Recommender System and Social Experiment

Wei Ai
University of Maryland
aiwei@umd.edu

4/19/2024

Today's plan

- Recommender System as Information Filtering
- Recommender System as Applied Machine Learning
- (Break)
- Recommender System as Contextualized Nudge
- Recommender System in Social Experiment
- (Discussion)

Short vs. Long Term Information Need

- Short-term information need (**Ad hoc retrieval**)
  - “Temporary need”, e.g., info about used cars
  - Information source is relatively static
  - User “pulls” information
  - Application example: library search, Web search
- Long-term information need (**Filtering**)
  - “Stable need”, e.g., new data mining algorithms
  - Information source is dynamic
  - System “pushes” information to user
  - Applications: news filter, recommender systems

Examples of Information Filtering

- News filtering
- Email filtering
- Movie/book recommenders
- Literature recommenders
- And many others …

Content-based Filtering vs. Collaborative Filtering

- Basic filtering question: Will user *U* like item *X*?
- Two different ways of answering it
  - Look at characters of what U likes ➔ **characterize X** ➔ **content-based filtering**
  - Look at who likes X ➔ **characterize U** ➔ **collaborative filtering**
- Can be combined (hybrid filtering)

Collaborative filtering is also called
"Recommender Systems"

IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

What is Collaborative Filtering (CF)?

- Making filtering decisions for an individual user based on the judgments of other users
- Inferring individual's interest/preferences from that of other similar users
- General idea
  - Given a user u, find similar users $\{u_1, \ldots, u_m\}$
  - Predict u's preferences based on the preferences of $u_1, \ldots, u_m$

IES Institute of Education Sciences · W AmplifyLearn.AI · UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS · UNIVERSITY OF OREGON

Collaborative Filtering: Assumptions

- Users with a common interest will have similar preferences
- Users with similar preferences probably share the same interest
- Examples
  - “interest is IR” ➔ “favor SIGIR papers”
  - “favor SIGIR papers” ➔ “interest is IR”
- Sufficiently large number of user preferences are available

CF: Intuitions

- User similarity (*Paul Resnick* vs. *Rahul Sami)*
  - Suppose Paul and Rahul viewed similar movies in the past six months …
  - If Paul liked the movie, Rahul will like the movie
- Item similarity
  - Since 90% of those who liked Star Wars also liked Independence Day, and, you liked Star Wars
  - You may also like Independence Day

The content of items “doesn’t matter”!

A Formal Framework for Rating

Objects: O

Users: U

	o_1	o_2	…	o_j	…	o_n
u_1	3	1.5	…. …		2	
u_2	2					
…						
u_i	1			?		
…						
u_m	3					

$X_{ij}=f(u_i,o_j)=?$

Unknown function
$f: U \times O \rightarrow R$

The task

- **Assume known f values for some (u,o)'s**
- **Predict f values for other (u,o)'s**
- **Essentially function approximation, like other learning problems**

Where are the intuitions?

- Similar users have similar preferences
  - If $u \approx u'$, then for all o's, $f(u,o) \approx f(u',o)$
- Similar objects have similar user preferences
  - If $o \approx o'$, then for all u's, $f(u,o) \approx f(u,o')$
- In general, f is "locally constant"
  - If $u \approx u'$ and $o \approx o'$, then $f(u,o) \approx f(u',o')$
  - "Local smoothness" makes it possible to predict unknown values by interpolation or extrapolation
- What does "local" mean?

Two Groups of Approaches

- Memory-based approaches
  - Also known as “Neighbor-based” approaches.
  - Find “neighbors” of u and combine g(u’)(o)’s
- Model-based approaches
  - Assume structures/model
  - $f(u, o) = f'(c_u, c_o)$, where c_u, c_o are model representations for u and o.
  - Estimation & Probabilistic inference

Predicting Score

1. Calculate correlation

$$r_{KL} = \frac{Cov(K,L)}{\sigma_K \sigma_L}$$

$$= \frac{\sum_i (K_i - \bar{K})(L_i - \bar{L})}{\sqrt{\sum_i (K_i - \bar{K})^2}\sqrt{\sum_i (L_i - \bar{L})^2}}$$

$$= \frac{-2 - 2 - 2 - 2}{\sqrt{10}\sqrt{10}} = -0.8$$

2. Aggregate predictions

$$K6_{Pred} = \bar{K} + \frac{\sum_{J \in raters} (J_6 - \bar{J}) r_{KJ}}{\sum_J |r_{KJ}|} =$$

$$3 + \frac{2r_{KM} - r_{KL}}{|r_{KM}| + |r_{KL}|} = 3 + \frac{2 - (-.8)}{|1| + |-.8|} = 4.56$$

message #	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	?	2	5	?

Figure 5: a sample matrix of ratings.

IES Institute of Education Sciences | W AmplifyLearn.AI | UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS | O UNIVERSITY OF OREGON

Memory-based Approaches
(Resnick et al. 94)

- General ideas:
  - x_{ij}: rating of object j by user i
  - n_i: average rating of all objects by user i
  - Normalized ratings: $v_{ij} = x_{ij} - n_i$
  - Memory-based prediction

$$v_{dj} = k \sum_{i=1}^{m} w(a,i) v_{ij} \quad k = 1 / \sum_{i=1}^{m} w(a,i) \quad \Rightarrow \quad x_{aj} = v_{aj} + n_a$$

- Specific approaches differ in *w(a, i)* -- the distance/similarity between user *a* and *i*

IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

User Similarity Measures

- Pearson correlation coefficient (sum over commonly rated items)

$$w_p(a,i) = \frac{\sum_j (x_{aj} - n_a)(x_{ij} - n_i)}{\sqrt{\sum_j (x_{aj} - n_a)^2 \sum_j (x_{ij} - n_i)^2}}$$

- Cosine measure

$$w_c(a,i) = \frac{\sum_{j=1}^{n} x_{aj} x_{ij}}{\sqrt{\sum_{j=1}^{n} x_{aj}^{2} \sum_{j=1}^{n} x_{ij}^{2}}}$$

- Many other possibilities!

IES Institute of Education Sciences AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

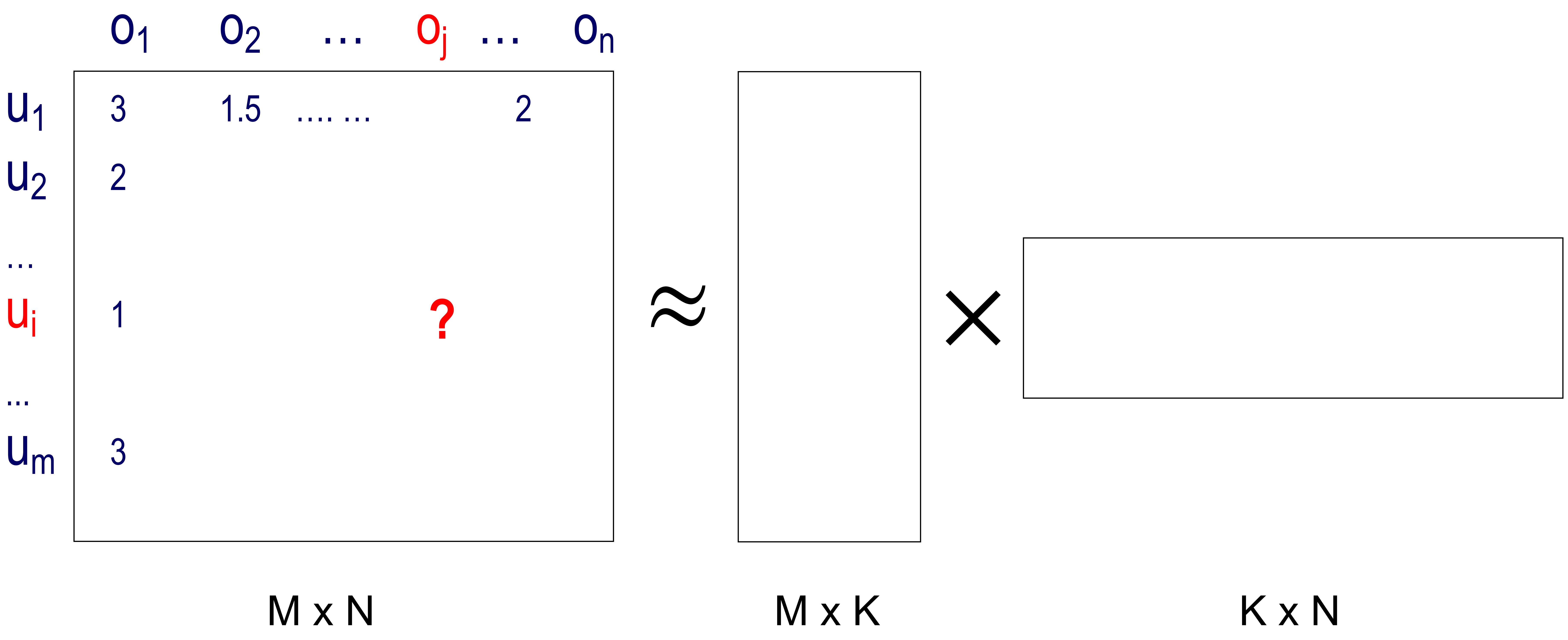
Problem with Memory-based Approaches

- Too many dimensions!
  - The rating matrix is *M*(user) x *N* (item)
- Too sparse observations!
  - Users rate very few items.
  - Recommendation not reliable

- Could we *model* users with fewer dimensions?
  - Latent factor models
  - Model users and items as *k*-dimension vectors

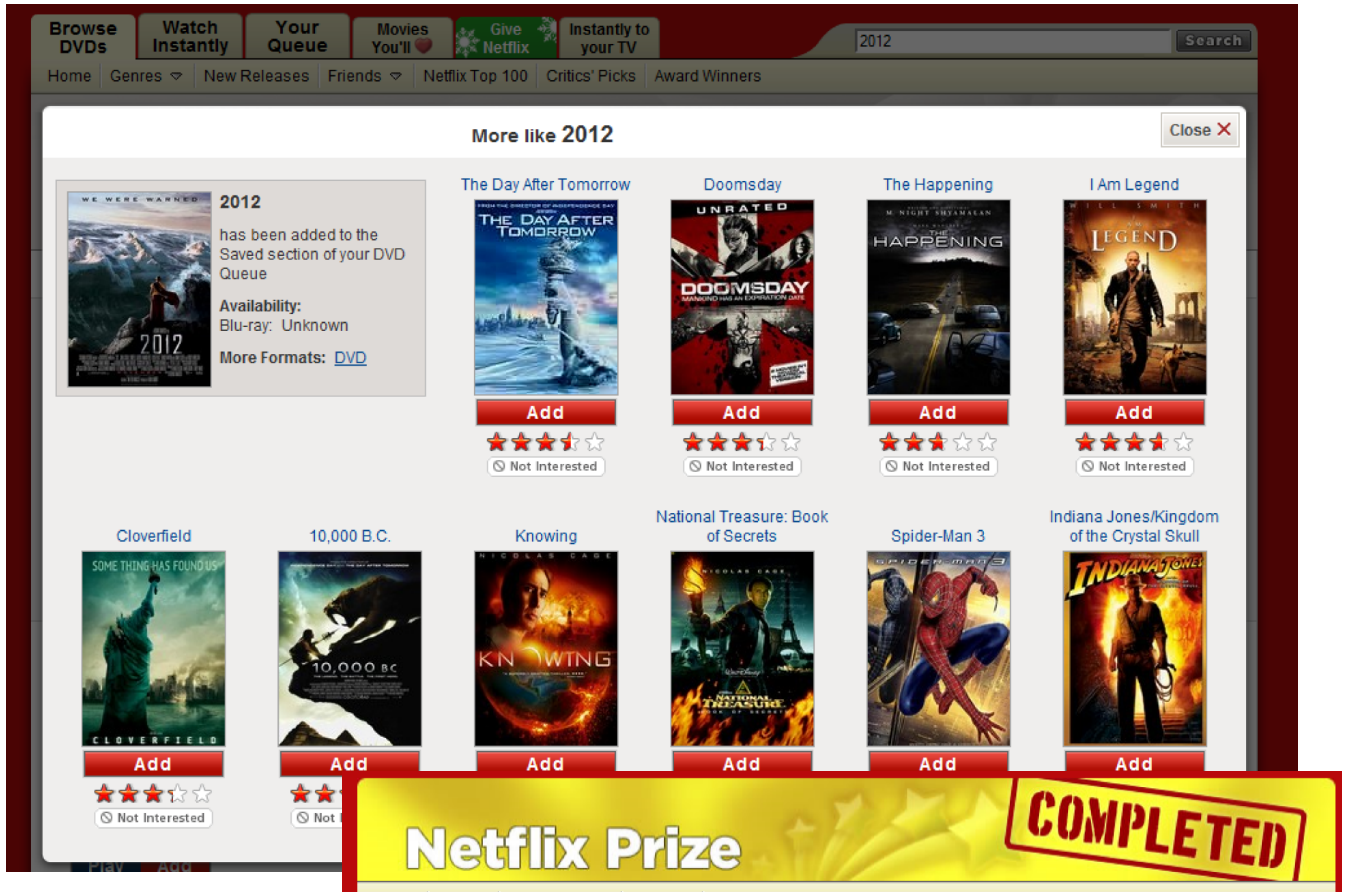
IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

Model-Based Approach — Matrix Factorization



* K dimensions can be interpreted as K interests, K concepts, K themes, etc.

Example: NetFlix



$$\hat{r}_{ui} = b_u + b_i + U_u \cdot I_i$$

IES Institute of Education Sciences
W AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Summary: RecSys as Information Filtering

- Information retrieval and information filtering are two ways to meet information need.
- Content based filtering and collaborative filtering are two major ways for information filtering.
- Recommender system most frequently refers to collaborative filtering.
- Two types of collaborative filtering:
  - Memory-based <——> Similarity of vectors
  - Model-based. <——> Patterns in Matrix (SVD)

RecSys as Applied Machine Learning

- Model-based Collaborative Filtering can be seen as a Machine Learning Prediction (regression) task.

$$RMSE = \sqrt{\sum (r_{ui} - \hat{r}_{ui})^2}$$

- But other problem formulation is also possible:
  - Classification: predict if a user like/dislike, purchase/not-purchase an item.
  - Sequential prediction: predict the next item in a sequence.
  - Ranking: provide a ranked list of items.

More ML algorithms can be adopted to RecSys

- Whenever a new ML algorithm trends, you will see them being used in $f(u, i)$
  - Neural network models (matrix factorization beyond dot product)
  - Graph network models (user-user graph, user-item graph, etc.)
  - Sequential neural network models (RNN/LSTM, attention, transformer, …)
- Open questions:
  - How will LLM be incorporated?
  - Will there be pre-trained foundation models for RecSys?

$$\hat{r}_{ui} = f(u, i)$$

IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

Evaluation is a first-class citizen in RecSys

- Different problem formulation means different evaluation metrics:
  - Classification: accuracy,
  - Sequential prediction: Precision/recall, …
  - Ranking: NDCG, MAP, …
- Evaluation beyond model performance:
  - Diversity and serendipity
  - Click-through rate vs. conversion rate
  - Short term revenue vs. long-term retention
  - …

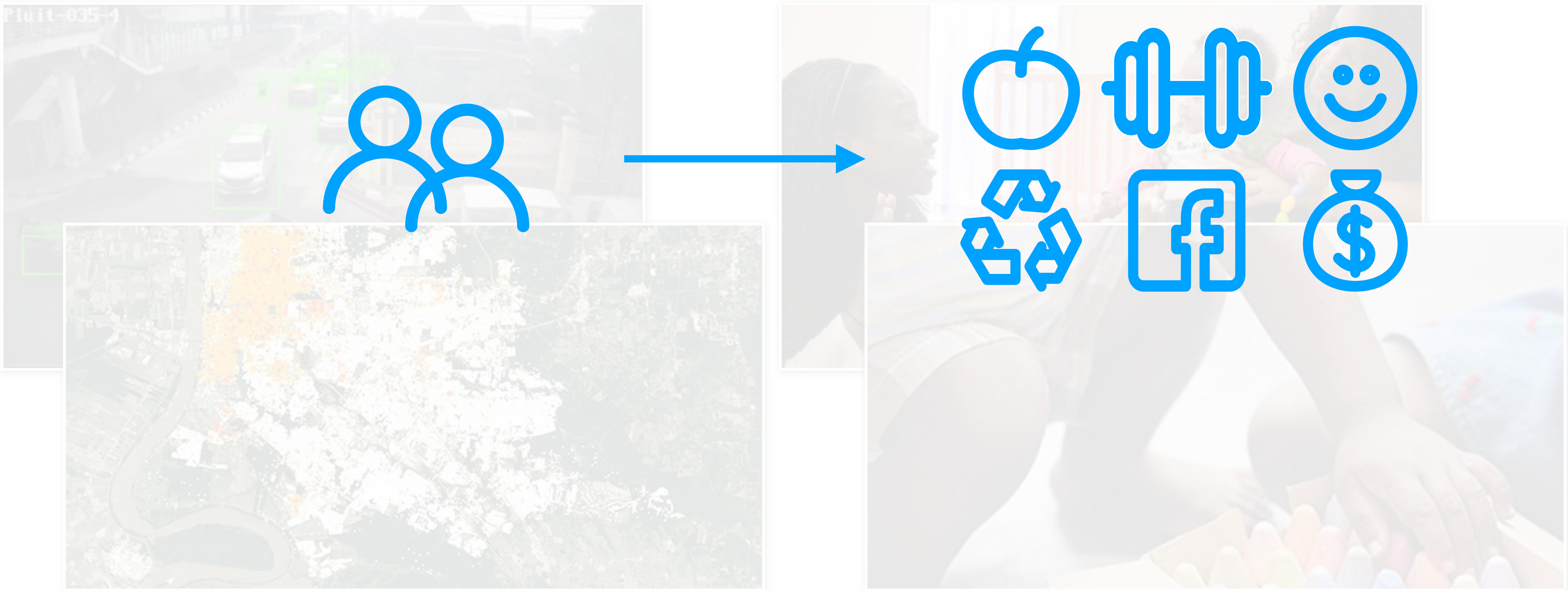
Recommender System in Social Experiment

IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

Data Science for Social Good

Traffic Safety

At-risk Moms



House Fires

Lead Poisoning

https://medium.com/pulse-lab-jakarta/using-deep-learning-to-tackle-traffic-safety-in-jakarta-a-collaboration-with-university-of-b5b4a4817e32

https://chihacknight.org/events/2018/06/05/lead-safe.html

https://datasmart.ash.harvard.edu/news/article/can-algorithms-predict-house-fires-990

https://gigaom.com/2013/10/08/how-data-science-is-helping-charities-save-lives-and-their-budgets/

Challenges in Promoting Behavioral Changes

- People may not make the most rational choice.
- People may not comply with unenforced policy.
- Change people's behavior through ***nudge***.

Can we *nudge* better with *data science*?

Nudge

… is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives.

Can we *nudge* better with *data science*?

End-to-End: from Data to Action

Data | Causal Inference | Recommender System | Field Experiment | Behavior

- Identify effective nudges through causal inference.
- Optimize the nudge using recommender system
- Evaluate the nudge with field experiment.

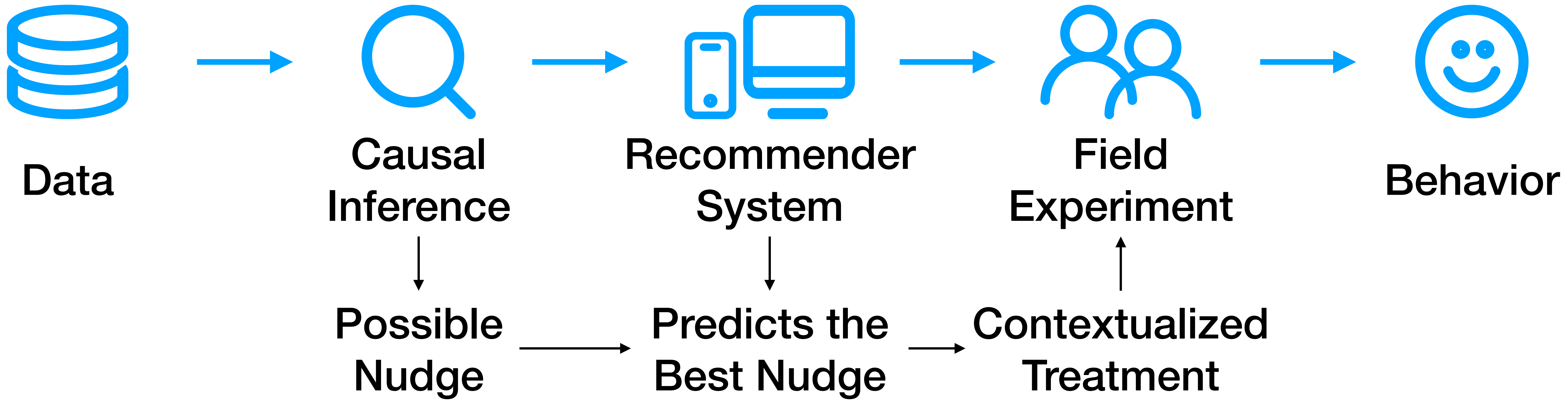
theory
→theory + data

rule based
→data/algorithm driven

algorithm assigned treatments
+ advanced result analysis

IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS O UNIVERSITY OF OREGON

An End-to-End Pipeline

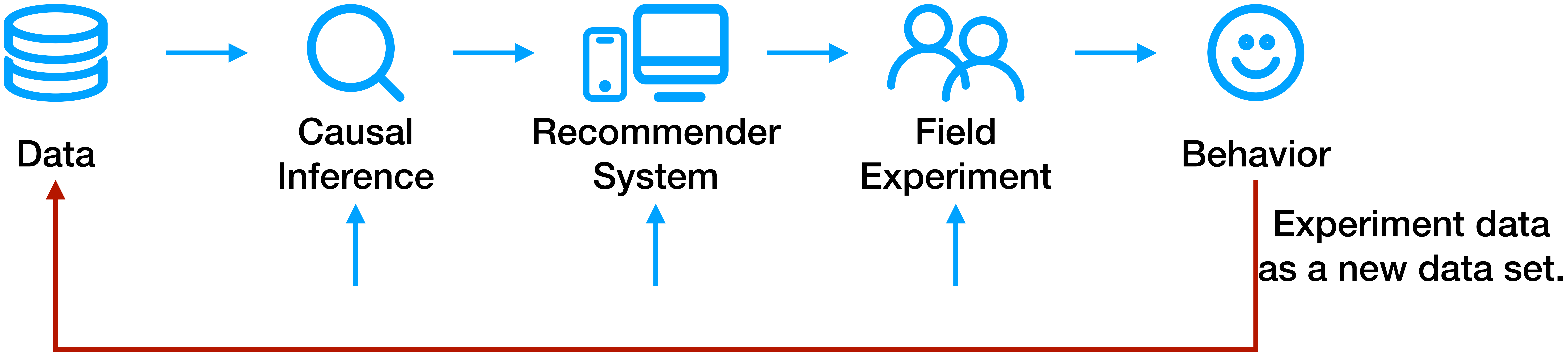


IES Institute of Education Sciences
W AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
O UNIVERSITY OF OREGON

Causal Inference + Machine Learning at Each Stage

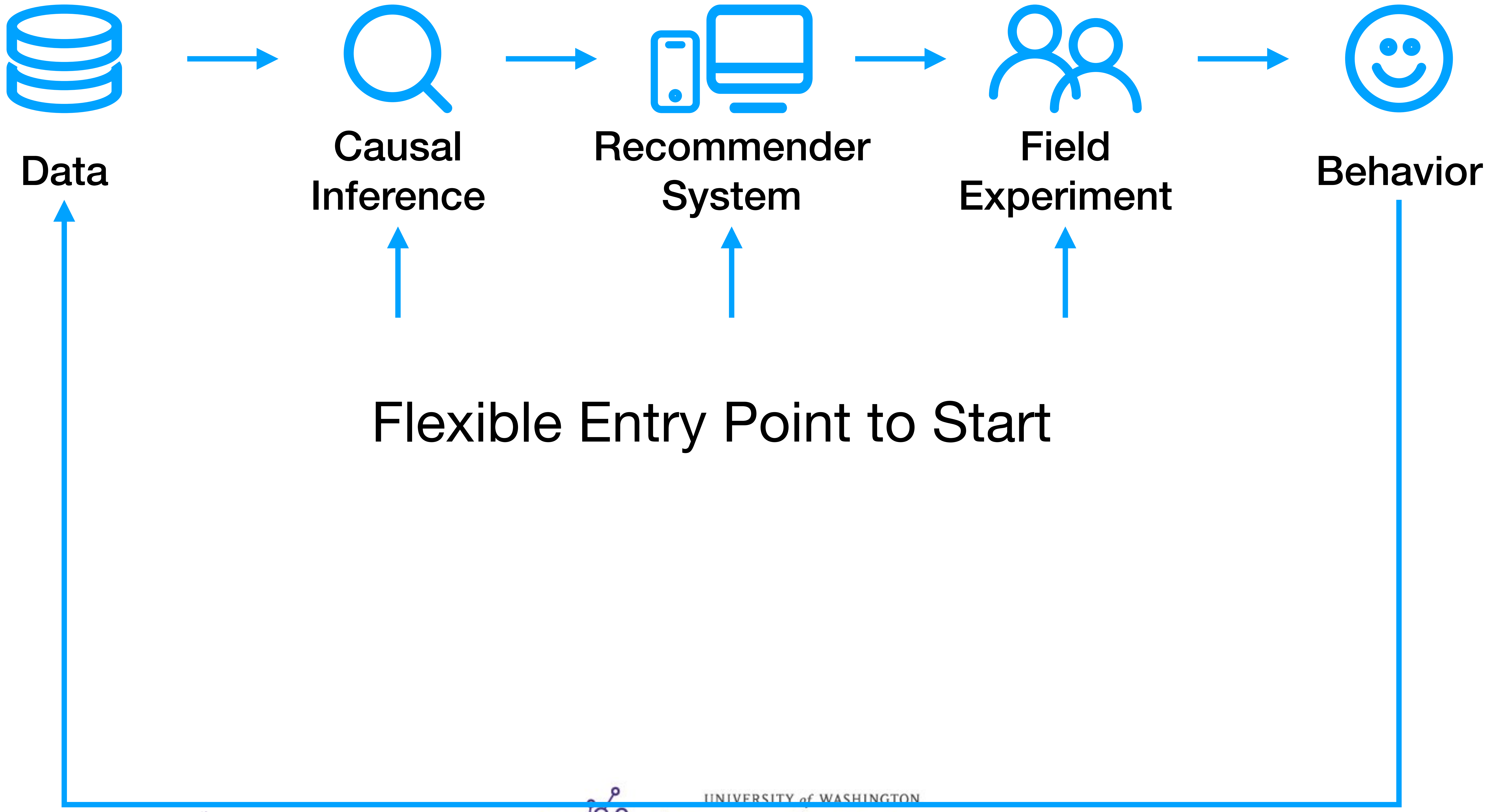
- Machine learning can help causal inference handle high-dimensional, complex, or unstructured data.
- Recommender System is a machine learning application, but with the goal of optimize the nudge.
- Randomized experiment is the gold standard for causal inference. Machine learning personalizes treatment and helps analyze heterogeneous treatment effect.

Connecting End to End as a Loop



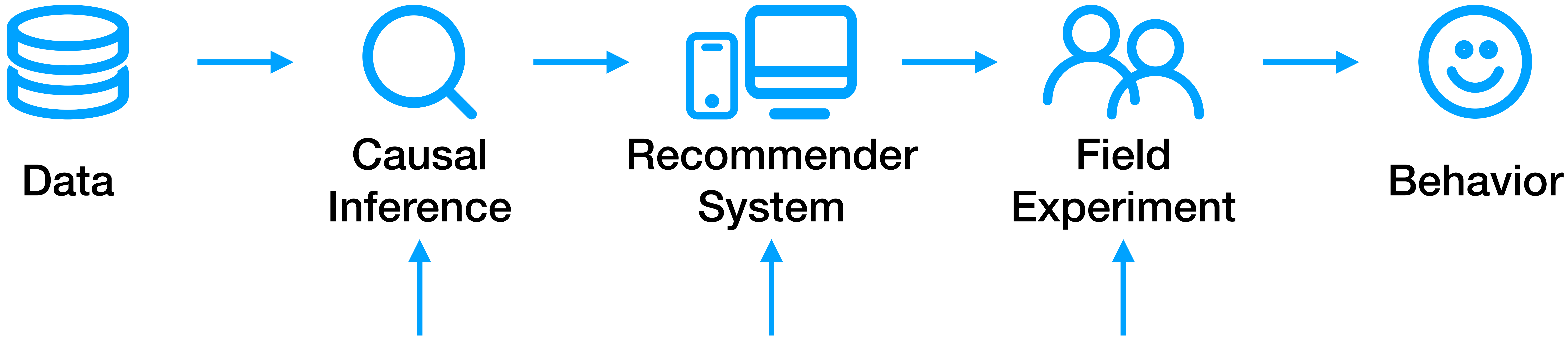
- Explore heterogeneous treatment effect to find insights for better nudge.
- Keep iterating the pipeline without an "end."
- Flexible Entry Point to Start

Connecting End to End as a Loop



Institute of Education Sciences
AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Promote Pro-social Behavioral Change through End-to-End Data Science



Emoji promotes developer participation on open source platform.

Recommending teams promotes pro-social lending in online microfinance.

Team competition increases driver productivity on ride-sharing platform.

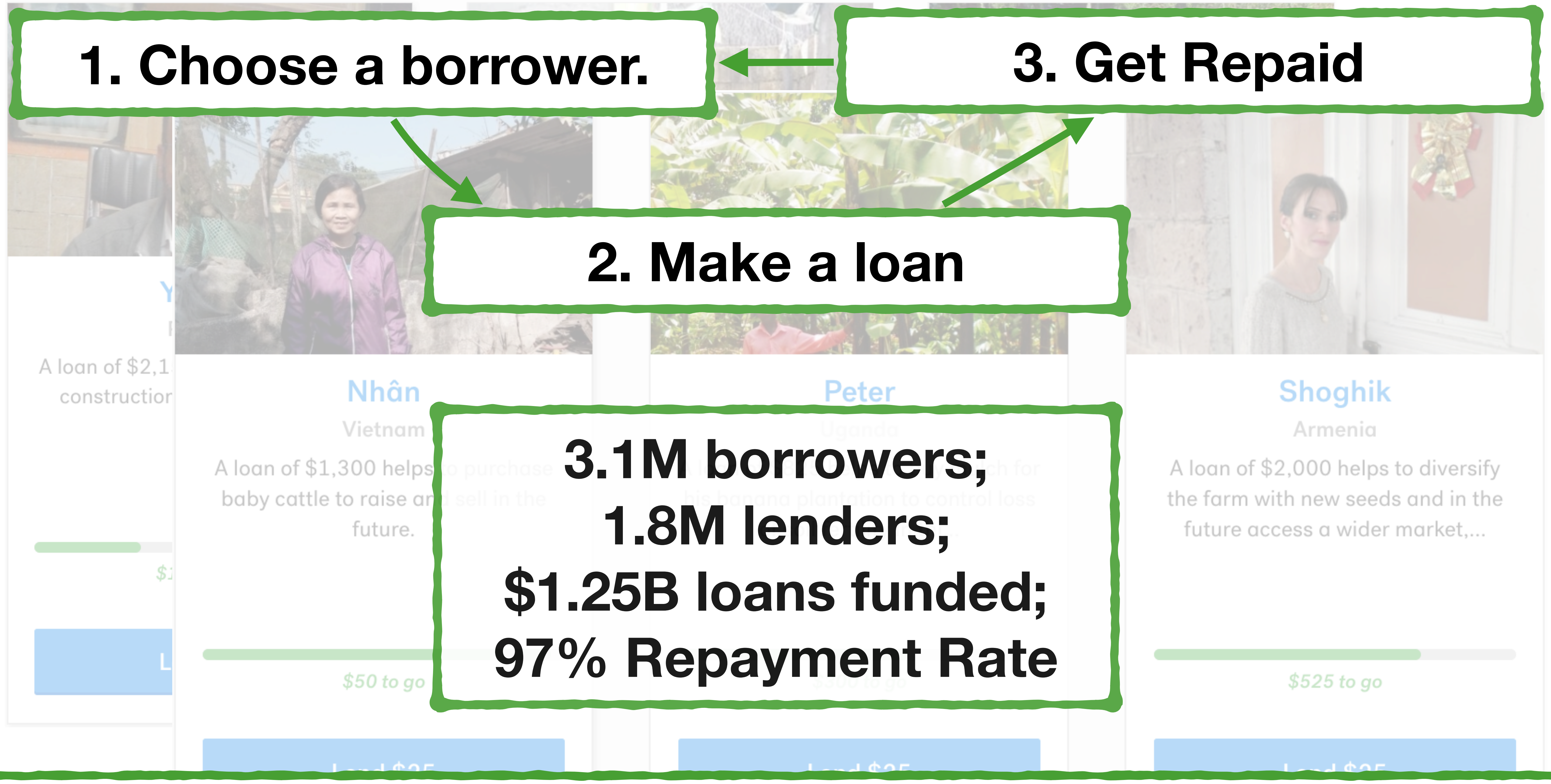
IES Institute of Education Sciences

AmplifyLearn.AI

UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

UNIVERSITY OF OREGON

kiva — Loans that changes lives



1. Choose a borrower.

2. Make a loan

3. Get Repaid

3.1M borrowers;
1.8M lenders;
$1.25B loans funded;
97% Repayment Rate

Challenge: How to increase lender participation?

IES Institute of Education Sciences | W AmplifyLearn.AI | UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS | O UNIVERSITY OF OREGON

Kiva Lending Team

Kiva Lending Team: University of Maryland

A Colleges/Universities team since Sep 6, 2008

ABOUT | LOANS 4.8K | MEMBERS 140 | GRAPHS | IMPACT

Kiva Lending Team: Team CANADA

A Local Area team since Aug 30, 2008

ABOUT | LOANS 250K | MEMBERS 10K | GRAPHS | IMPACT

We loan because...
So little means so much. And because we are so fortunate to be able to lend with the luxury of not worrying about whether we ever see that money again, while the clients borrow with the hope and determination that they will be able to repay, and improve their lives along the way.

Check out fundraising loans already being supported by Team Canada:
www.kiva.org/team/team_canada/loans?status=fundRaising

About us
We're Canadian, eh?

TEAM LEADERBOARDS

Amount funded

THIS MONTH | LAST MONTH | ALL TIME

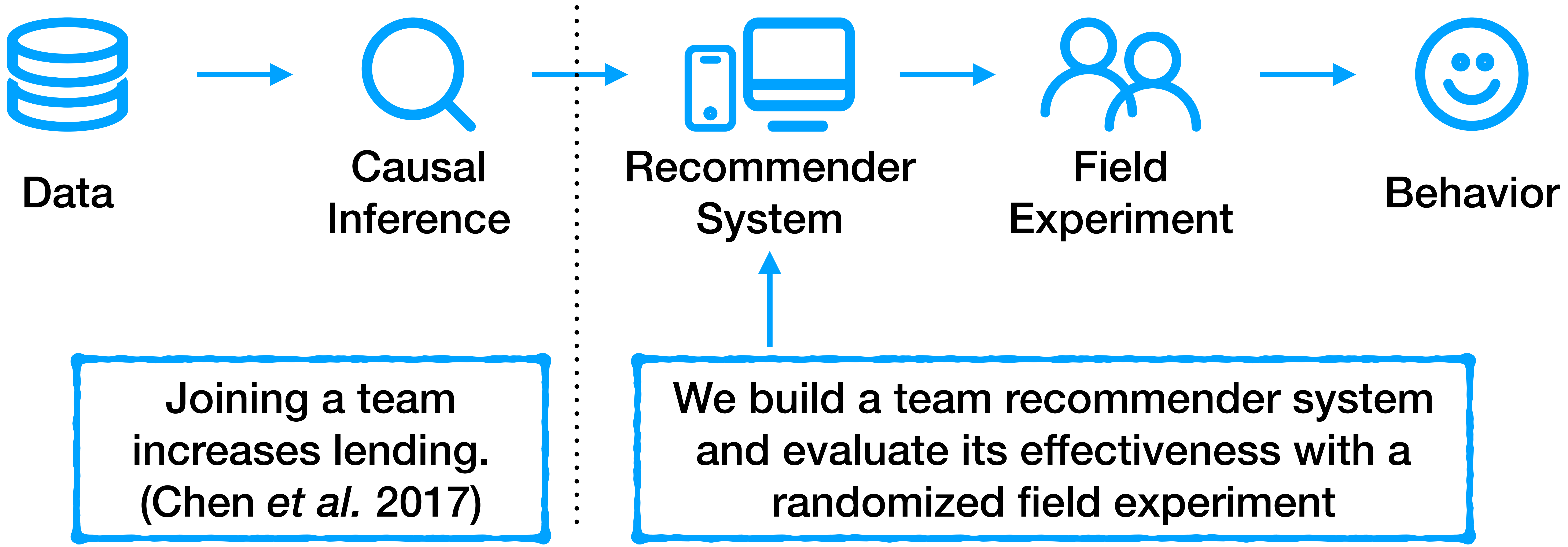
Team	Amount
Kiva Christians	$53,303,350
(A+) Atheists, Agnostics, Skeptics,…	$42,006,500
The Mindful Bunch	$18,838,950
InsideFlyer	$16,078,150
HP	$13,123,250
Nerdfighters	$11,555,900
Tieks by Gavrieli	$10,604,575
Team CANADA	$9,744,700
Friends of Bob Harris	$9,194,700

https://www.kiva.org/team/university_of_maryland

https://www.kiva.org/team/team_canada

https://www.kiva.org/teams

Team Recommender System



- H1: Lenders will be more likely to join teams if we make “good” recommendations.
- H2: Lenders will lend more after they join teams.

Experiment Design: 3 x 2 factorial

		Explanation	
		Explanation	No Explanation
Algorithm	Location	Location-Explanation	Location-NoExplanation
	Loan History	History-Explanation	History-NoExplanation
	Leaderboard	Leaderboard-Explanation	Leaderboard-NoExplanation
Control		No Contact	
		Teams Exist	

Institute of Education Sciences AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

“Teams Exist” Email

KIVA

Hi Wei,

Since you're such an awesome Kiva lender, we wanted to let you know about a fun feature of the Kiva experience: Kiva Lending Teams!

Lending Teams are self-organized groups around shared interests – location, alumni orgs, social causes, you name it. You can connect with other lenders, discover loans you might be interested in, and track your collective impact.

Check out some of the thousands of lending teams to find the right one for you.

Thanks for being a part of the Kiva community and making a difference around the world.

Best Wishes,
The Kiva Team

"Team Recommendation" Emails

KIVA

Hi Wei,

Since you're such an awesome Kiva lender, we wanted to let you know about a fun feature of the Kiva experience: Kiva Lending Teams!

Lending Teams are self-organized groups around shared interests – location, alumni orgs, social causes, you name it. You can connect with other lenders, discover loans you might be interested in, and track your collective impact.

Based on your past lending, people who have made similar loans enjoy being a part of these teams:

España - Spain

We loan because: Kiva ofrece un medio ideal para participar activamente en el apoyo a emprendedores sin recursos que no pueden acceder a los canales normales de financiación y que, gracias a los...

Join Team

Team Europe

We loan because: We think Kiva is a unique opportunity for people all over the world to assist entrepreneurs in improving their businesses and communities.

Join Team

Emprendedores Desencadenado.com

We loan because: We believe that entrepreneurship is the only way to fight poverty.

Join Team

Or check out the thousands of other lending teams to find the right one for you.

Thanks for being a part of the Kiva community and making a difference around the world.

Best Wishes,

"Other lenders who live near you enjoy being a part of these teams"

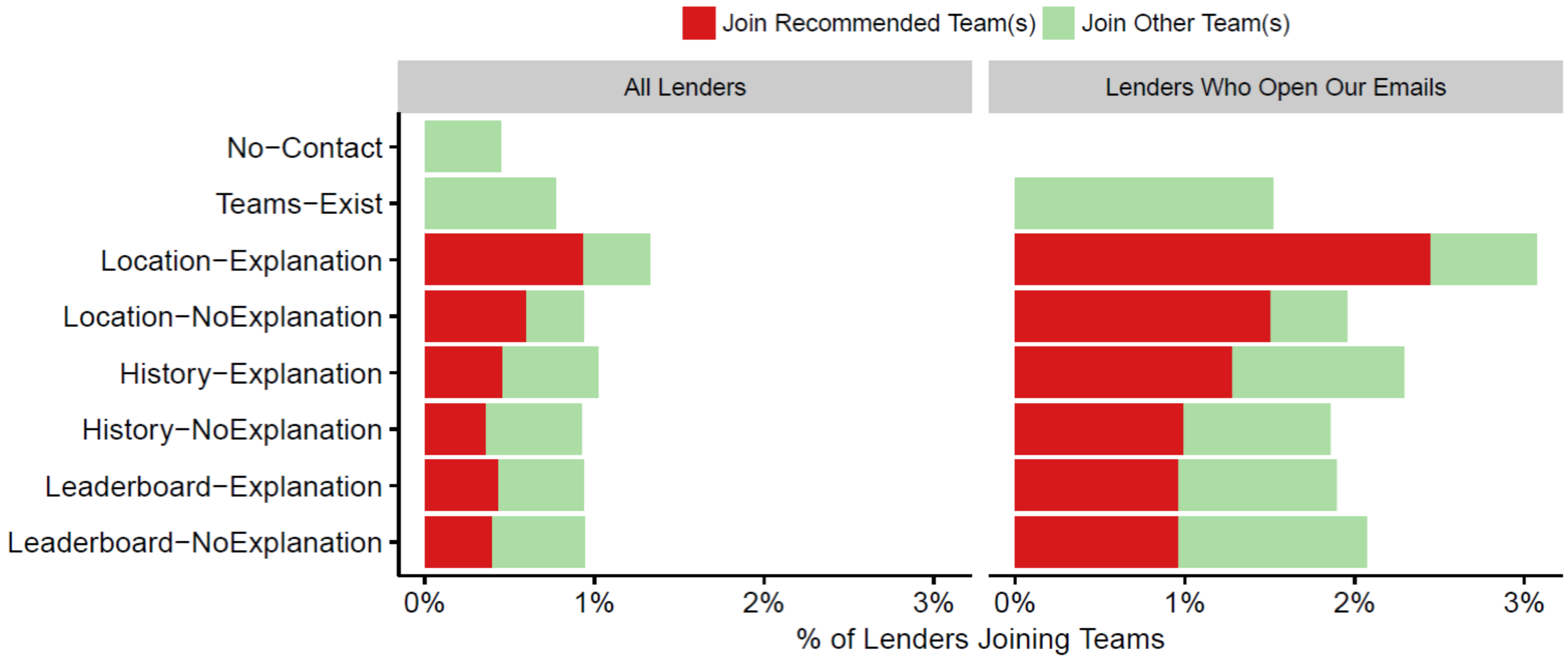
"Based on your past lending, people who have made similar loans enjoy being a part of these teams"

"Some of the most popular teams are"

"Here are a few teams you may want to check out"

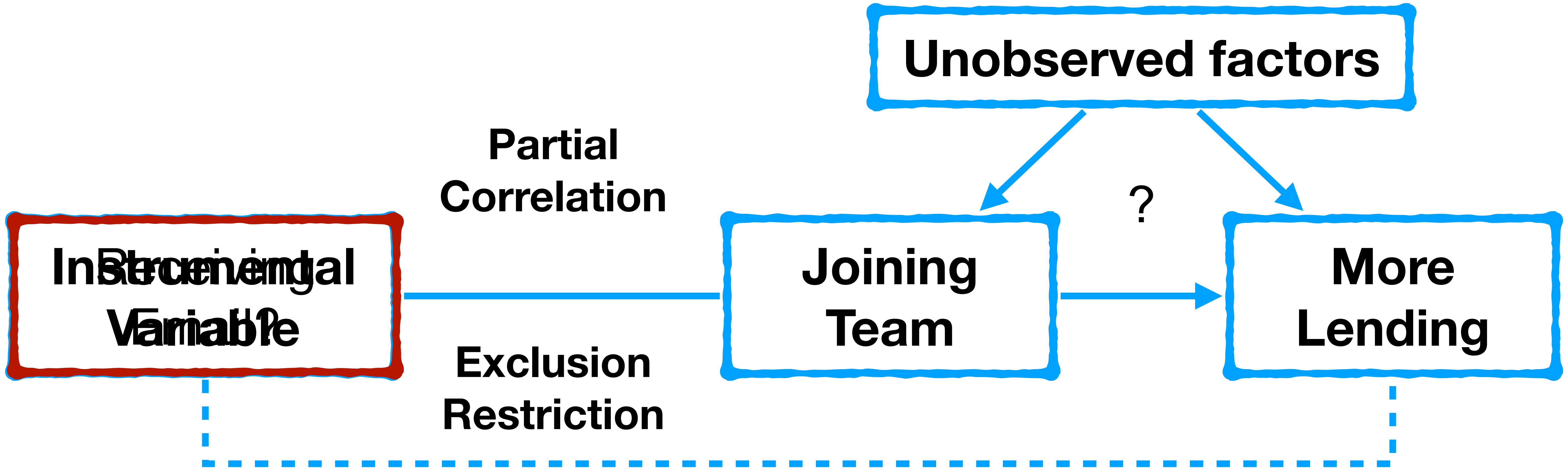
Effectiveness on Joining a Team

- H1: Lenders will be more likely to join teams if we make "good" recommendations



Does Joining Team Increase Lending?

- *Instrumental Variable* to tease out confounders.
- Treatment assignment (receiving email) as IV.
  - Partial correlation: receiving email ~ joining team. F = 23.55
  - Exclusion restriction: Email by itself doesn't increase lending
    - Chen et al. (2017)

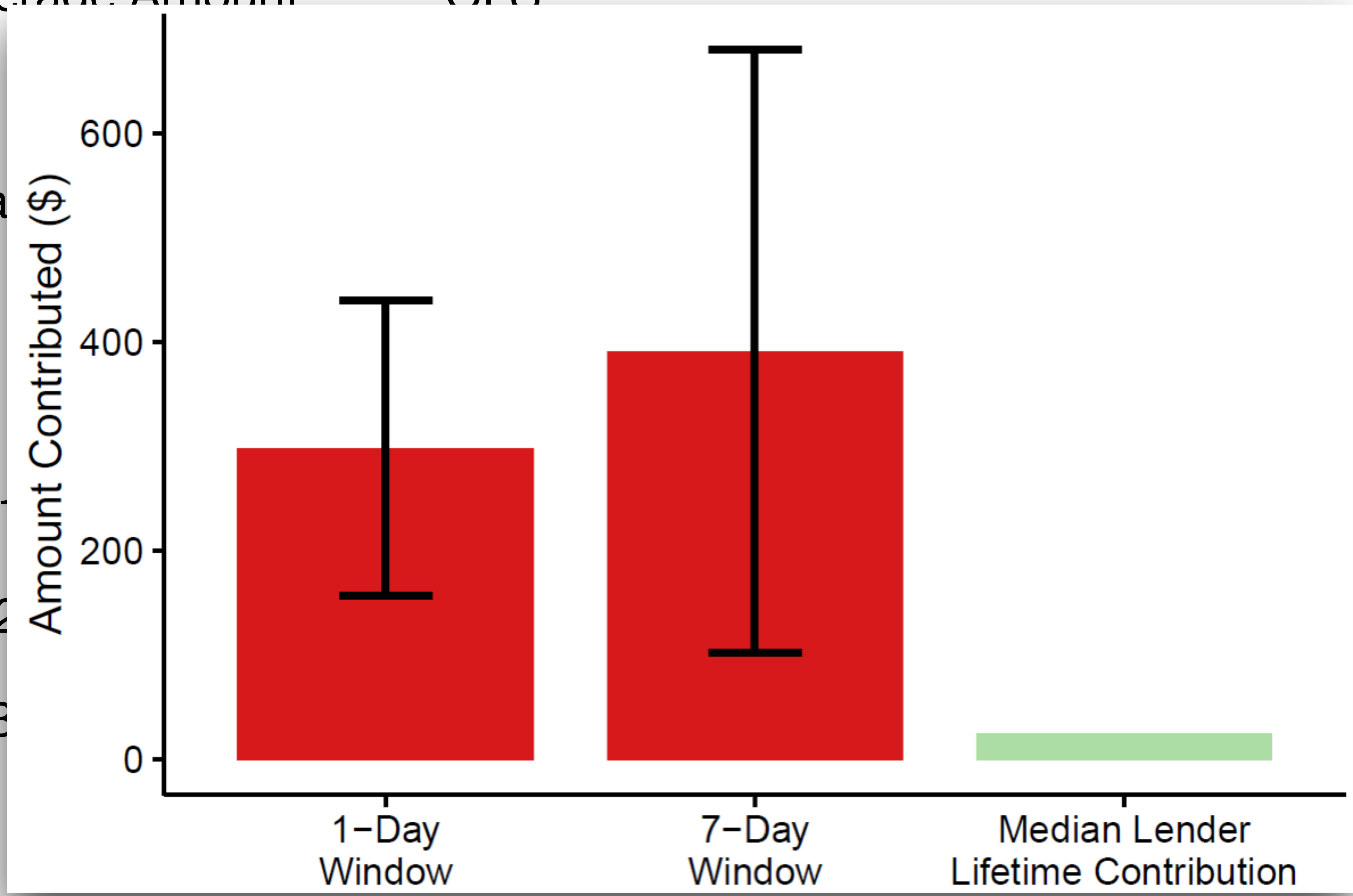


IES Institute of Education Sciences
AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Does Joining Team Increase Lending?

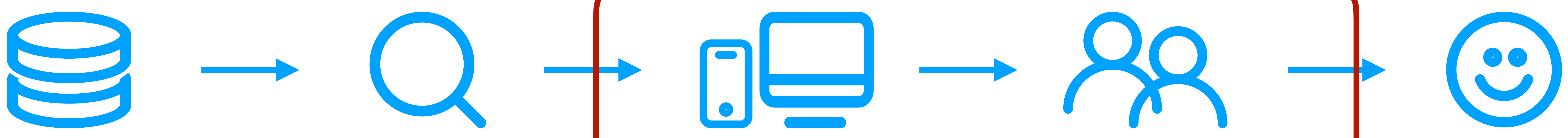
Difference-in-differences regressions of average daily lending amount (2SLS)

	IV 1st Stage	IV 2nd Stage: Average Amount		OLS
	(1)	(2)		
		1-Day	7-Da	
Email	0.0053***			
	(0.001)			
Join Team		298.5579***	55.9	
		(72.283)	(2	
Constant	0.0045***	-2.6593***	-0.93	
	(0.001)	(0.670)		



Note: Standard errors in parentheses, Significant at the: * 10%, ** 5%, and *** 1% levels.

Takeaway: Recommending Teams Promotes Pro-social Lending on Kiva.org



- **User Compliance increased**
- **Pro-social lending promoted**

- Can we apply a similar nudge (teams) to other online platforms?

Sharing Economy and its Problems

- Many view the sharing economy as the future of work.
- But for drivers:
  - Lack of Identity
  - Feeling Unaffiliated
  - Sense of achievement
- Can we help with driver teams?

- *"I have no interaction or relationship with other colleagues." (A driver, The Curiosity Daily, 2019)*
- *"These are jobs that don't lead to anything." (A TaskRabbit worker, The New Yorker, 2017)*

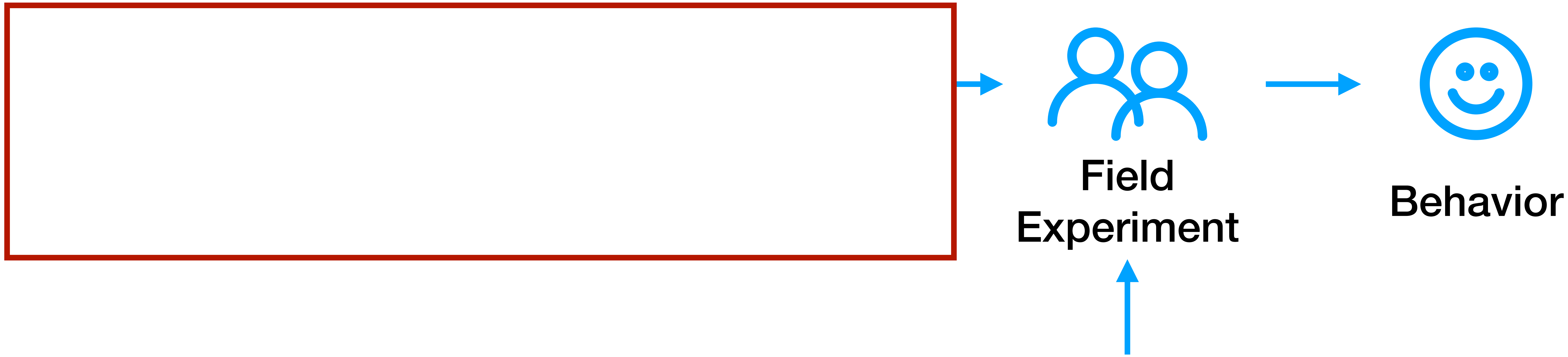
Team Formation → **Intra-team Coordination** → **Team Competition**

Improving Driver Performance through Team Competition

Team Formation → Intra-team Coordination → Team Competition

- Does team contest increase driver productivity?
  - Social identity theory. (Akerlof & Kranton 2000, 2010)
  - Contest theory. (Konrad 2010, Fu et al. 2015)
- Does team composition make a difference?
  - Homephily vs. Diversity?

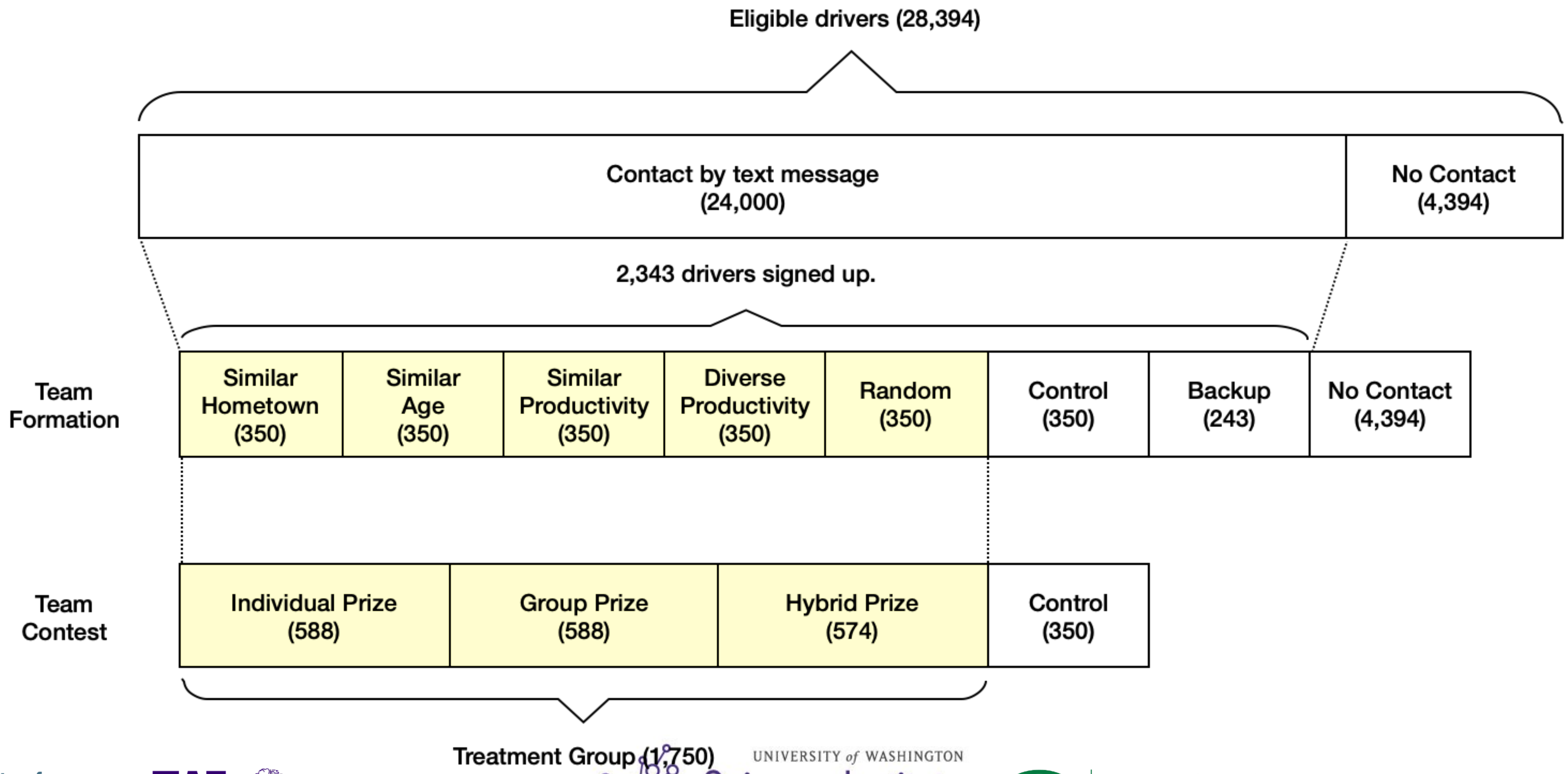
Team Contest on the Platform



Data-driven recommender system is not available

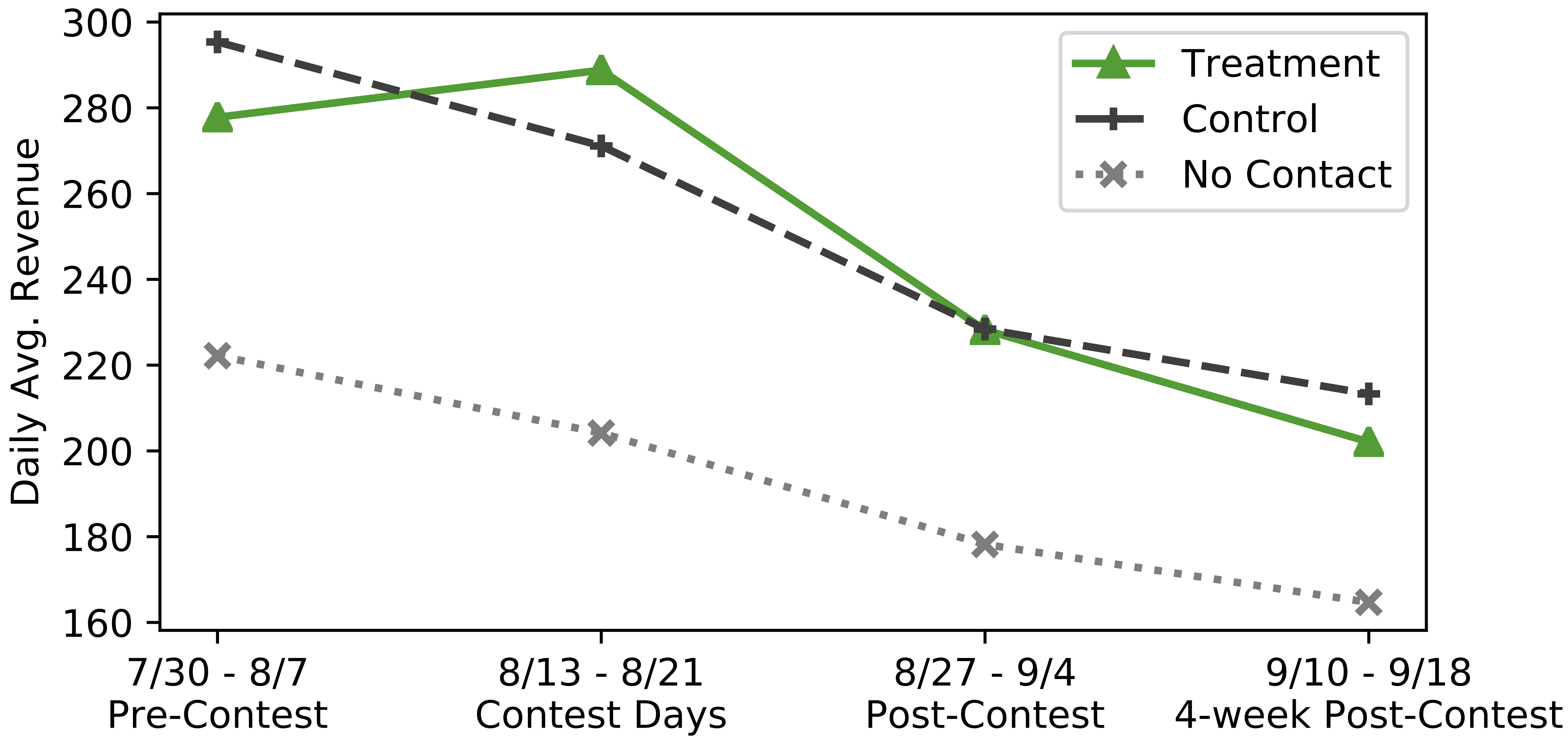
We can start from field experiments!

Experiment Design



Ai et al. <u>*Putting Teams into the Gig Economy: A Field Experiment at a Ride-sharing Platform*</u>, *Management Science*, forthcoming

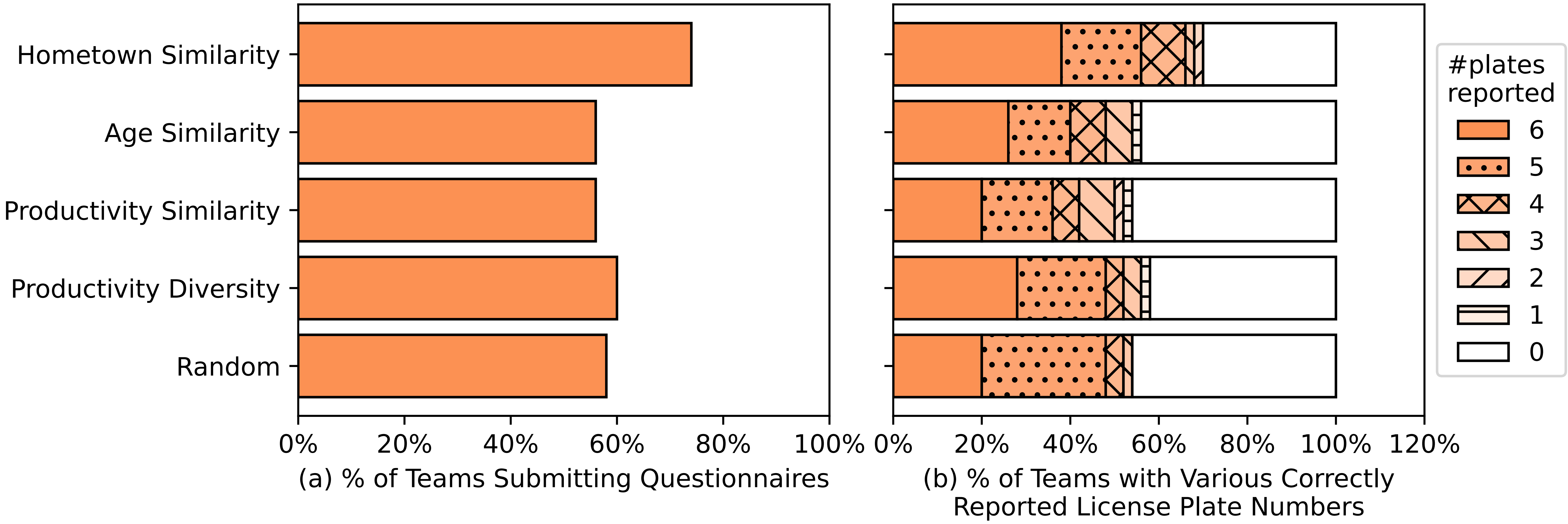
Average and Heterogeneous Treatment Effect



- Treated drivers earn 35 CNY (12%) more than those in the control group.
- Effects are stronger for those in responsive teams (56 CNY, 19%), and persist two weeks after contest.

IES Institute of Education Sciences
AmplifyLearn.AI
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Team Responsiveness



60.8% of team captains submitted the survey: responsive teams

Institute of Education Sciences

AmplifyLearn.AI

UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

UNIVERSITY OF OREGON

Similarity and Diversity on Driver Productivity

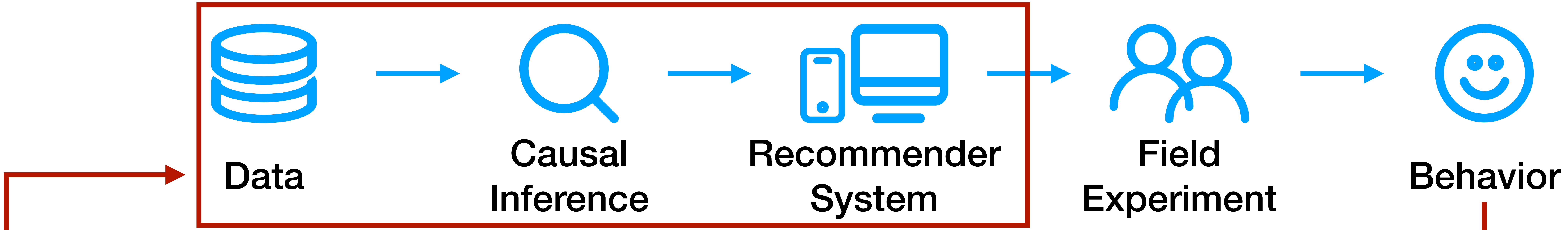
	Dependent variable: Δ Daily Revenue (CNY)					
	By Treatment Group			By Diversity Metrics		
	(1)	(2)	(3)	(4)	(5)	(6)
Time Period	Contest	2 weeks Post Contest	4 weeks Post Contest	Contest	2 weeks Post Contest	4 weeks Post Contest
Age Similarity	0.933	33.19**	9.806			
	(16.91)	(12.70)	(11.05)			
Hometown Similarity	5.838	20.70	17.12			
	(18.35)	(13.16)	(13.62)			
Productivity Similarity	-14.65	21.47*	13.85			
	(17.15)	(12.04)	(12.67)			
Productivity Diversity	-17.50	17.50	11.33			
	(15.62)	(12.25)	(13.09)			
Age Stdev				-0.417	-3.357**	-0.123
				(1.647)	(1.346)	(1.279)
Avg. Hometown Distance				0.0297	-0.00706	-0.0196
				(0.0242)	(0.0227)	(0.0203)
Productivity Std.				0.0953	-0.0347	-0.00401
				(0.122)	(0.0882)	(0.0961)
DiDi Age Std.				-0.0646	-0.0370	-0.0852
				(0.0914)	(0.0852)	(0.0799)
Constant	16.07	-68.17***	-86.12***	4.701	-15.89	-48.15**
	(13.69)	(9.377)	(8.566)	(29.68)	(21.04)	(22.52)
# Driver	1,750	1,750	1,750	1,750	1,750	1,750
Observations	8,750	8,750	8,750	8,750	8,750	8,750

Standard errors in parentheses are clustered at the contest (individual) level for treatment (control) conditions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Experiment Result Summary

- Team contest increases driver productivity
  - Driven by responsive teams.
- Team Composition makes a difference:
  - Hometown-similar teams are more likely to be responsive.
  - Age-similar teams are more active after the contest
- **Heterogeneity** in treatment Effects.

Platform-wide Implementation



- Recommender System is built upon experimental data
  - Zhang *et al.* 2019 (CIKM'19)
- Shipped into product with *HUGE* product impact
  - In 2018: **2.08 Million** drivers participated in **1,548** team contests across **52** cities
- Supported new field experiments
  - Ye *et al.* 2020 (working paper)

Zhang et al. 2019. Recommendation-based Team Formation for On-demand Taxi-calling Platforms, CIKM'19

Limitation - Teams Dismissed After Contest

- Wasted opportunities:
  - Team identify should have long-term effects.
- Short-term contests costly:
  - Status contest without monetary rewards.

- Will bonus-free longer-term team leaderboard improve worker revenue and retention?

IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

Problems and Opportunities

- Heterogeneity in treatment effects:
  - Why does a design work in one city but not in another?
  - What types of drivers and teams benefit more from team contest?
- Heterogeneity in contest design:
  - e.g. Prize structure, team size, and Design What contest designs better increase driver performance?
- In 2018: 2.08 Million drivers participated in 1,548 team contests across 52 cities
  - Each contest is a “mini” experiment.

IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

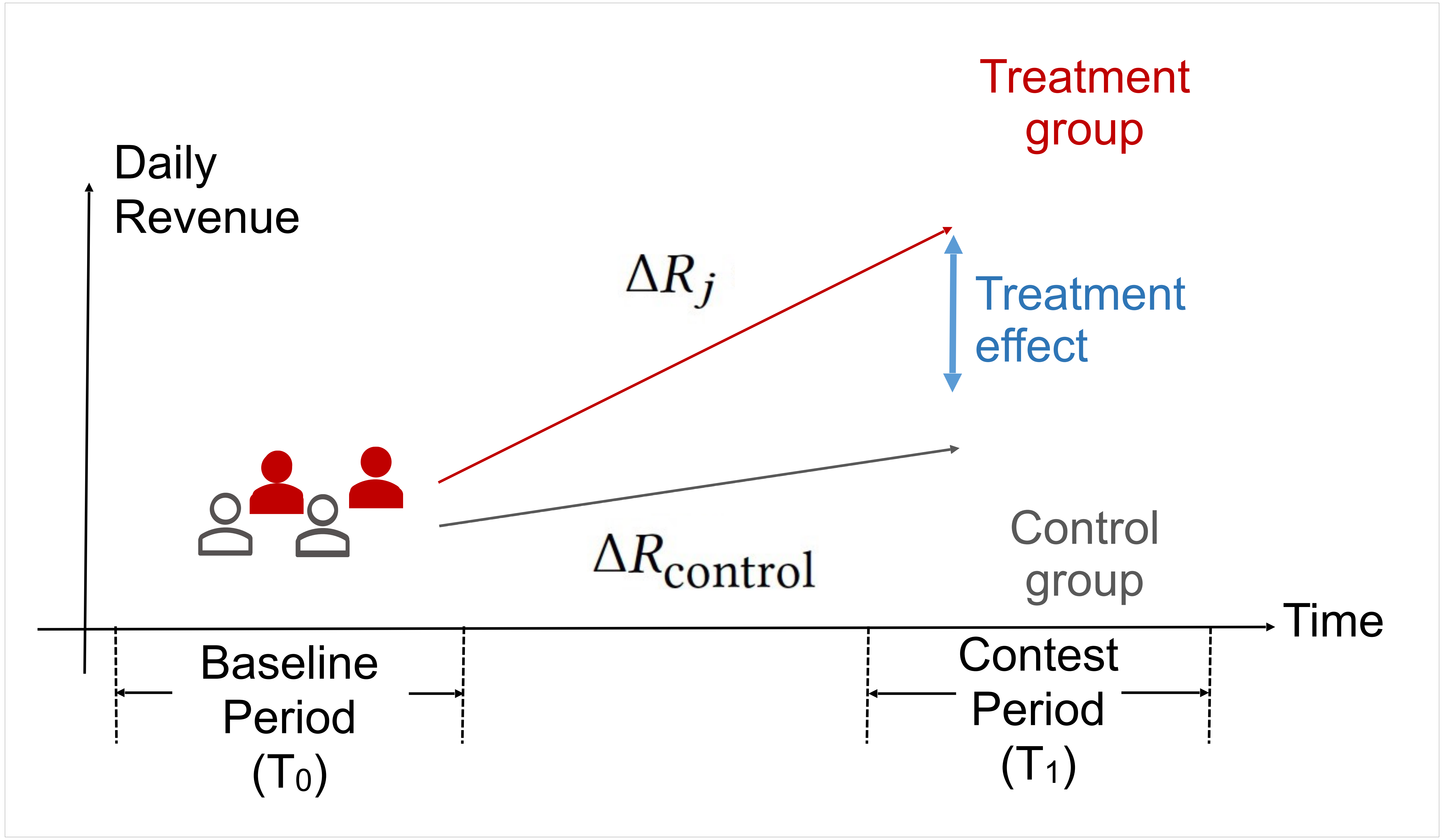
Ye et al. Predicting Individual Treatment Effects of Large-scale Team Competitions in a Ride-sharing Economy. KDD’2020

Machine Learning Analysis — A Prediction Task

- We want to understand how different factors **predict** the outcomes of individual drivers
  - The Individual Treatment Effect (ITE):
  - Revenue increase of a driver who team up and participate in the contests compared to the solo drivers in the control group

IES Institute of Education Sciences · W AmplifyLearn.AI · UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS · UNIVERSITY OF OREGON

Individual Treatment Effect Estimation



Within-driver revenue change:

$$\Delta R_j = R_{j,T_1} - R_{j,T_0}$$

Average revenue change of control group:

$$\Delta R_{\text{control}} = \frac{1}{|\text{control}|} \sum_{i \in \text{control}} \Delta R_i .$$

Individual treatment effect:

$$\Delta R_j^{\text{ITE}} = \Delta R_j - \Delta R_{\text{control}}$$

What predicts individual treatment effect?

- City properties
- Contest design
- Driver properties
- Team properties

- 555 features designed based on theories and domain knowledge:
  - e.g., virtual teams, social influence, social identity

Table S40. List of features

Level	Feature Selected
Team-level	Pre-contest team ranking in leaderboard
	Teammate average pairwise # of pre-contest team-ups
	Pre-contest average pairwise similarity in driving area
	Team average pairwise hometown distance
	Team age standard deviation
Driver-level	Driver age
	Driver platform age
	Driver pre-contest revenue difference with teammates' average
City-level	# of snow days during contest
	# of rainy days during contest
	Amount of bonus for the winning team
	Has individual threshold bonus during contest
	City-level pre-contest average daily order-fulfillment rate
	# of drivers on this platform in contest city

IES Institute of Education Sciences W AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

Method

- Model: Lasso, Gradient Boosting Regression Tree
  - Capture both linear and non-linear effects of features
  - Easy to interpret the features
- 520 contests, 143 cities, > 0.5 million drivers
- Data split
  - Train (70%), Validation (15%), Test (15%)
- Evaluation
  - RMSE $= \sqrt{\frac{1}{N}\sum(\hat{y}_t - y_t)^2}$

Institute of Education Sciences AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

Results - Model Performance

- Best-performing models reduce the prediction error (RMSE) by > 11.5%

	All Teams	
Model	Validation RMSE	Testing RMSE
Random Forest	172.64	172.68
Uniform baseline	-	195.10
Random baseline	-	277.42

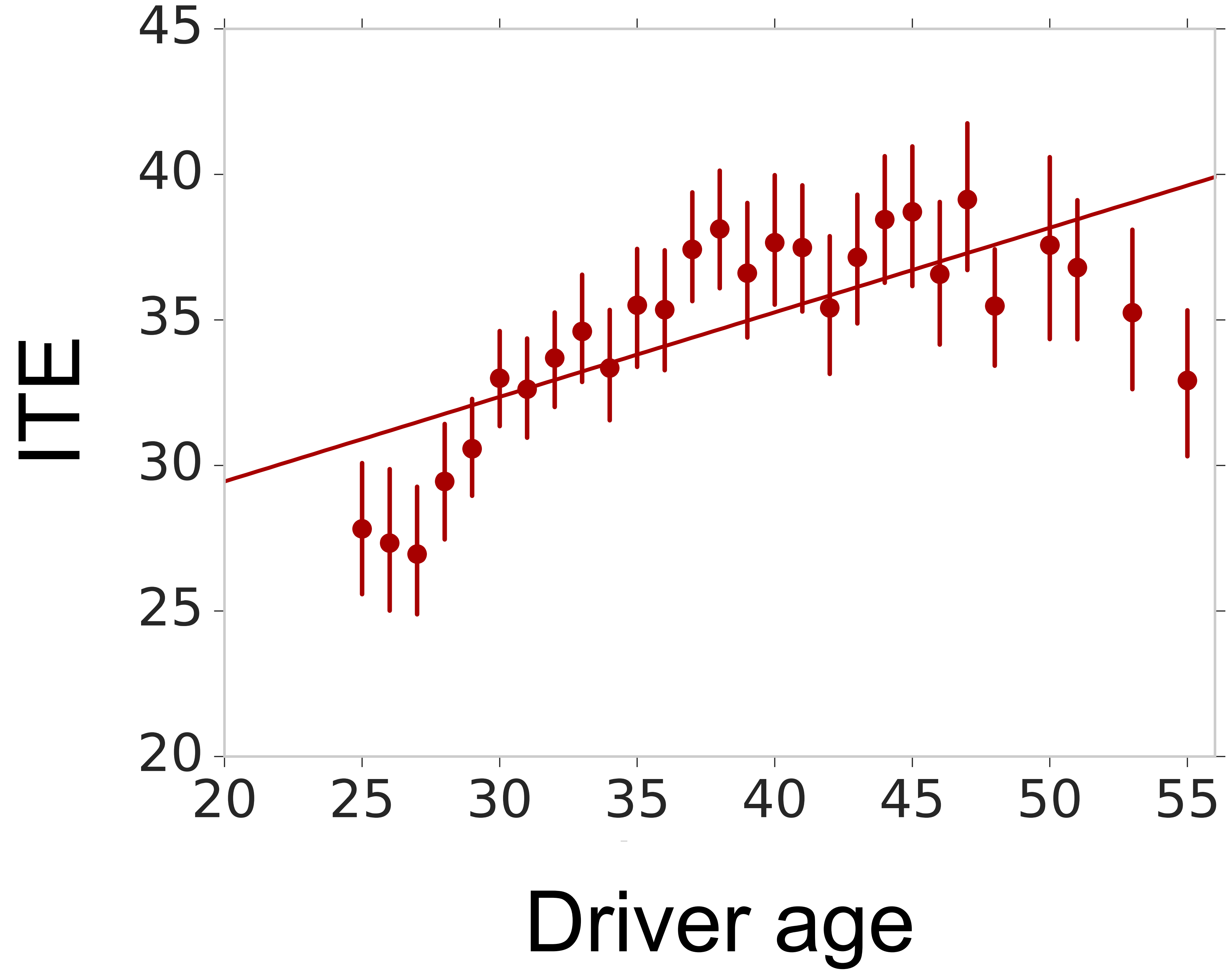
Feature Importance
— robustness check and explaining heterogeneity

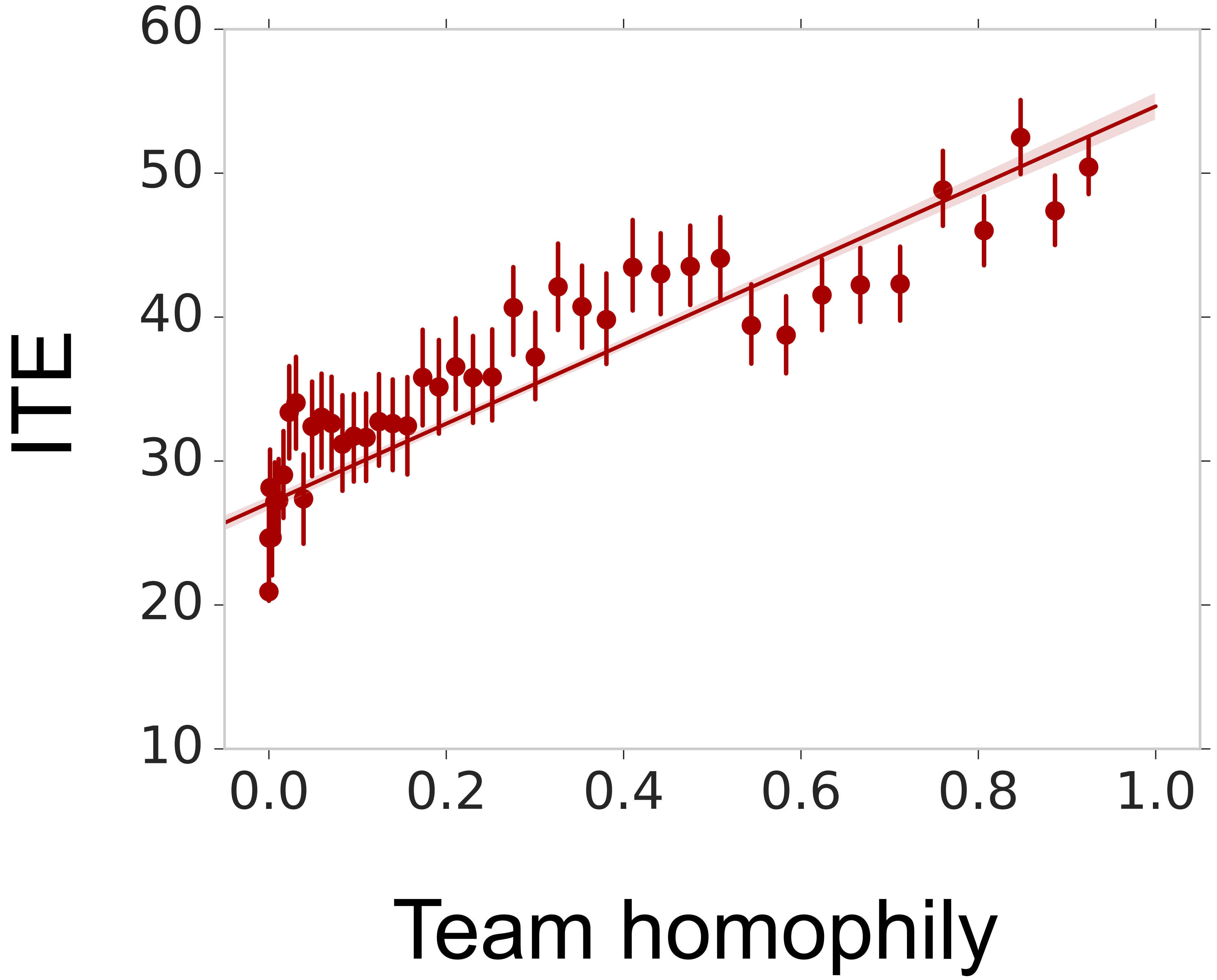
Table S43. Feature importance of the best-performing Random Forest Model.

Ranking	Feature name	Feature Importance: System-formed	Feature Importance: All teams
1	Driver pre-contest revenue difference with teammates' average	0.32	0.34
2	Driver Platform age	0.13	0.12
3	Pre-contest average pairwise similarity in driving area	0.10	0.10
4	Team age standard deviation	0.10	0.09
5	Team average hometown distance	0.09	0.09
6	Driver age	0.08	0.07
7	Teammate average pairwise # of pre-contest team-ups	0.04	0.05
8	# of drivers on this platform in the contest city	0.04	0.04
9	City-level pre-contest average daily order-fulfillment rate	0.03	0.03
10	Pre-contest team ranking in leaderboard	0.03	0.03
11	Prize amount for the winning team	0.02	0.02
12	# of rainy days during contest	0.02	0.02
13	Has individual threshold bonus during contest	0.001	0.001
14	# of snow days during contest	0.0004	0.001

Note: Feature importance scores reflect the proportion of node impurity reduction explained by a given feature. Features are ranked by importance scores using the dataset of the system-formed teams and the same ranking is obtained using the all-team dataset.

Examples of Intriguing Findings: Age and Team Homophily





Ye et al. *Predicting Individual Treatment Effects of Large-scale Team Competitions in a Ride-sharing Economy. KDD'2020*

Takeaway (RecSys in Social Experiment)

- Driver team contest increases drivers' productivity and improves their emotional well-being.
- End-to-end data science allows iteratively building, experimenting, and analyzing.
- Machine learning joins forces with causal inference.
- The end-to-end pipeline allows integration of domain expertise and it is necessary to do so.

Carlson School of Management
University of Minnesota
College of Information Studies
School of Information
University of Michigan
IES Institute of Education Sciences
AmplifyLearn.AI
University of Washington eScience Institute
Advancing Data-Intensive Discovery in All Fields
University of Oregon

References

- Ai et al. Recommending teams promotes prosocial lending in online microfinance. PNAS 2016.
- Ai et al. Putting Teams into the Gig Economy: A Field Experiment at a Ride-sharing Platform. Management Science, 2023.
- Ye et al. Virtual Teams for the Modern Workforce. PNAS. 2022.
- Ye et al. Predicting Individual Treatment Effects of Large-scale Team Competitions in a Ride-sharing Economy. In SIGKDD 2020.

IES Institute of Education Sciences AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

Takeaway

- Recommender system as Information Filtering
  - Content based vs. Collaborative Filtering
  - Memory vs. Model based Filtering
- Recommender system can be considered as applications of machine learning algorithms (e.g. classification, regression, ranking.)
- Recommender system can be used to provide contextualized nudges to promote prosocial behaviors
- Evaluation should be the first-class citizen in designing RecSys.

IES Institute of Education Sciences AmplifyLearn.AI UNIVERSITY of WASHINGTON eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS UNIVERSITY OF OREGON

Discussion

- How will LLM change the landscape of recommender system?

- Can you think about an application of recommender system in the education setting?
  - Will it be content-based filtering or collaborative filtering?
  - Can it be formulated as a machine learning problem?
  - How will you evaluate the recommender system?
  - What do you see as the biggest challenge in implementing the RecSys?