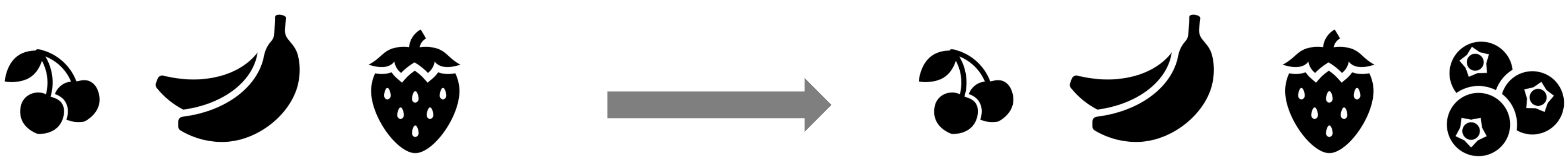
Our research. New models that yield intuitive, interpretable, and statistically testable context effects from passively collected choice data, with favorable computational properties. [Tomlinson & Benson 20]

A basic model is the multinomial logit (MNL) random utility model (RUM).

- **RUMs.** Item i has random utility V_i. Choose item in choice set C with largest utility.
- **MNL.** $V_i = U_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i$ are i.i.d. Gumbel $\longrightarrow \Pr(i \mid C) = \exp(U_i) / \sum_{j \in C} \exp(U_j)$.
  Parameterize as $U_i = \boldsymbol{\theta}^{\top} x_i$, where x are item covariates.

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

Satisfies independence of irrelevant alternatives



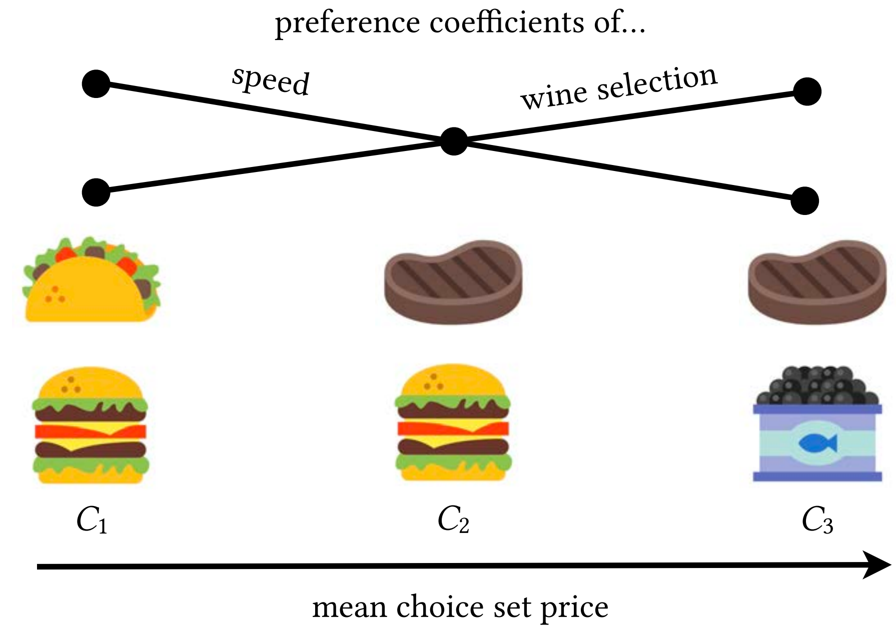
¼ ¼ ½ ¼ ¼ ¼ ¼

Violates independence of irrelevant alternatives.

We propose the linear context logit (LCL) model.

Tomlinson & Benson, ICML, 2020.

- **MNL.** $U_i = \boldsymbol{\theta}^\top x_i$, where x_i are item i covariates.
- **LCL.** $\boldsymbol{\theta} \longrightarrow \boldsymbol{\theta} + Ax_c$, where x_c is the mean feature vector in choice set. Still a RUM!



As the mean price increases, we might place less weight on service speed and more on wine selection.

Interpretable context effects.

- $A_{pq} > 0$. Choice sets with larger mean feature q increases preference of feature p.
- $A_{pq} < 0$. Choice sets with larger mean feature q decreases preference of feature p.

1. MNL is a special case, so we can use likelihood ratio tests for context effects.
2. Can derive from first principles (see paper).
3. Negative log likelihood is easy to optimize.
4. Identifiability conditions are simple.

The LCL reveals context effects in sushi choices.

Effect (q on p)	A_{pq}	$\overline{A}_{pq}$
oiliness on *oiliness*	-0.29	-0.30
popularity on *availability*	0.24	0.17
availability on *is maki*	0.23	0.07
is maki on *is maki*	0.22	0.15

- More oily sushi options makes oily options less appealing (similarity effect).
- More maki options makes maki more appealing (asymmetric dominance).

The LCL reveals context effects in hotel bookings.

Effect (q on p)	A_{pq}	$\bar{A}_{pq}$
location score on *price*	-0.38	-0.10
on promotion on *price*	0.20	0.13
review score on *price*	-0.12	-0.13
star rating on *price*	0.12	0.17

- Cheaper options are preferred when there is a good location.
- Hotels on discount lead to more willingness to pay.

The LCL often significantly fits the data better and has better out-of-sample predictions.

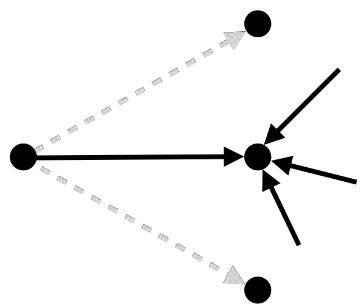
	MNL NLL	LCL NLL	MNLL MRR	LCL MRR
district-smart	3426	3343*	0.40	0.39
expedia	839505	837620*	0.39	0.37
sushi	9821	9774*	0.27	0.27
car-a	1702	1697	0.36	0.35
car-alt	7393	6697*	0.29	0.27

*Significant likelihood-ratio test vs. MNL ($p < 0.001$)

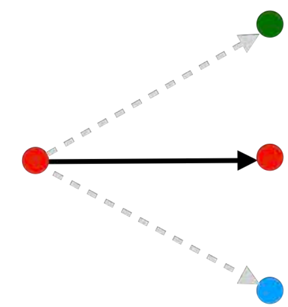
NLL = negative log likelihood
MRR = mean relative rank

We also used the LCL to study which factors drive edge formation in social networks.

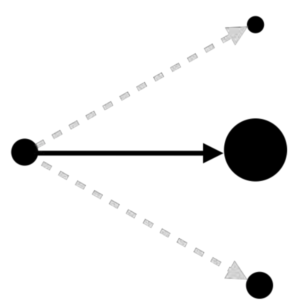
Preferential attachment
(Barabási & Albert, *Science* 1999)



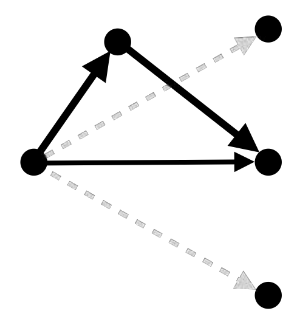
Homophily
(McPherson et al., *Annual Review of Sociology* 2001)
(Papadopoulos et al., *Nature* 2012)



Fitness
(Bianconi & Barabási, *Europhysics Letters* 2001)
(Caldarelli et al., *Physical Review Letters* 2002)

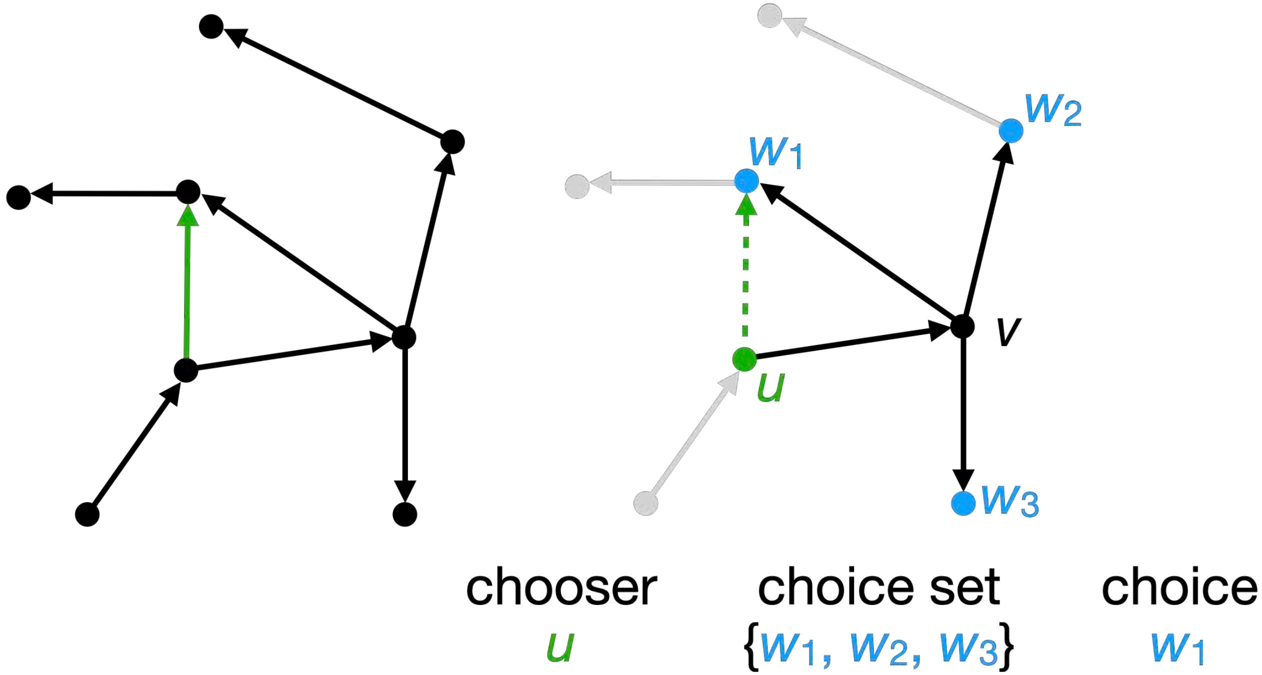


Triadic closure
(Rapoport, *Bulletin of Mathematical Biophysics* 1953)
(Jin et al., *Physical Review E* 2001)



See *Choosing to grow a graph: Modeling network formation as discrete choice.*
Jan Overgoor, Austin R. Benson, and Johan Ugander. Proc. of the World Wide Web Conference, 2019.

We studied how people choose to close triangles.



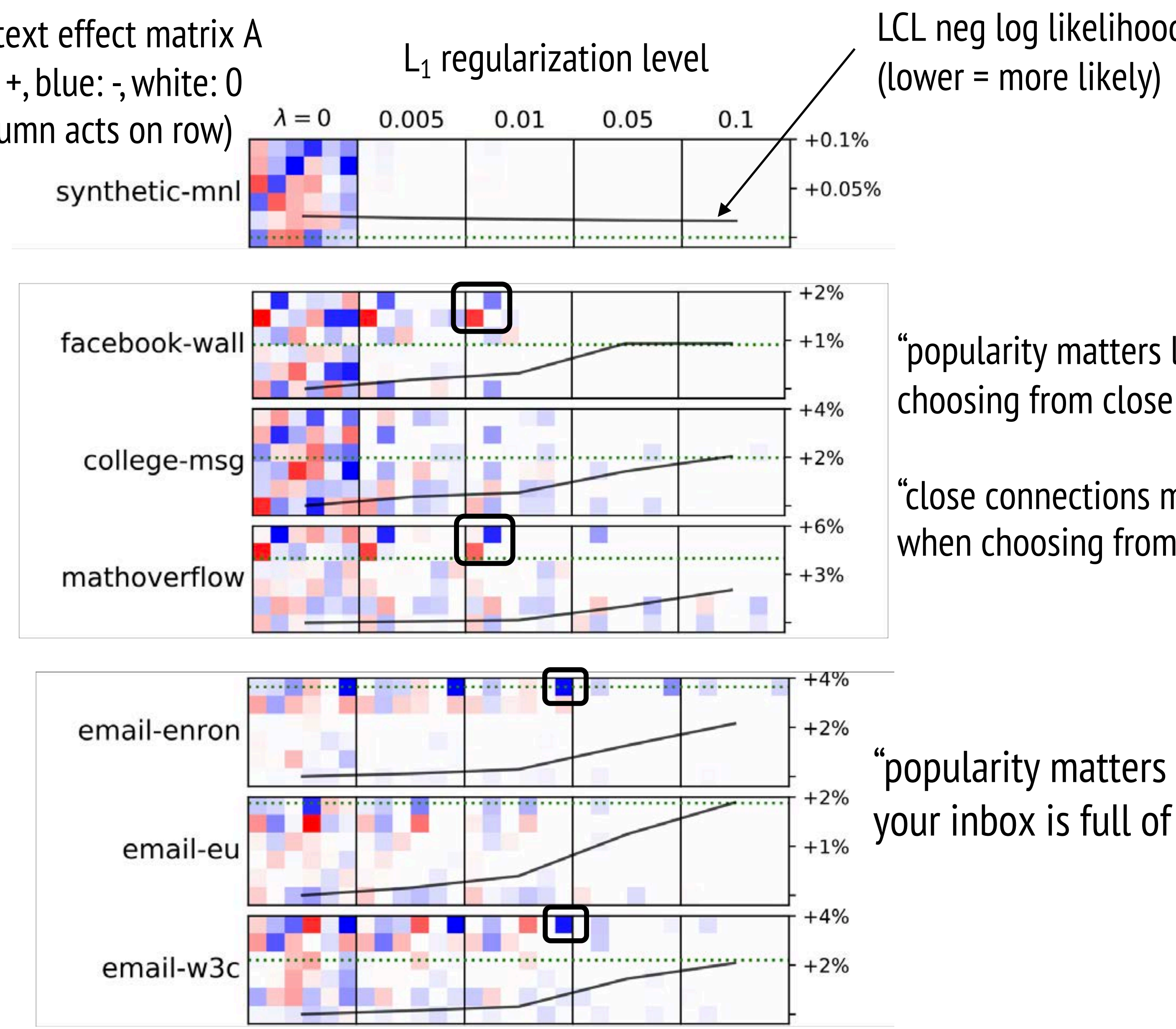
We collected a sequence of timestamped directed edges (i, j, t) from 13 social networks.

Features.

- in-degree of w
- # shared neighbors of u, w
- weight of edge $w \longrightarrow u$
- time since last edge into w
- time since last edge out of w
- time since last $w \longrightarrow u$ edge

The LCL reveals interpretable context effects.

context effect matrix A
red: +, blue: -, white: 0
(column acts on row)



Features (left-right, top-bottom).

1. in-degree of w
2. # shared neighbors of u, w
3. weight of edge w $\longrightarrow$ u
4. time since last edge into w
5. time since last edge out of w
6. time since last w $\longrightarrow$ u edge

"popularity matters less when choosing from close connections"

"close connections matter more when choosing from the popular"

"popularity matters less when your inbox is full of recent emails"

We can learn behavioral biases or context effects directly from choice data.

1. The linear context logit (LCL) model makes it easy to identify interpretable context effects from raw choice data.
2. We have other models in our paper that help distinguish context effects from population heterogeneity.
3. **Key takeaway.** Passively collected data on decisions can be used to identify possible behavioral biases.

Future work.

- How can we disentangle population heterogeneity and context effects?
- How do recommender systems influence context effects?

Learning Interpretable Feature Context Effects in Discrete Choice.
K. Tomlinson and A. R. Benson. arXiv, 2020.

github.com/tomlinsonk/feature-context-effects

Data-driven modeling and experimental evaluations: Choice-set effects and habit formation

1. Can we learn biases (context effects) directly from data?
   New tractable and interpretable choice models reveal many context effects.

2. Can we optimize choice sets to steer behavior?
   Making groups agree or disagree can be harder than promoting one decision.
   Kiran Tomlinson and Austin Benson, ICML 2020.

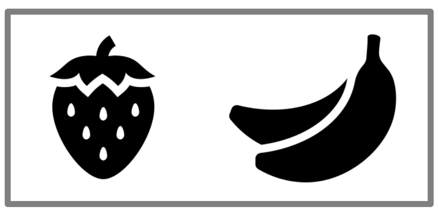
3. Can changes in choice sets let us understand habits?
   Store closures as a "shock" where choice sets change and habits are tested.

4. How do choices and comparisons unknowingly reveal habits and skill?
   Informal text online reveal many tendencies; simple pairwise preferences reveal skill distributions.

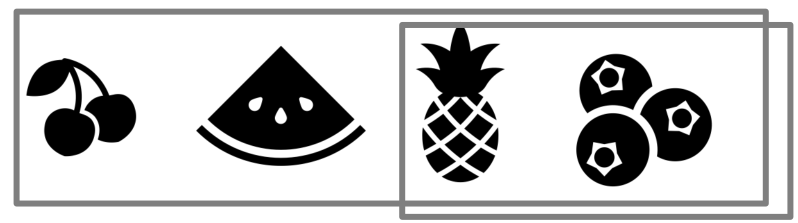
So far we have looked at predictive models to understand decisions and possible biases.

- Given models, can we understand how decisions are subject to influence?
- Here we consider influence by simply changing the choice set C.

base choice set C



Other alternatives in universe

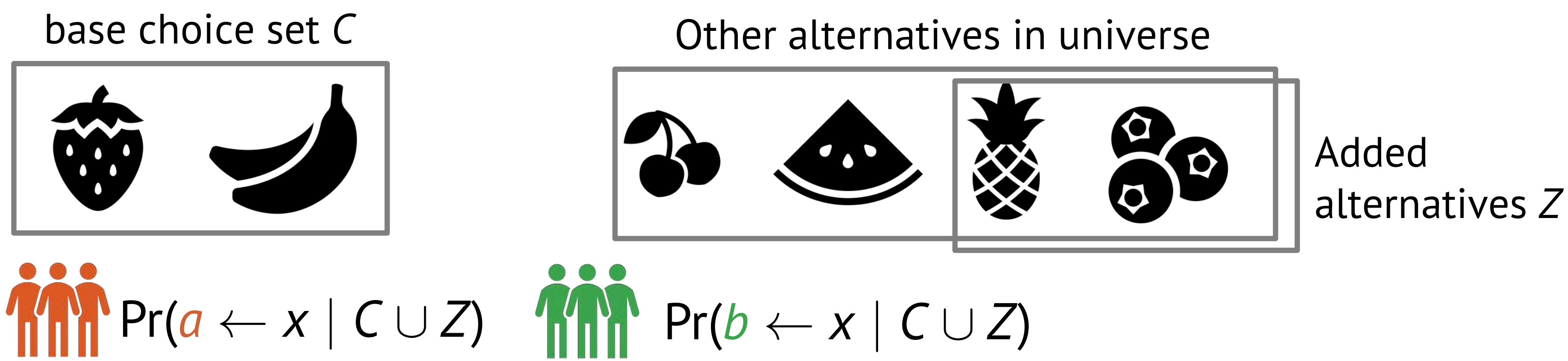


Added alternatives Z

New choice set ⟶ new decision problem.

- For analysis, assume we know Pr(choose item x | choice set C), and ignore features.

We aim to modify preferences just with new context.



1. Can we make the groups agree on the base choice set?
   Agreement. $\min_Z \sum_{\{a,b\} \subseteq A} \sum_{x \in C} |\Pr(a \leftarrow x \mid C \cup Z) - \Pr(b \leftarrow x \mid C \cup Z)|$

2. Can we make the groups disagree on the base choice set?
   Disagreement. $\max_Z \sum_{\{a,b\} \subseteq A} \sum_{x \in C} |\Pr(a \leftarrow x \mid C \cup Z) - \Pr(b \leftarrow x \mid C \cup Z)|$

3. Can we promote a target decision from the base choice set?
   Promotion. $\max_Z |\{a \subset A \mid \Pr(a \leftarrow x^* \mid C \cup Z) > \Pr(a \leftarrow x \mid C \cup Z), x \in C, x \neq x^*\}|$

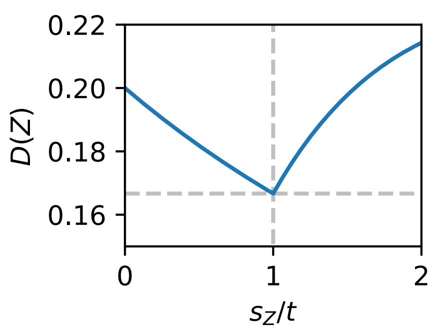
Making just two groups agree or disagree is hard.

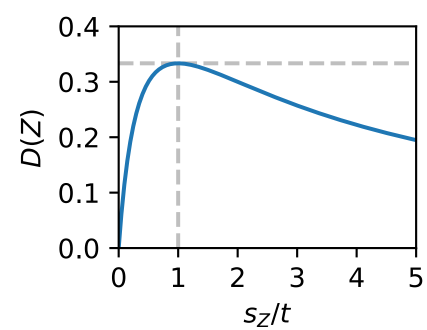
Multinomial logit (MNL). Choice probability is proportional to the exponential of utility.

$$\Pr(a \leftarrow x \mid C \cup Z) = \frac{\exp(u_{ax})}{\sum_{z \in C \cup Z} \exp(u_{az})} \qquad \Pr(b \leftarrow x \mid C \cup Z) = \frac{\exp(u_{bx})}{\sum_{z \in C \cup Z} \exp(u_{bz})}$$

Theorem. Under MNL, the agreement and disagreement problems are NP-hard, even with just two groups where everyone has equal utilities on the items not in C.

Proof. By reduction from Partition.





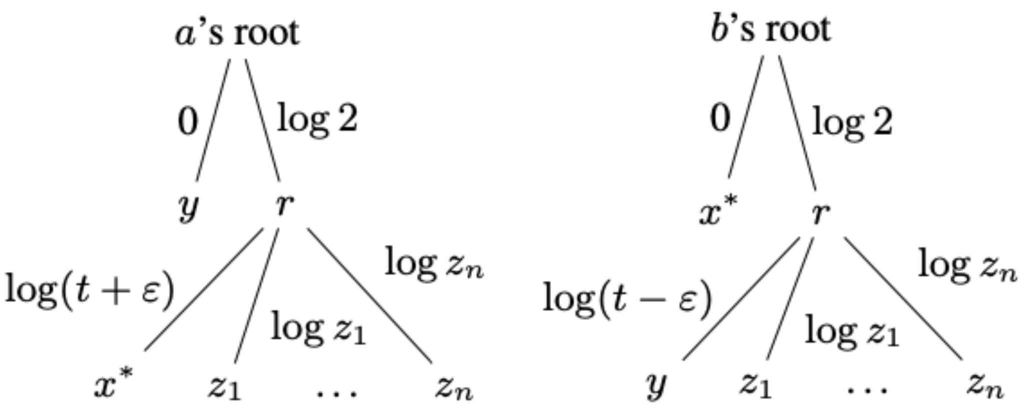
Corollary. Also NP-hard under models that subsume MNL, such as nested logit (NL), elimination-by-aspects (EBA), and context-dependent utility model (CDM).

Promotion is impossible for MNL and hard for models with context effects.

$$\Pr(a \leftarrow x \mid C \cup Z) = \frac{\exp(u_{ax})}{\sum_{z \in C \cup Z} \exp(u_{az})}$$

Adding new alternatives Z does not change relative preferences within C.

Theorem. Promotion is NP-hard in NL, EBA, and CDM.



- Nested logit (NL)
- Reduction from subset sum.
- Need to add z_i to promote x^* for group b, but this makes x^* less attractive to group a.

However…

Theorem. Promotion becomes tractable when for NL a and b have the same tree.

Natural restrictions make promotion easy but keep agreement and disagreement hard.

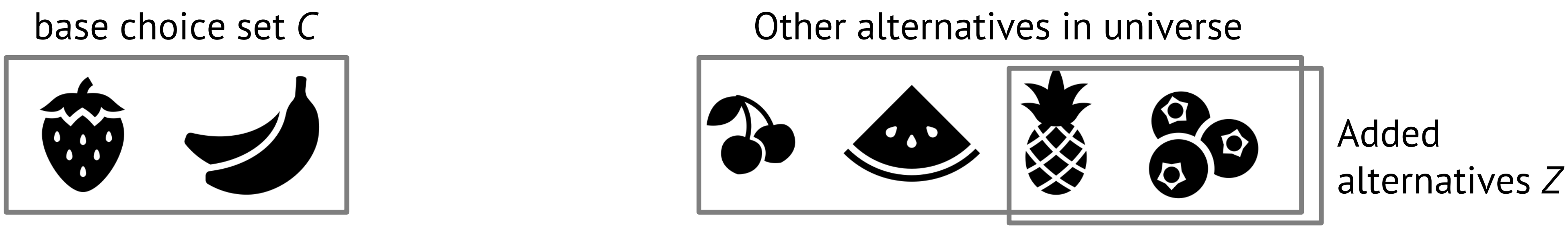
Promotion is tractable when

- in nested logit (NL) [McFadden 74]
  the groups have the same decision tree structure
- in elimination-by-aspects (EBA) [Tversky 1972]
  the outside alternatives share no aspects with the promoted item
- in context-dependent utility model (CDM) [Seshadri+ 19]
  the outside alternatives have the same context effect “pulls”

… and all of these restrictions keep agreement and disagreement NP-hard.

Takeaway. Encouraging consensus or sowing discord among groups is more difficult than promoting a particular choice.

We have approximation algorithms for NP-hard problems, based on FPTAS algorithms for subset sum.



$\Pr(a \leftarrow x \mid C \cup Z)$ $\Pr(b \leftarrow x \mid C \cup Z)$

Can we make the groups agree on the base choice set?

Agreement. $\min_Z \sum_{\{a,b\} \subseteq A} \sum_{x \in C} |\Pr(a \leftarrow x \mid C \cup Z) - \Pr(b \leftarrow x \mid C \cup Z)|$

1. $\boldsymbol{\varepsilon}$-additive approx. for Agreement.
2. Greedy: build Z one by one based on largest gains.

The approximation algorithm and greedy algorithm tends to give different solutions in practice.

SFWork data.

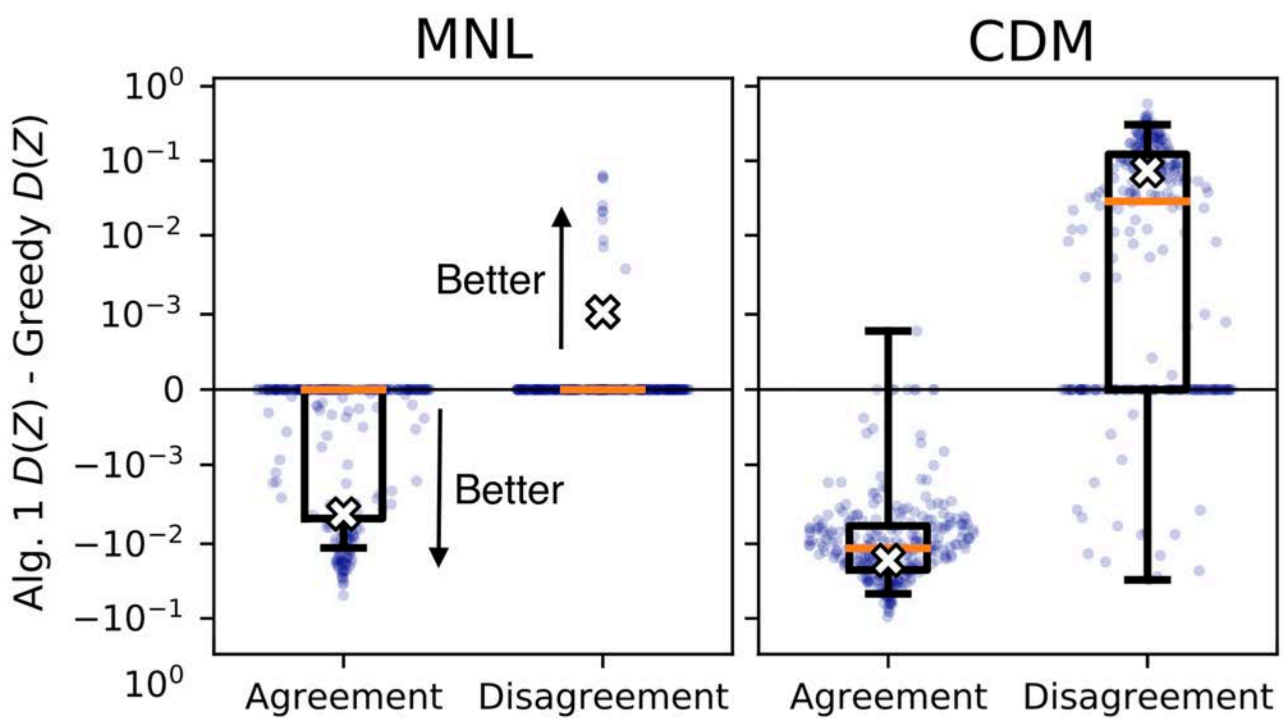
- Choice sets = transportation options for getting to work.
- Two groups = live in city center or live in suburbs.
- From data, estimate $\Pr(a \leftarrow x \mid C \cup Z)$ and $\Pr(b \leftarrow x \mid C \cup Z)$ (here, using CDM).
- For a given choice set C, optimize Z.

C = { drive alone, public transit }
Z = { carpool } is given by greedy algorithm for Agreement.
Z = { bike, walk } is given by approx. algorithm (and is optimal).

The approximation algorithm outperforms greedy on two-item choice sets.

Allstate data.

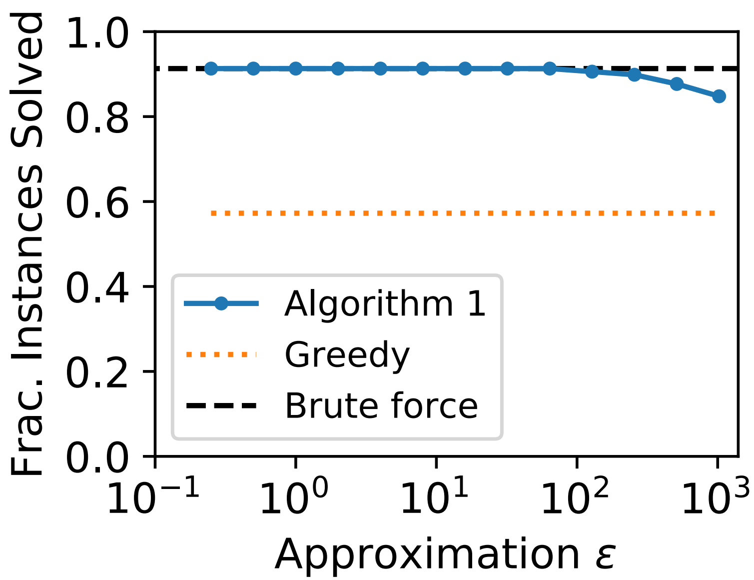
- Choice sets = insurance policies.
- Two groups = homeowners and non-homeowners.
- Goal = encourage agreement or disagreement.

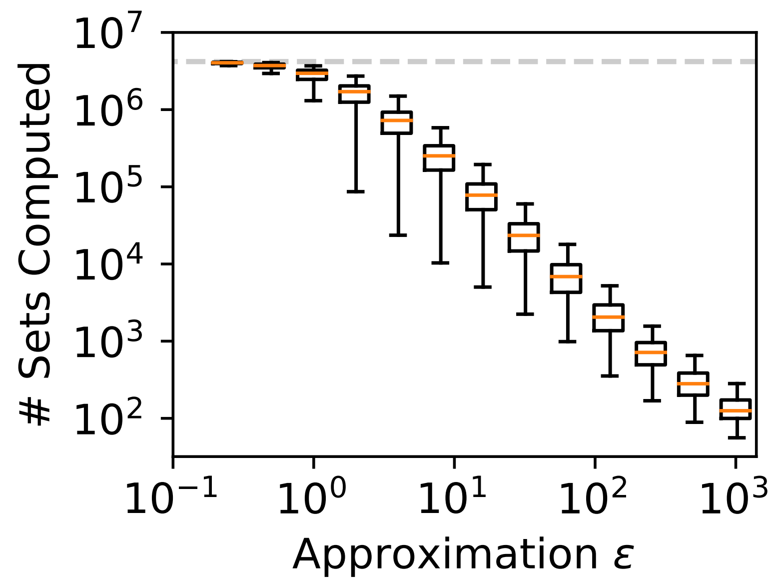


In practice, the approximation algorithm outperforms the theoretical guarantees.

Allstate data.

- Choice sets = insurance policies.
- Two groups = homeowners and non-homeowners.
- Goal = promote a certain policy under CDM decision model.





Modifying choice sets is a subtle intervention that can have major effects.

1. We can influence how groups agree or disagree simply by changing the choice set.
2. In general, NP-hard to maximize / minimize consensus or promote a particular decision.
3. Under some model restrictions, promotion becomes easier.
4. Even with NP-hard problems, approximation algorithms work well in practice.

Future work.

- How do we designing new alternatives to steer influence rather than just selecting more from the current universe?

Choice Set Optimization Under Discrete Choice Models of Group Decisions.
K. Tomlinson & A. R. Benson. Proc. of the International Conference on Machine Learning, 2020.
github.com/tomlinsonk/choice-set-opt

Data-driven modeling and experimental evaluations: Choice-set effects and habit formation

1. Can we learn biases (context effects) directly from data?
   New tractable and interpretable choice models reveal many context effects.

2. Can we optimize choice sets to steer behavior?
   Making groups agree or disagree can be harder than promoting one decision.

3. Can changes in choice sets let us understand habits?
   Store closures as a "shock" where choice sets change and habits are tested.
   Amir Tohidi, Ali Jadbabaie, and Dean Eckles.

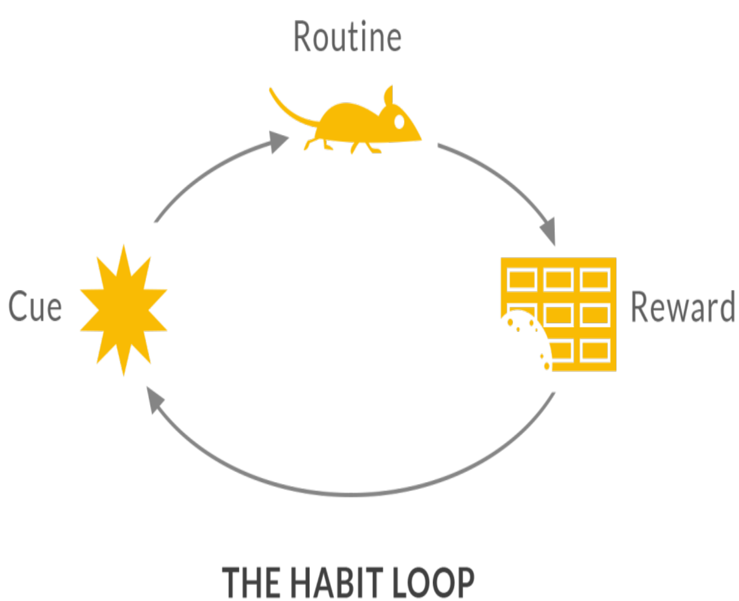
4. How do choices and comparisons unknowingly reveal habits and skill? Simple pairwise preferences reveal skill distributions; informal text online reveal many tendencies.

Choice sets and behavioral biases: Habit formation

Main Question:

How do formed habits affect our choices?

How formed habits by repeated purchases at grocery stores affect our behavior?



https://gdt.stanford.edu/the-habit-loop/

Habit formation

- Habits are defined as ***automatic processes*** that are learned from ***repeated responses*** and are triggered through various ***contextual or mental cues***.
- **Habits are very common!**
  More than 1/3 of students' behavior could be classified as habits done daily in the same location and context [Wood, Quinn, Kashy, 2002]
  Similar results for user behavior on the Web [Benson+ 16, 18]
- **+** By forming habits we free up our mental resources for more important deliberative tasks.
- **-** Habits are persistent, even when the outcome structure is changed and the behavior is no longer optimal.

State Dependence or Inertia

- Guadagni and Little [1983] added a loyalty variable in consumers' utility function to capture reliance on past behavior

$$U_k = X_k \beta + h_k \alpha + \epsilon_k$$

where X_k is the set of features for brand k (e.g., price) and h_k is exponentially weighted sum of past purchases.

- With Markov assumption added, loyalty is called **State-Dependence**. Lots of evidence of for state-dependence consumer behavior even with more flexible models [Keane 97; Dube, Hitsch, & Rossi 10].

Habit Discontinuity

- Context changes in people's personal, social, or professional circumstances provides opportunities for conscious, planned behavior change [Schwanen, Banister, & Anable, 12; Verplanken & Wood, 06]
- **Our idea.** use the disruption due to store closures to measure the effect of habits and distinguish from brand-specific loyalty.

https://www.ballinaadvocate.com.au/subscriptions/premium-offer/

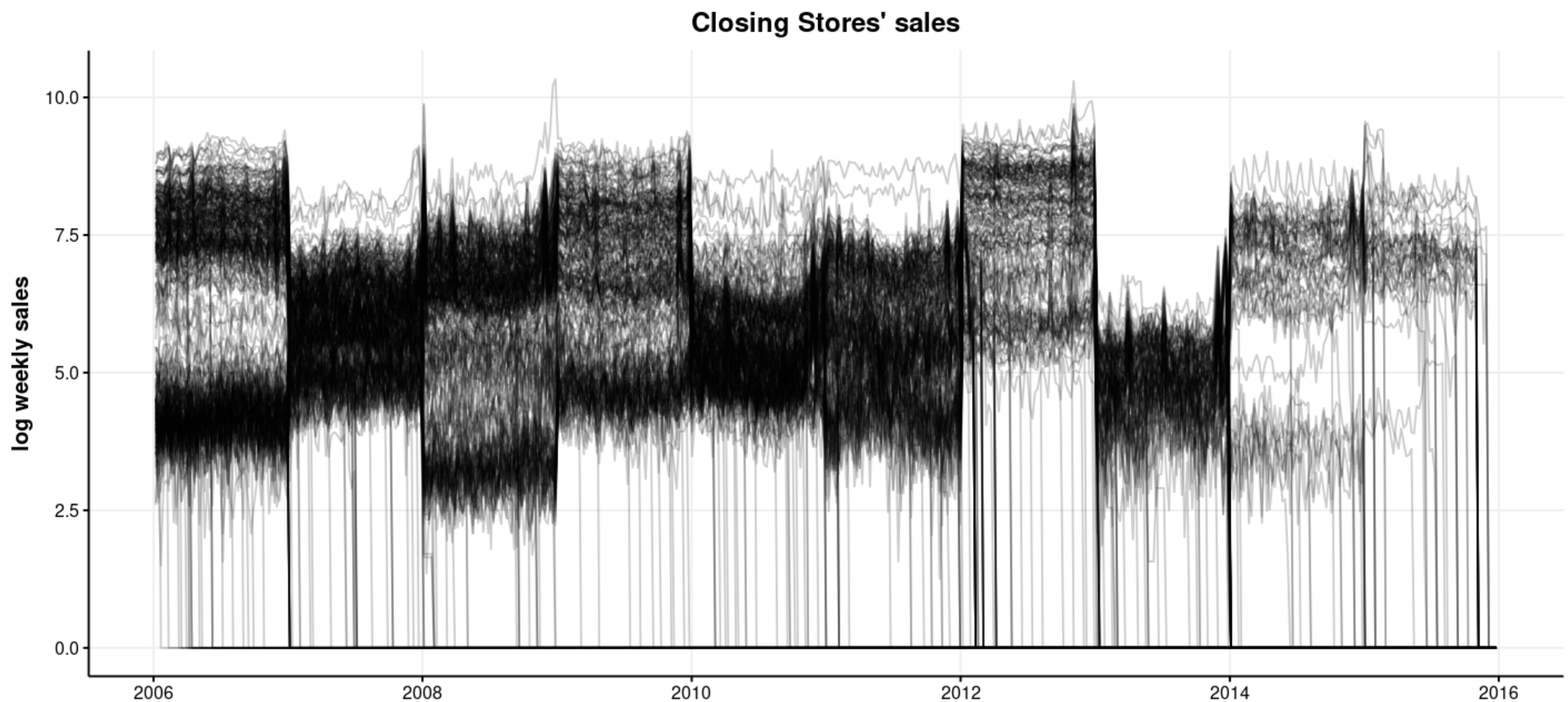
Nielsen retail scanner data

1. **Retail Scanner data.**
   weekly pricing, volume, and store merchandising conditions generated by approximately 35,000 participating grocery, drug, mass merchandiser, and other. From 2006 to 2018.
2. **Consumer Panel data.**
   Tracks purchases of a panel of 40,000–60,000 households. The data describes when, where, and what the panelists purchase, and at what price.

By combining the two datasets, one can find what each consumer purchased, and what alternatives they had.

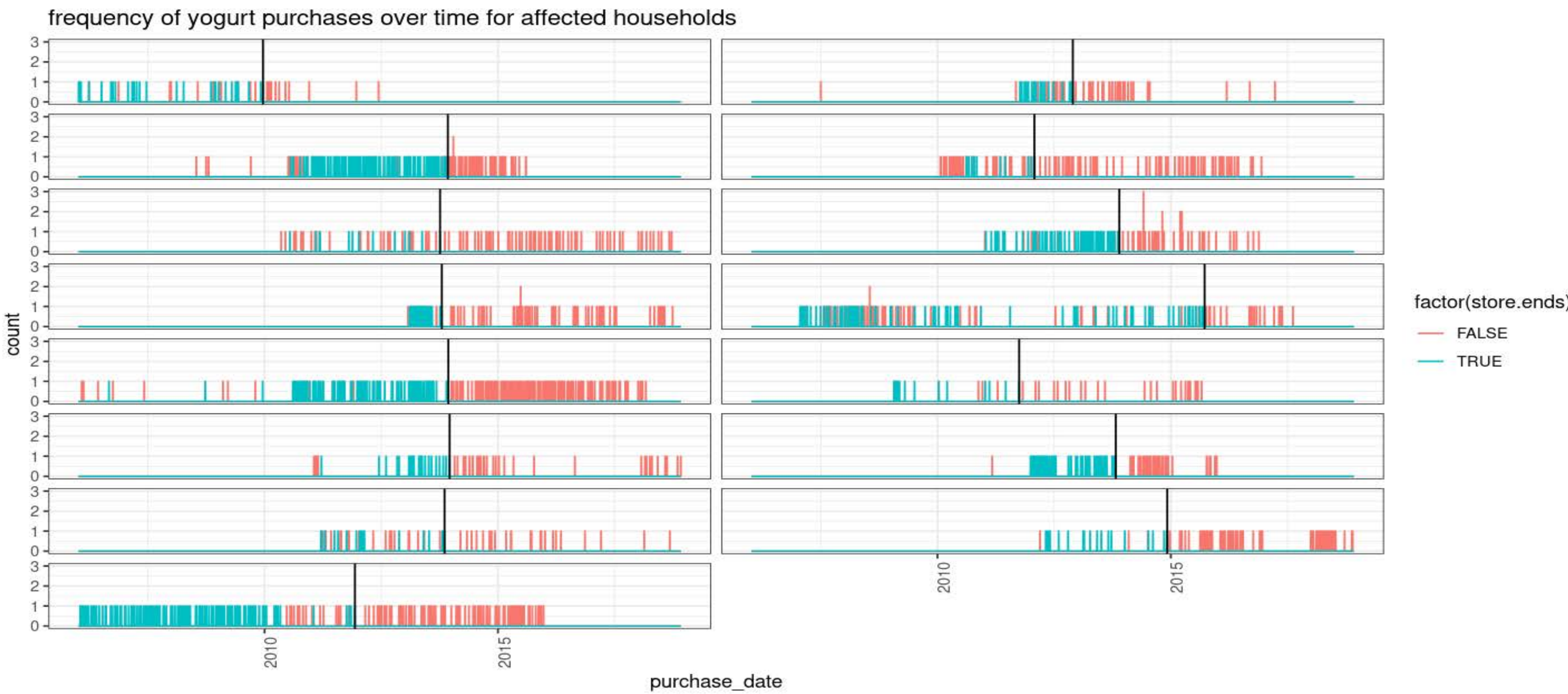
Identifying Store Closings

We identify store closure points by computing the aggregate weekly sales over time and find the ones that become zero at some point.



Selection of affected households

We need to choose households who did a substantial part of the shopping in the closing stores, so they have formed strong habits and a significant part of their behavior is disrupted.



Mixed Logit model (MIXL)

$$\boldsymbol{U_{i,b,t} = \alpha_b + X'_{i,b,t}\beta_i + I[period = 1]X'_{i,b,t}\ \gamma_i + \epsilon_{i,b,t}} \qquad \boldsymbol{(\beta_i, \gamma_i) \sim MVN(\mu, \Sigma)}$$

$\boldsymbol{U_{i,b,t}}$: the utility of consumer **i**, for brand **b**, at trip **t**

α_b: brand specific intercepts

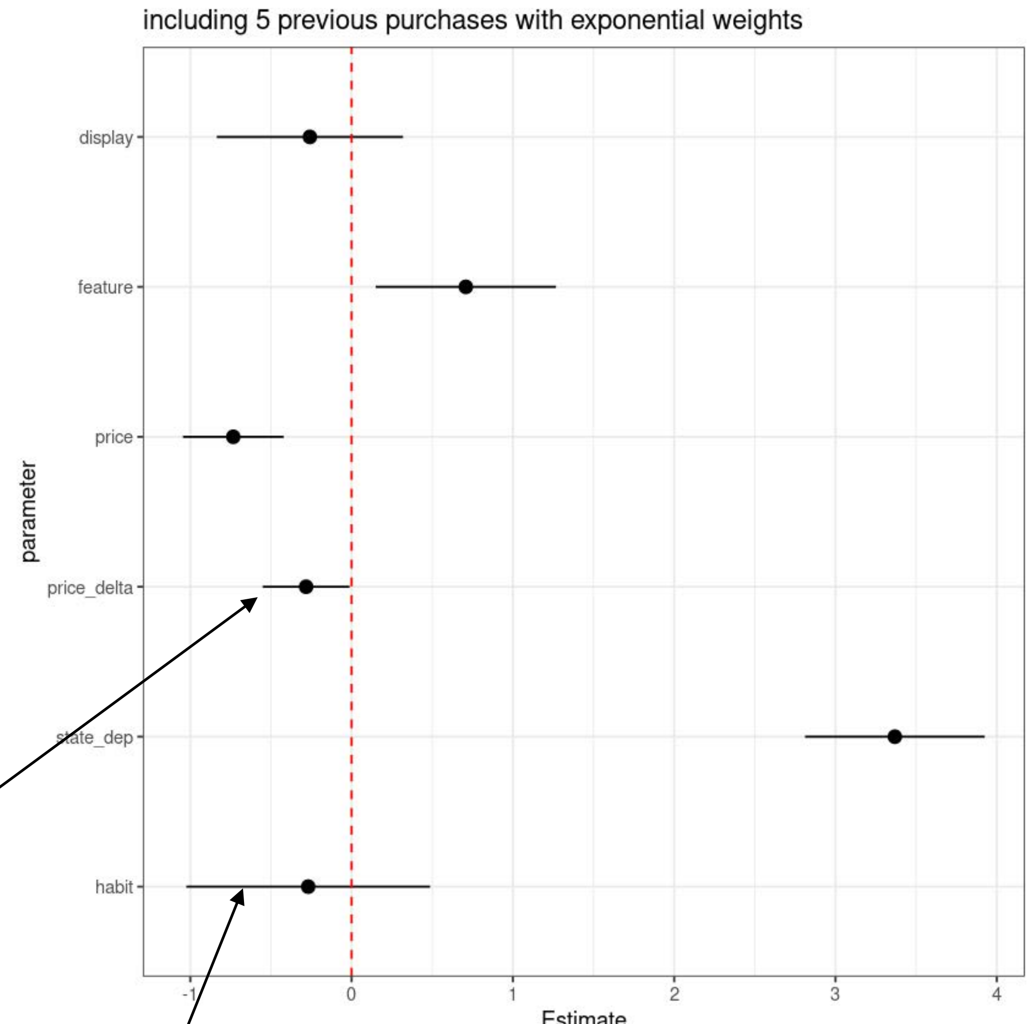
$\boldsymbol{X_{i,b,t}}$: price, display ads indicator, featured product indicator, state dependence

Before closure: β_i after closure: $\beta_i + \gamma_i$

$$P_i = \int \prod_{t=1}^{T_i} \prod_{b=1}^{B} \frac{e^{U_{i,b,t}}}{\sum_{b=1}^{B_{i,t}} e^{U_{i,b,t}}}\ f(\beta_i)\, d\beta_i$$

The parameters are estimated using simulated maximum likelihood.

Results for yogurt category

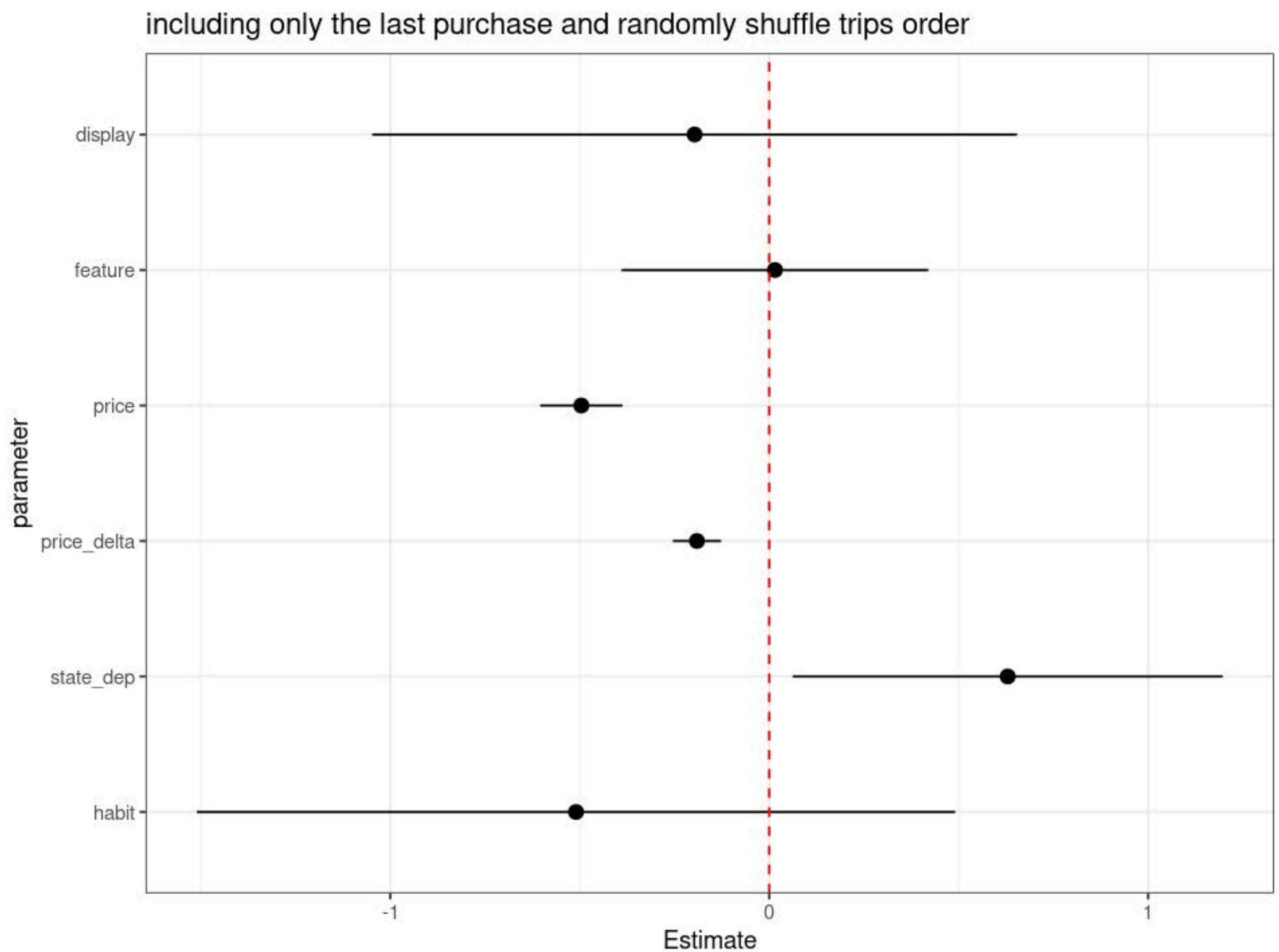


Consumers become more sensitive to price changes

Habits matter less as consumers become more variety-seeking

Robustness

- If the measured state dependence is not an artifact of unobserved heterogeneity, it shouldn't remain for reshuffled trips [Dube, Hitsch, & Rossi 10]



Discussion and Future work

Takeaways.

1. We provide a quantitative framework to measure strength of habits in consumer brand inertia.
2. The results point out the importance of context on behavior, and can inform policy makers and people who seek behavioral change.

Future work.

- Expanding the analysis to other product categories.
- Applying the model to other datasets, particularly disruptions due to Covid-19.
- Adjusting the model to include multiple choices at the same trip. [Benson et al., 2018]
- Study the effect of order of temporal purchases.

Data-driven modeling and experimental evaluations: Choice-set effects and habit formation

1. Can we learn biases (context effects) directly from data?
   New tractable and interpretable choice models reveal many context effects.

2. Can we optimize choice sets to steer behavior?
   Making groups agree or disagree can be harder than promoting one decision.

3. Can changes in choice sets let us understand habits?
   Store closures as a "shock" where choice sets change and habits are tested.

4. How do choices and comparisons unknowingly reveal information?
   Informal text online reveal habits; Simple pairwise preferences reveal skill distributions;
   Katie Van Koevering, Austin Benson, and Jon Kleinberg. Web Conference 2020.
   Ali Jadbabaie, Anuran Makur, and Devavrat Shah Neurips 2020. Oral presentation (top 100 / 10,000).

Decisions and choices are not always obvious or consciously deliberate, and they reveal habits.

"The Conference assembles scholars, researchers, policymakers, practitioners, and end-users..."

red and blue
peaches and plums
jogging or running
bakeries or grocers
still and quiet
quickly and loudly

fire and ice
bread and butter
husband and wife
mother and child
men and women
more or less

These are examples of *binomials:* pairs of words separated by "or" or "and"

[Bolinger 1962; Cooper & Ross 1975; Allen 1987; Sobkowiak 1993; Benor 2006; Mollin 2012; Haggerty 2015]

Given two words (or lists in general), we must ***choose*** their order

Our idea. Study word ordering at scale with 15 years of Reddit data.

There are changing habits in binomials.

1. “fire and ice” is classically considered a “frozen binomial” (always in that order) but we saw an increase in the ordering “ice and fire” over time… why?

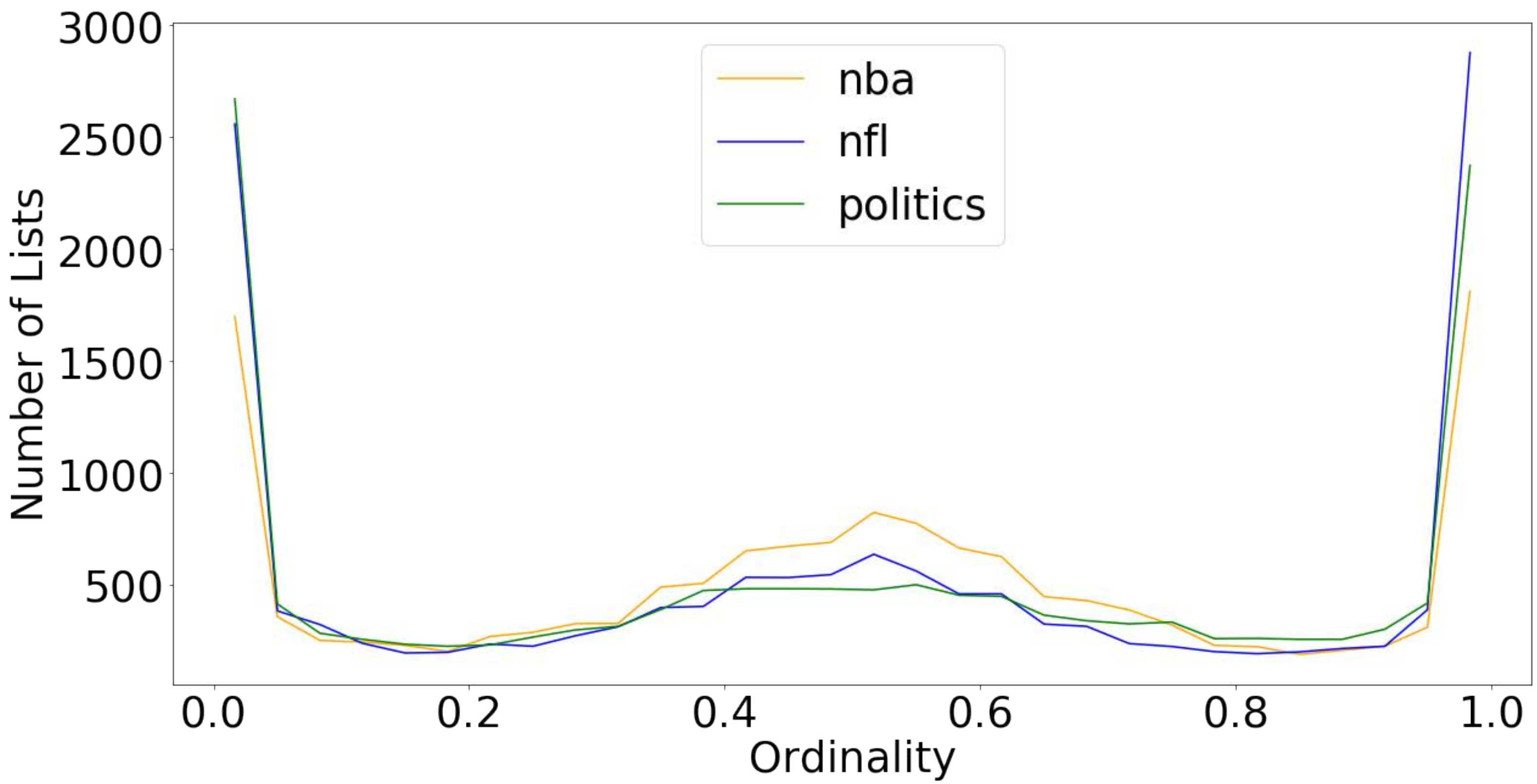
   Answer: Game of Thrones (A Song of Ice and Fire)

2. In 2008, “son and daughter” was frozen (always in that order). By 2018, it was “daughter and son” 36% of the time… why?

   Answer? Changing culture?

Word ordering at scale is revealing.

"brother and sister": 273
"sister and brother": 42
Ordinality (w/r/t alphabetical ordering) = 273 / (273 + 42) ≈ 0.87.



Many (nearly) frozen binomials, but many orders are close to 50/50.

We found evidence for a "proximity principle", where words closer to your community are listed first.

- My team before your team
  "Bulls and Clippers"

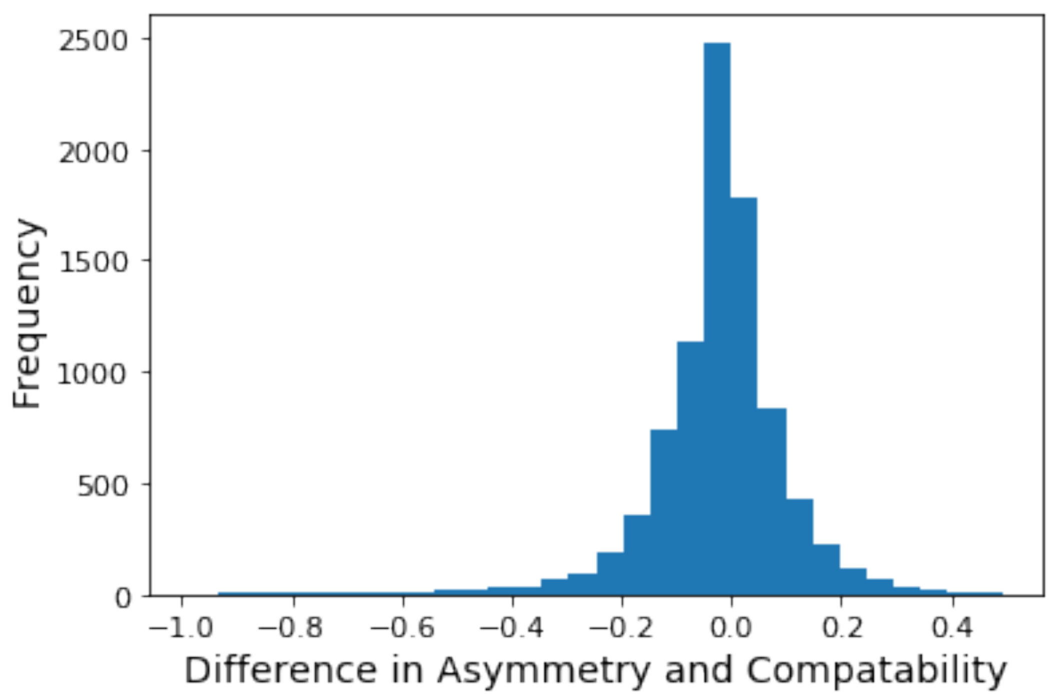
- My subreddit before your subreddit
  "r/nba and r/hockey"

- My ideology before your ideology
  "democrats and republicans"

Multinomials also have structure.

- Many trinomials only appear in a single ordering [a, b, c] with most other cases appearing as [b, a, c] (third word in the list remains the same)
  "Fraud, Waste, and Abuse": 34
  "Waste, Fraud, and Abuse": 20

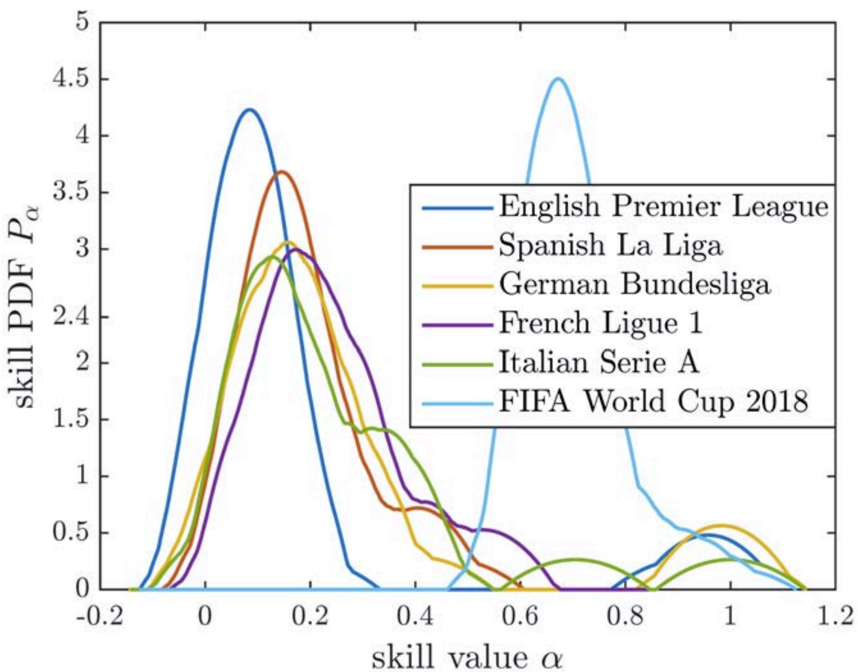
- Independence of irrelevant alternatives in trinomials?

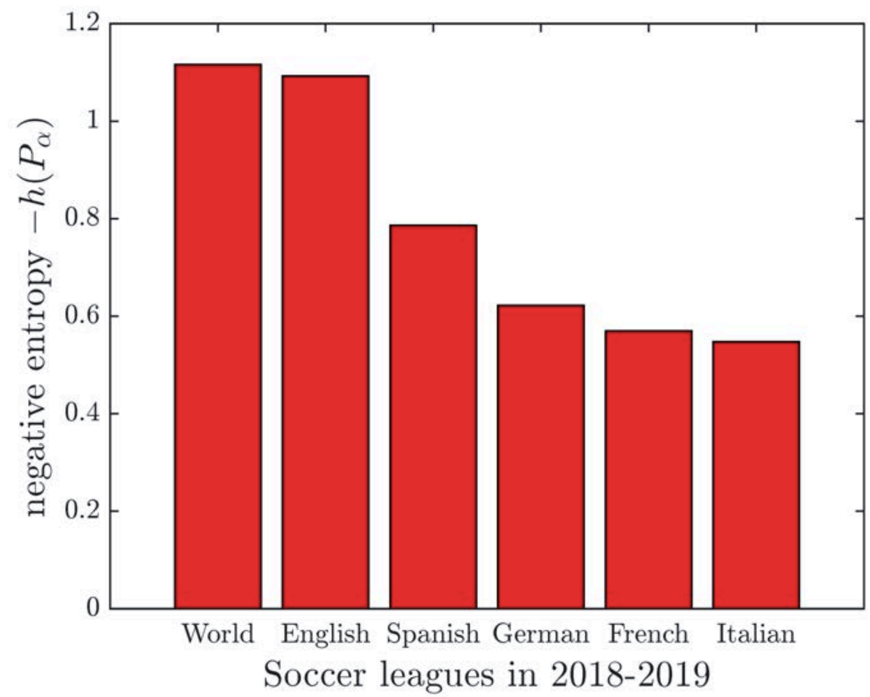


For pairs of words appearing in both binomials and trinomials, the ordering frequencies are roughly the same.

Discrete choice models can be used to learn skill distributions

Key Question. Can we estimate the distribution of skill levels based on partially observed pair-wise comparisons (e.g., as in tournament data)?





1. Suppose there is an unknown skill density of interest.
2. n agents play tournament with skill levels drawn from this density.
3. Outcomes of pairwise games determined by Bradley-Terry-Luce or MNL model.
4. Negative entropy of skill density measures overall skill score.
5. Approach can be used to measure distribution of skills/quality for any set of choices with pairwise comparisons

Discrete choice models can be used to learn skill distributions from tournament data.

Key Question. Can we estimate the distribution of skill levels based on partially observed win-loss tournament data?

Minimax optimal algorithm using spectral learning and kernel density estimation.
Estimation of Skill Distribution from a Tournament. Jadbabaie, Makur & Shah, NeurIPS, 2020.
[selected for oral spotlight presentation (top 100 out of 10,000 submission)]

Estimation Algorithm.

1. Estimate skill levels of agents using **rank centrality** [Negahban-Oh-Shah 2017].
2. Construct density estimator using **Parzen-Rosenblatt method** with appropriate choice of bandwidth.

Minimax Estimation Results.

Estimation problem	Loss function	Upper bound	Lower bound
Smooth skill density	mean squared error	$\tilde{O}(n^{-1+\varepsilon})$	$\Omega(n^{-1})$ [IK82, Tsy09]
BTL skill parameters	relative ℓ^{∞}-norm	$\tilde{O}(n^{-1/2})$ [CFMW19]	$\tilde{\Omega}(n^{-1/2})$
BTL skill parameters	ℓ^{1}-norm	$O(n^{-1/2})$ [CFMW19]	$\tilde{\Omega}(n^{-1/2})$

With decision models and choices or preferences, we can reveal information that is often considered hidden.

1. Our habits in how we order lists (on social networks, in email, in publications or technical documents, etc.) can have signal.
2. Skill distributions are learnable with relatively few comparisons.
3. The availability of large-scale data online offers opportunity for analysis.

Future work.

- Learn underlying skill distributions with broader choice models and varying-size choice sets.
- Word orderings in other contexts such as the news.

- Estimation of skill distribution from a tournament.
  A. Jadbabaie, A. Makur & D. Shah, Advances in Neural Information Processing, 2020.
- Frozen Binomials on the Web: Word Ordering and Langauge Conventions in Online Text.
  K. Van Koevering, A. R. Benson, and J. Kleinberg. Proc. of The Web Conference, 2020.

github.com/ktvank/Frozen-Binomials

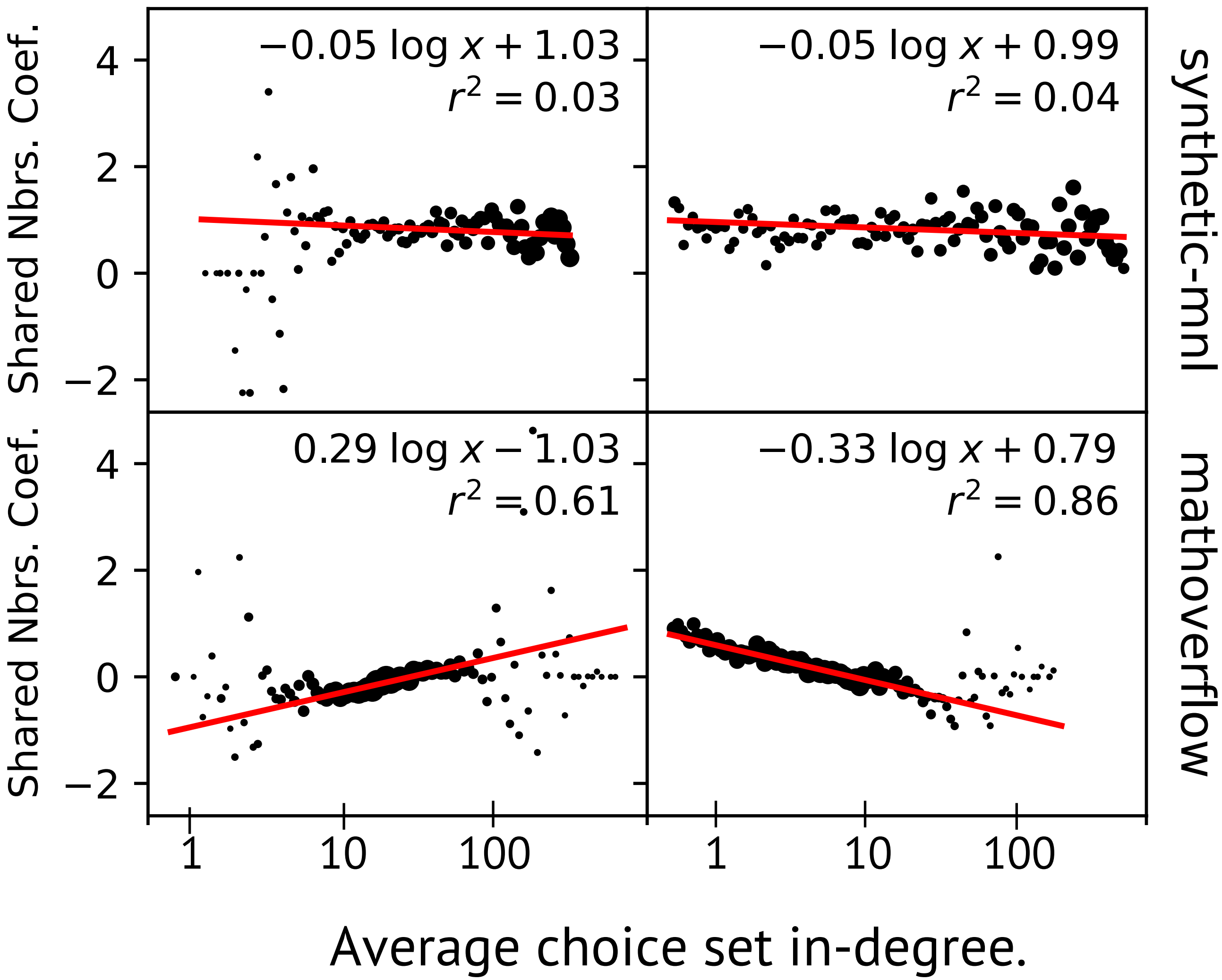
Thrust III: Data-driven modeling and experimental evaluations

Using explicit choice data, historical outcome data, or revealed preference data, we can learn behavioral biases, susceptibility, habits, culture, and skill.

Future directions.

- Causal inference techniques for disentangling population heterogeneity and "true" behavioral biases via context effects.
- Designing new alternatives or comparisons to elicit more information, or to get more information with less data.
- Data-driven modeling of habits for multiple selection such as product bundling.
- More on the process of groups arriving at a single decision (might need new experimental data).

We see linear context effects in real-world data.



Synthetic data
⟶ no context effects.

Commenting network
⟶ linear context

Binomials and word ordering have a long history.

[Jespersen 1905; Scott 1913; Bolinger 1962; Cooper & Ross 1975; Allen 1987; Sobkowiak 1993; Benor 2006; Mollin 2012; Haggerty 2015]

1. **Phonological.** Short before long, vowel quality, rhythm, rhyming.
2. **Semantic.** Men before women, now before later, here before there.
3. **Context.** Frequent before infrequent, recency, primacy.

Focus has been on ***frozen*** binomials: pairs of words in the same order.

Word ordering can be leaking information about relative importance, cultural conventions, unconscious bias, etc.

Our idea. Study choices in word ordering in large-scale text on the Web, using the Reddit online social network.

The habitual consumer

- Consumers tend to buy the same brands of products across different shopping episodes [Seetharaman 04; Wood & Neal 09]
- In a study, for 67% of the products consumers went straight to the product they seemed to be looking for, grabbed it and put it in their shopping basket or cart, without making any comparison among products. [Machín et al., 2020]

https://news.nd.edu/news/impulse-purchases-rise-the-longer-shoppers-are-in-store/