Fairness in Two-Sided Platforms

Abhijnan Chakraborty

Max Planck Institute for Software Systems
Saarbrücken, Germany

https://www.mpi-sws.org/~achakrab

MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

Algorithmic Decision Making in Practice

Algorithms being used to assist or replace human-decision making in several domains

Banking: Loan approval
Employment: Filtering and ranking applicants
Judiciary: Bail decisions
Healthcare: Determining high-risk patients

Benefits

- Higher accuracy, effectiveness
- Lower cost, higher efficiency
- Consistency
- Preventing certain human biases and prejudices
- Better access to opportunities and resources
- ...

Challenges and Risks

- Higher unfairness: unequal allocation of benefit or harm
- Stereotyping: denigration, unequal representation
- Opaqueness: inexplicability
- Accountability: due process
- Recourse: right to dispute/appeal
- Invasion of privacy: surveillance
- ...

Public Safety

Predictive Policing:

Predicting patterns in criminal activity for police placement

Predictive policing is a scam that perpetuates systemic bias

TNW

- Allocation of scarce resources with higher precision
- Reduce the role of human instincts and prejudices
- Perpetuate biases against racial groups

Criminal Justice

Recidivism Risk Assessment:
Predicting risk of future crime for bail or sentencing decisions

- Reducing crimes committed by released defendants
- Making consistent decisions across different judges
- Discrimination towards racial groups

Employment

Job Candidate Screening:

Predicting who will be a successful hire

Amazon scraps secret AI recruiting tool that showed bias against women

REUTERS

- Better matching at a lower cost
- Reduce impact of human affinity biases
- Replicate gender bias in past decisions

Healthcare

Identifying High Risk Patients:
Predicting who will need additional care

Healthcare algorithm used across America has dramatic racial biases

The Guardian

- Allocation of manpower to who need it most
- Reduce arbitrariness of human scheduling
- Replicate historical neglect towards poorer groups

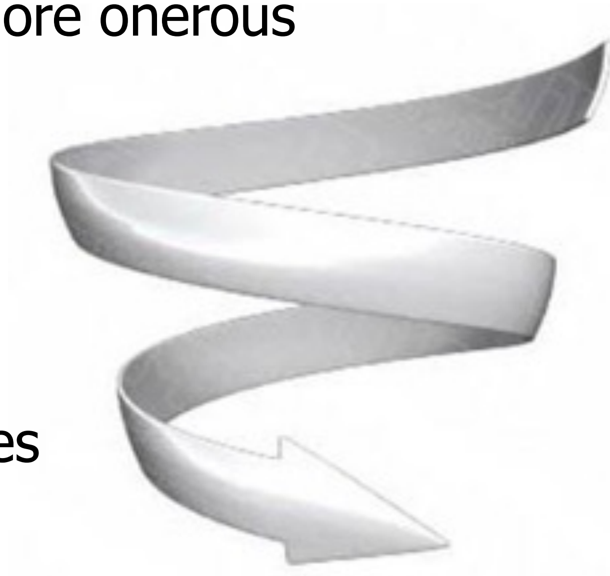
Self-perpetuating Algorithmic Biases

Credit scoring algorithm suggests Joe has high risk of defaulting

Hence, Joe needs to take a loan at a higher interest rate

Hence, Joe has to make payments that are more onerous

Hence, Joe's risk of defaulting has increased



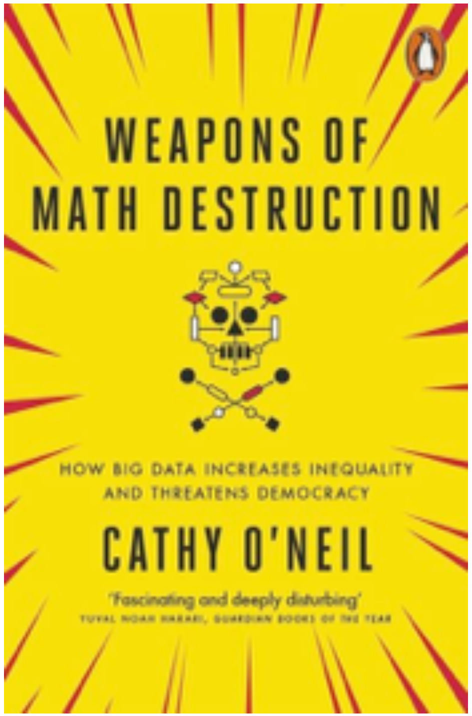
Same happens with stop-and-frisk of minorities
Further increasing incarceration rates

Lesson

Where Did We Go Wrong?

Misconception: Data and ML-Tools Are Neutral!

Where Did We Go Wrong?



- Data at best reflects the current state of the world
  - Acts as a social mirror
- Proxies
  - Protected attributes redundantly encoded in observables
- Correctness
  - Noise in training labels
- Incomplete/Sample size disparity
  - More data from one group

M. Hardt (2014): "How big data is unfair". Medium.

The Achilles Heels of Traditional ML

Even assuming no training data biases, ML decisions

1. Often optimize for a single decision outcome goal, ignoring
   - Fairness: Equal prediction accuracy for all salient social groups
   - Worst-cases: Lower bound worst-case prediction accuracy
   - Norms: Should use or not use data in a specific manner

The Achilles Heels of Traditional ML

Even assuming no training data biases, ML decisions

1. Often optimize for a single decision outcome goal, ignoring
   - Fairness: Equal prediction accuracy for all salient social groups
   - Worst-cases: Lower bound worst-case prediction accuracy
   - Norms: Should use or not use data in a specific manner

2. Optimal for a static NOT an evolving society, because
   - Training data becomes unrepresentative
   - Feedback loops are not accounted for in the first place
   - Decision outcome goals change over time!

Can We Guard the Achilles Heels?

- Can we account for fairness & other norms in ML decision making?
  - Possibly yes!
    - Lots of ongoing research on specifying multiple decision objectives to algorithms

- Can we design ML decision making for an evolving society?
  - Not sure!
    - Continuing engagement with social scientists & legal scholars
    - Focus on procedures than outcomes

Fairness in Machine Learning

- Very recent and emerging field
- Two broad categories of fairness
  - Group Fairness:
    Decision should equally impact different groups
  - Individual Fairness
    Similar individuals should be treated similarly
- A prime example of group unfairness is Discrimination

Defining Discrimination

- A first approximate normative / moralized definition:

  wrongfully impose a **relative disadvantage** on persons based on their membership in some **salient social group**

- Challenge: How to operationalize the definition?
  - How to make it clearly distinguishable, measurable, and understandable in terms of empirical observations

Need to Operationalize Two Fuzzy Notions

1. What constitutes a salient social group?

2. What constitutes a wrongful relative disadvantage?

Need to Operationalize Two Fuzzy Notions

1. **What constitutes a salient social group?**

   Depends on existing legislations

2. What constitutes a wrongful relative disadvantage?

Regulated Domains in the US

- Credit (Equal Credit Opportunity Act)
- Education (Civil Rights Act of 1964; Education Amendments of 1972)
- Employment (Civil Rights Act of 1964)
- Housing (Fair Housing Act)
- ‘Public Accommodation’ (Civil Rights Act of 1964)

Regulated Domains in the US

- Credit (Equal Credit Opportunity Act)
- Education (Civil Rights Act of 1964; Education Amendments of 1972)
- Employment (Civil Rights Act of 1964)
- Housing (Fair Housing Act)
- ‘Public Accommodation’ (Civil Rights Act of 1964)

Extends to marketing and advertising; not limited to final decision

Legally Recognized ‘Protected Classes’

Race (Civil Rights Act of 1964); **Color** (Civil Rights Act of 1964); **Sex** (Equal Pay Act of 1963; Civil Rights Act of 1964); **Religion** (Civil Rights Act of 1964); **National origin** (Civil Rights Act of 1964); **Citizenship** (Immigration Reform and Control Act); **Age** (Age Discrimination in Employment Act of 1967); **Pregnancy** (Pregnancy Discrimination Act); **Familial status** (Civil Rights Act of 1968); **Disability status** (Rehabilitation Act of 1973; Americans with Disabilities Act of 1990); **Veteran status** (Vietnam Era Veterans' Readjustment Assistance Act of 1974; Uniformed Services Employment and Reemployment Rights Act); **Genetic information** (Genetic Information Nondiscrimination Act)

Need to Operationalize Two Fuzzy Notions

1. What constitutes a salient social group?

   Depends on existing legislations

2. **What constitutes a wrongful relative disadvantage?**

Three Measures of Discrimination

Disparate treatment: Targets **direct discrimination**

Requires: $P(\hat{y} \mid \mathbf{x}, z) = P(\hat{y} \mid \mathbf{x})$

Disparate impact: Targets **indirect discrimination,**
when **historical labels are biased**

Requires: $P(\hat{y} = 1 \mid \text{♀}) = P(\hat{y} = 1 \mid \text{♂})$

Disparate mistreatment: Targets **indirect discrimination,**
when **ground truth available**

Requires: $P(y \neq \hat{y} \mid \text{♂}) = P(y \neq \hat{y} \mid \text{♀})$

Also for other misclassification rates

Broad Categories of Fairness

- Group Fairness:
  Decision should equally impact different groups

- Individual Fairness
  Similar individuals should be treated similarly

Most of the focus has been on supervised classification with some form of objective ground truth

Going Beyond Classification

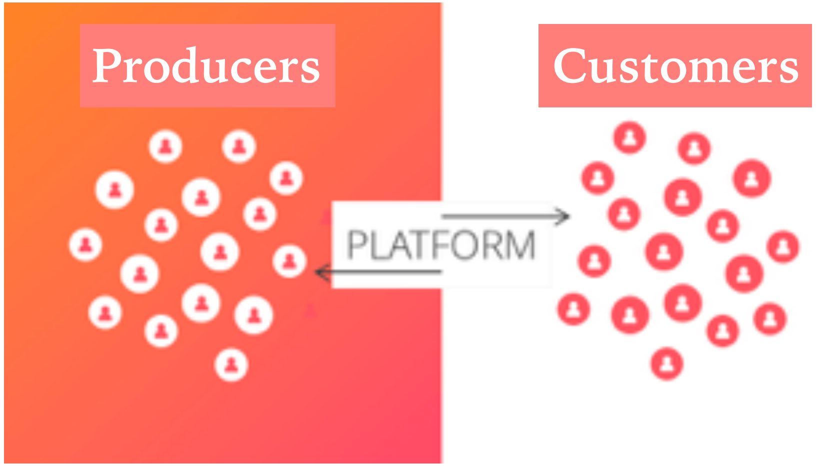
- In multiple learning systems, no objective true label exists, rather many subjective personal preferences
    - Ranking
    - Recommendation
    - Matching

Going Beyond Classification

- In multiple learning systems, no objective true label exists rather many subjective personal preferences
  - Ranking
  - Recommendation
  - Matching

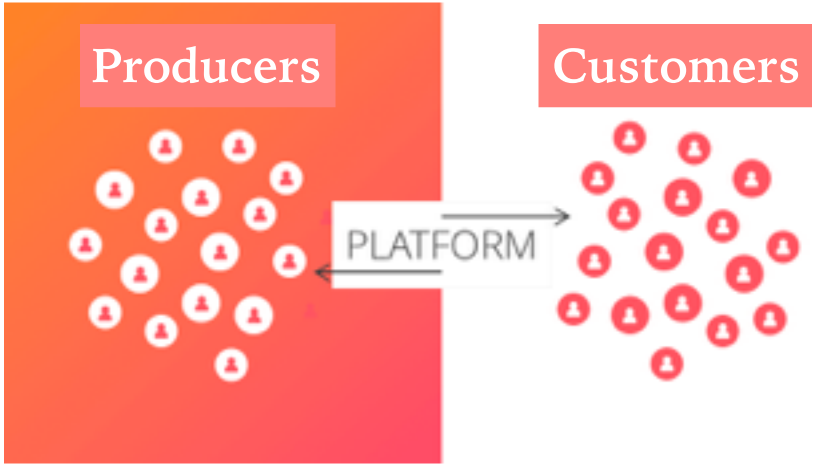
How to ensure fairness in decision making that consider preferences?

Fairness in Two-Sided Platforms



- Ecommerce (Amazon, Flipkart): Sellers & Buyers
- Ride-hailing (Uber, Ola): Drivers & Passengers
- Content streaming (Spotify, Youtube): Artists & Listeners
- Donation (DonorsChoose, Kickstarter): Donors & Recipients

Fairness in Two-Sided Platforms



- Ecommerce (Amazon, Flipkart): Sellers & Buyers
- Ride-hailing (Uber, Ola): Drivers & Passengers
- Content streaming (Spotify, Youtube): Artists & Listeners
- Donation (DonorsChoose, Kickstarter): Donors & Recipients

Fairly consider preferences of one or both sides while developing search, recommendation or matching systems

Today's Focus

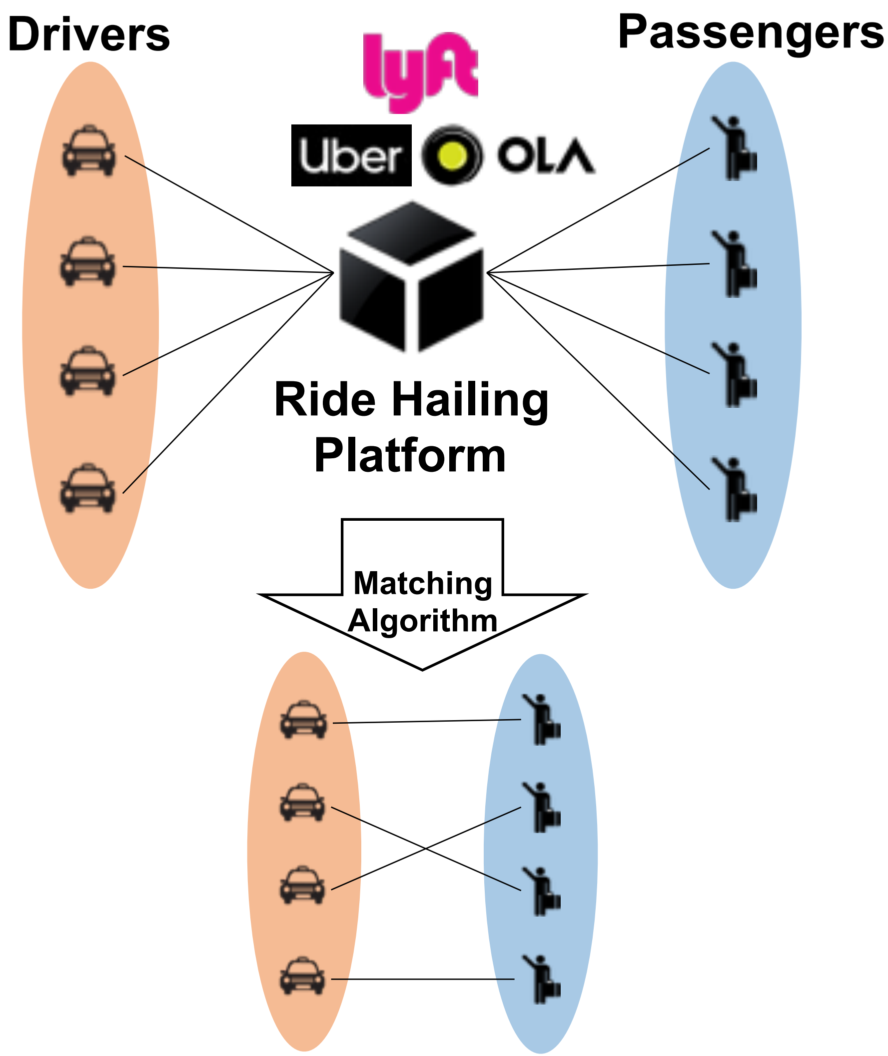
Fair Matching of Drivers & Passengers

Two-Sided Fairness for Repeated Matchings in Two-Sided Markets: A Case Study of a Ride-Hailing Platform

Tom Sühr, Asia J. Biega, Meike Zehlike,
Krishna P. Gummadi and Abhijnan Chakraborty

ACM KDD 2019

Ride Hailing Platforms



Ride Hailing Platforms

- Ride hailing industry is now valued at $61.3 billion and expected to grow to $218 billion by 2025
- Uber and Lyft have launched mega-IPOs in recent years

Ride Hailing Platforms

- Ride hailing industry is now valued at $61.3 billion and expected to grow to $218 billion by 2025
- Uber and Lyft have launched mega-IPOs in recent years

What about the drivers in these platforms?

Concerns about Drivers

Uber, Lyft Drivers Stage Strike To Draw Attention To Drivers' Plight

CBS

The Uber drivers forced to sleep in parking lots to make a decent living

The Guardian

They were sold a fantasy of middle-class life. Now Ola and Uber drivers face a brutal reality

QUARTZ

Dataset Gathered

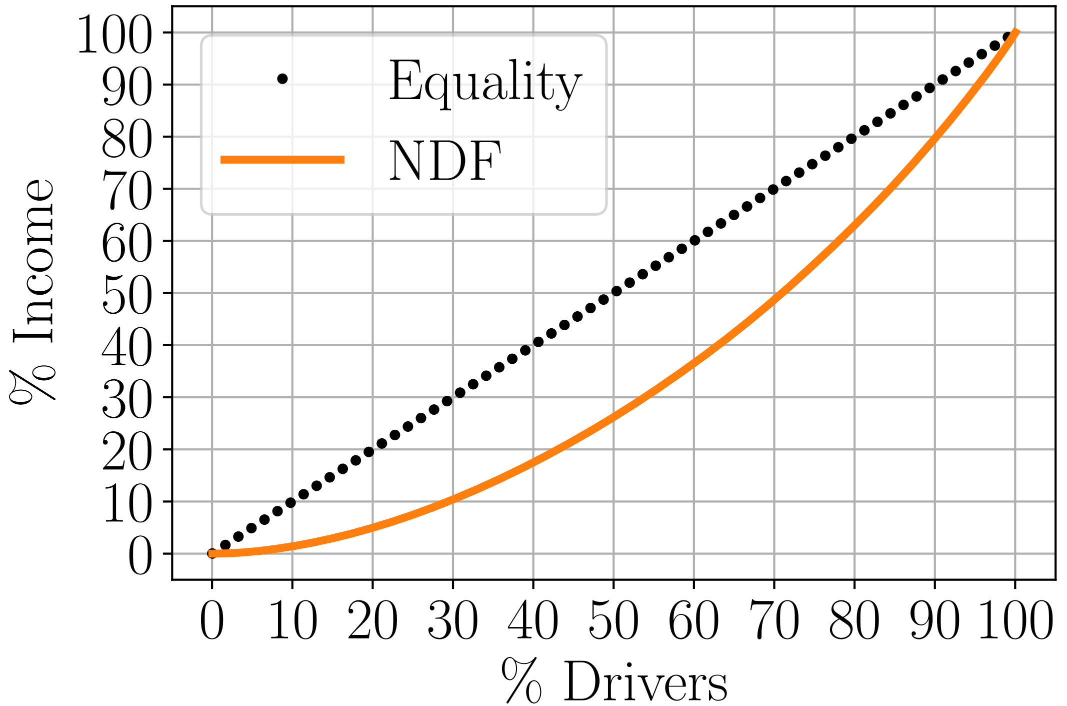
- Got data from an Asian taxi riding platform for a particular city for one month
    - About 15,000 drivers
    - 4.6 million ride assignments
- Always more drivers available than ride requests
    - Supply exceeds demand

Dataset Gathered

- Got data from an Asian taxi riding platform for a particular city for one month
  - About 15,000 drivers
  - 4.6 million ride assignments
- Always more drivers available than ride requests
  - Supply exceeds demand

What is the distribution of driver income?

Distribution of Driver Income

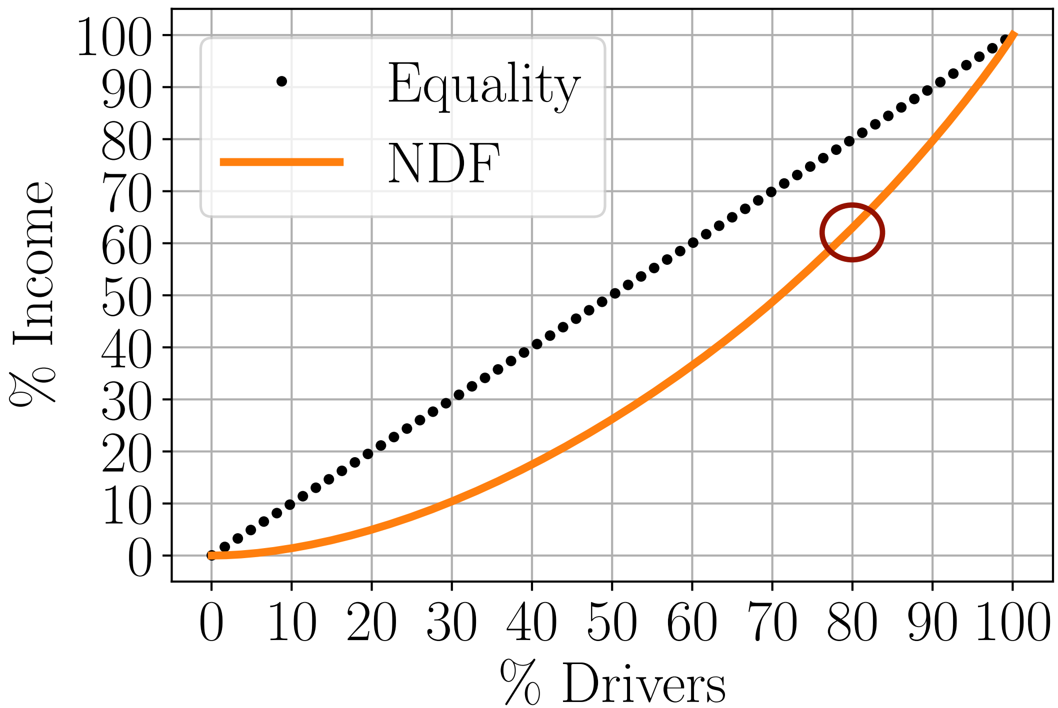


Lorenz Curve

Y-axis: Cumulative % of total income

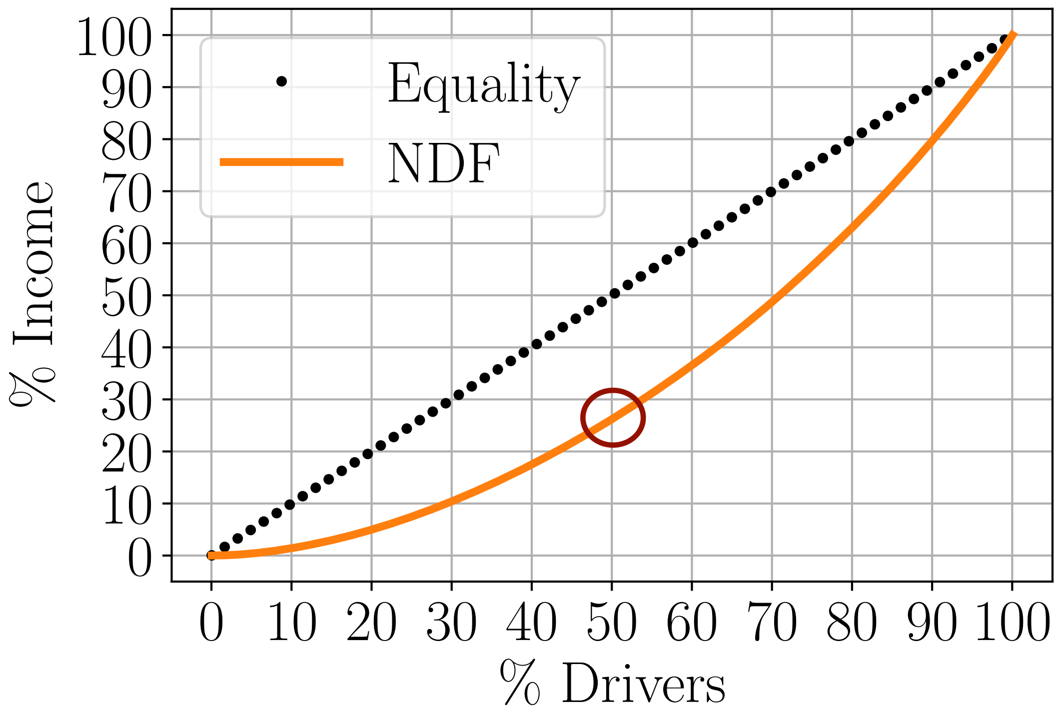
X-axis: Cumulative % of the corresponding drivers

Inequality in Driver Income



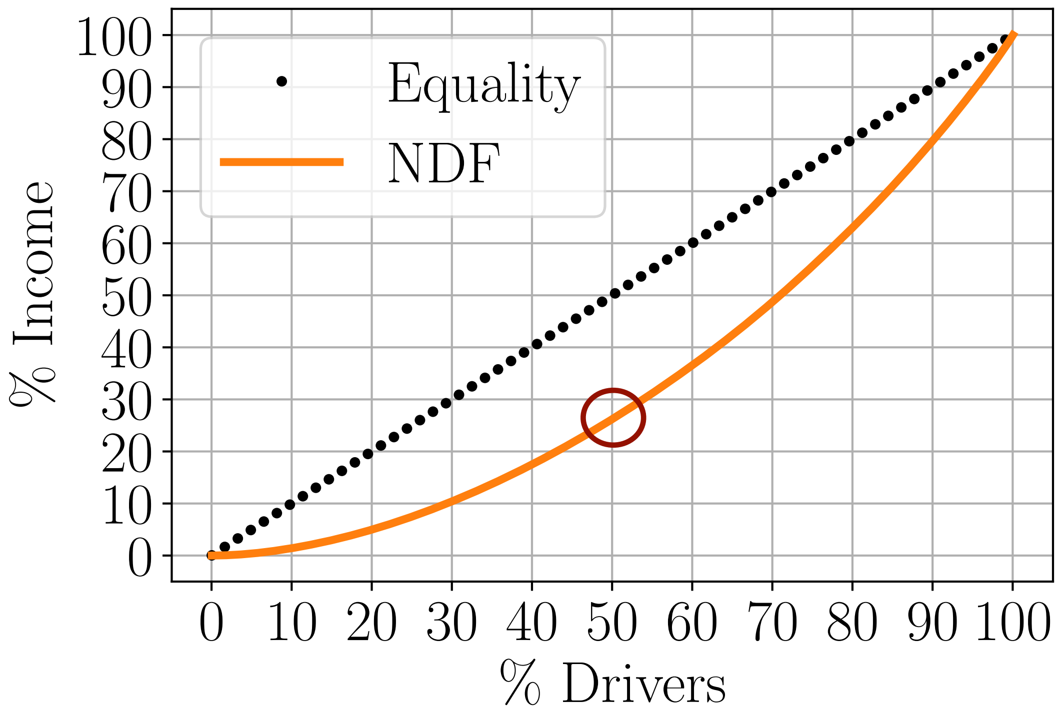
- Most successful 20% of the drivers earned 40% of total income

Inequality in Driver Income



- Most successful 20% of the drivers earned 40% of total income
- 50% of the drivers only earned 27% of total income

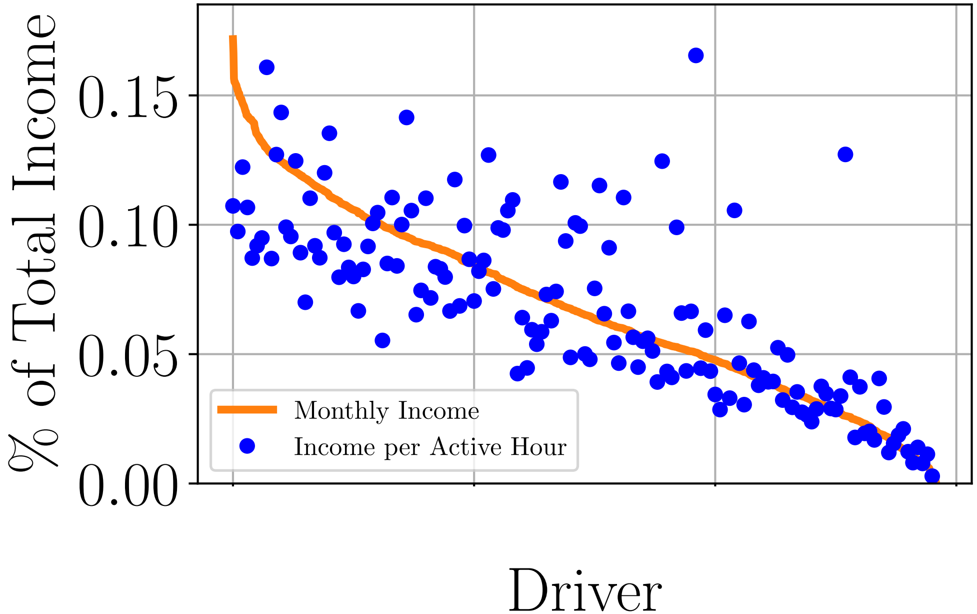
Inequality in Driver Income



- Most successful 20% of the drivers earned 40% of total income
- 50% of the drivers only earned 27% of total income

Is it due to the difference in activity levels?

Inequality in Driver Income



Similar pattern even after considering per-hour income
i.e., by normalizing w.r.t. number of hours different drivers are active

Towards Fairer Ride-Hailing

California just passed a landmark law to regulate Uber and Lyft

Vox

California Bill Makes App-Based Companies Treat Workers as Employees

The New York Times

With new tax and minimum wage, Seattle is latest battleground in Uber and Lyft's feud with regulators

GeekWire

Towards Fairer Ride-Hailing

California just passed a landmark law to regulate Uber and Lyft

Vox

California Bill Makes App-Based Companies Treat Workers as Employees

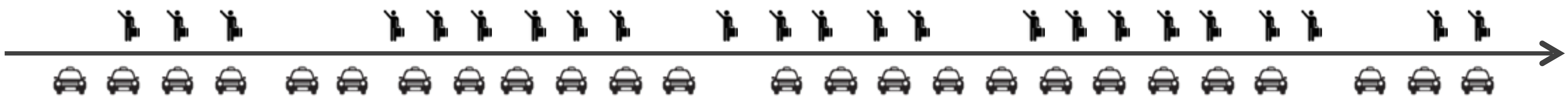
The New York Times

With new tax and minimum wage, Seattle is latest battleground in Uber and Lyft's feud with regulators

GeekWire

What would be a fair distribution of income on such platforms?

Modeling a Ride Hailing Platform

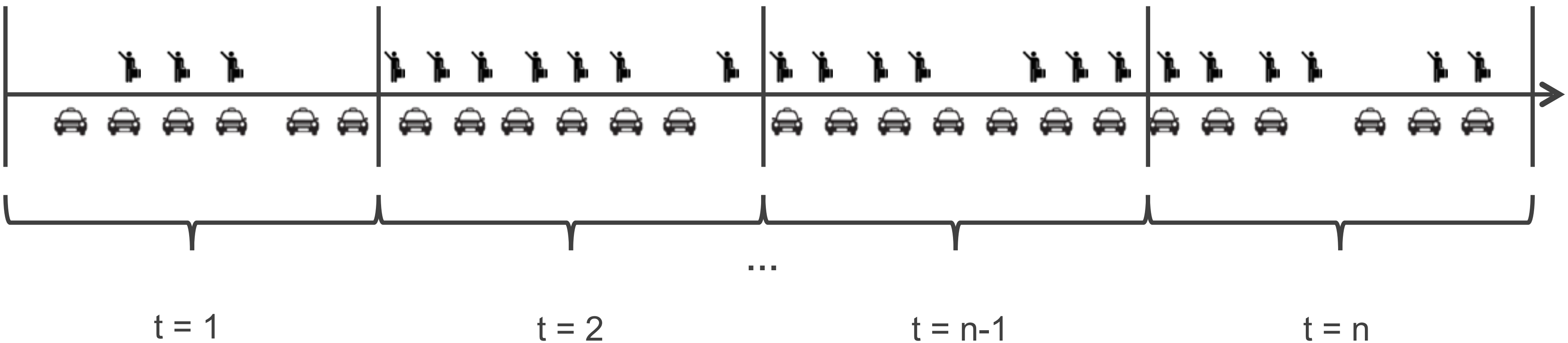


Platform produces a sequence of matches between drivers and passengers over time

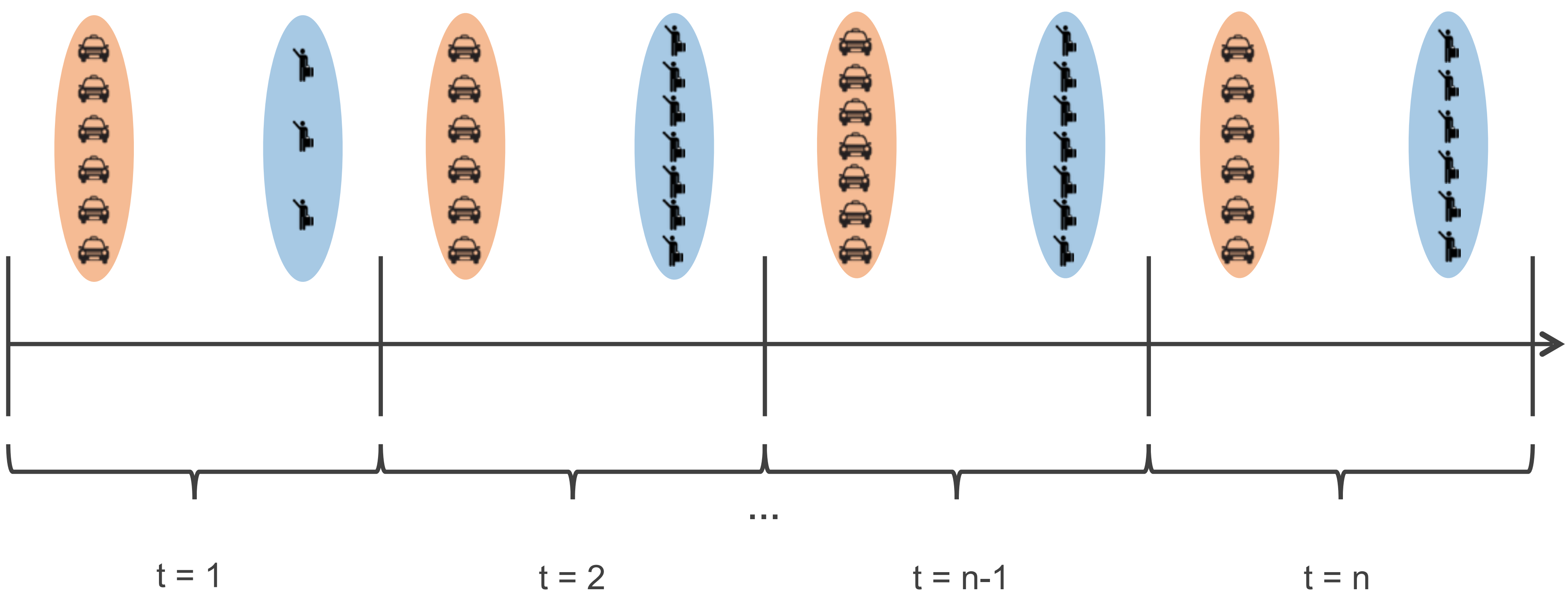
Modeling a Ride Hailing Platform

Platform groups incoming passenger requests within a short period of time to form matching rounds

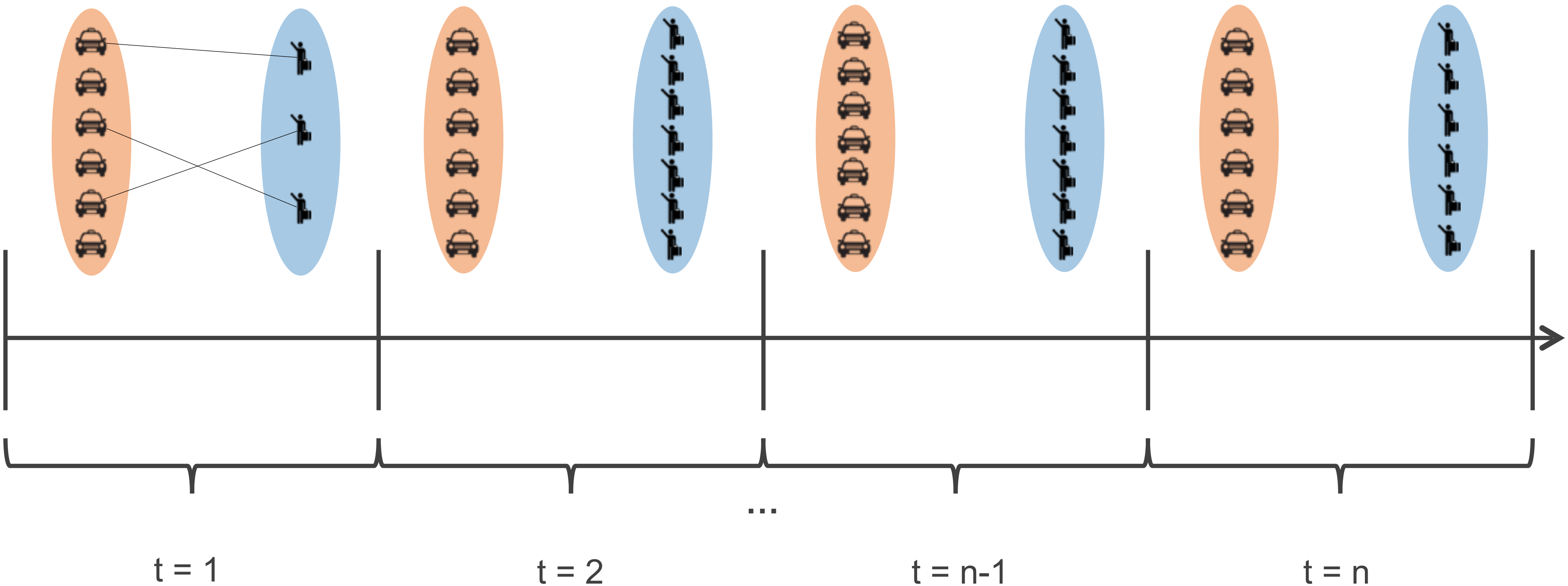
Modeling a Ride Hailing Platform



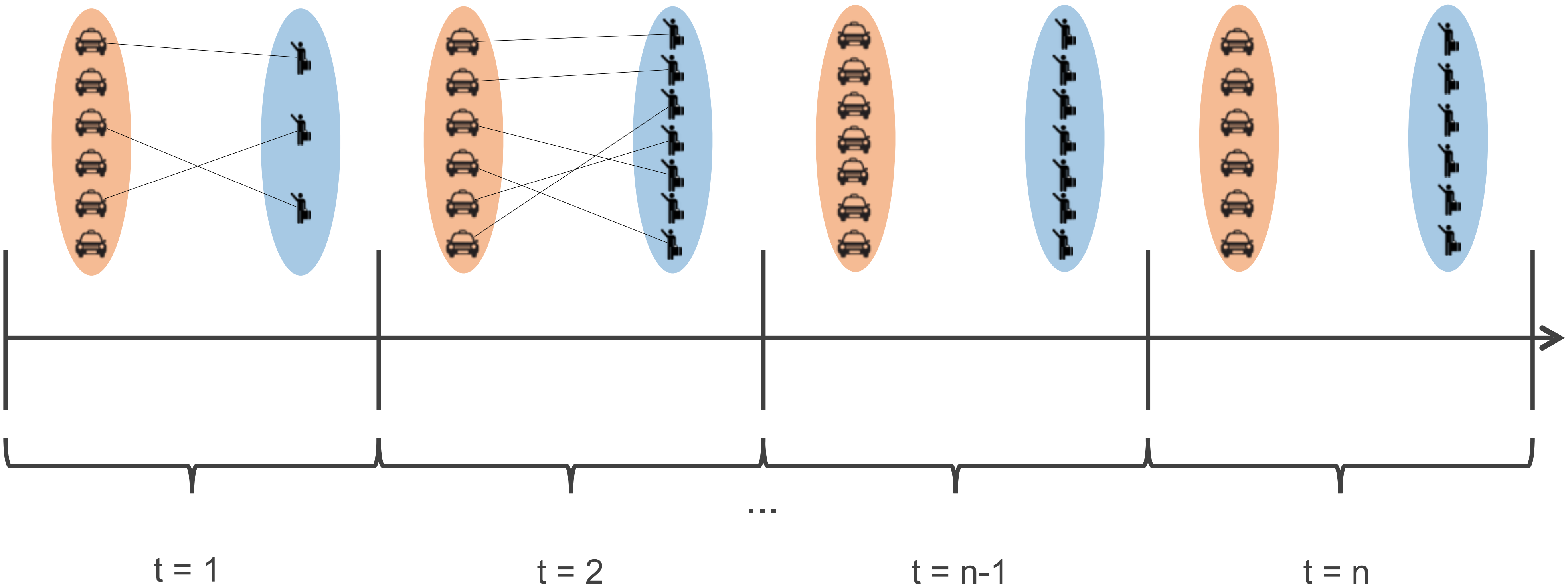
Modeling a Ride Hailing Platform



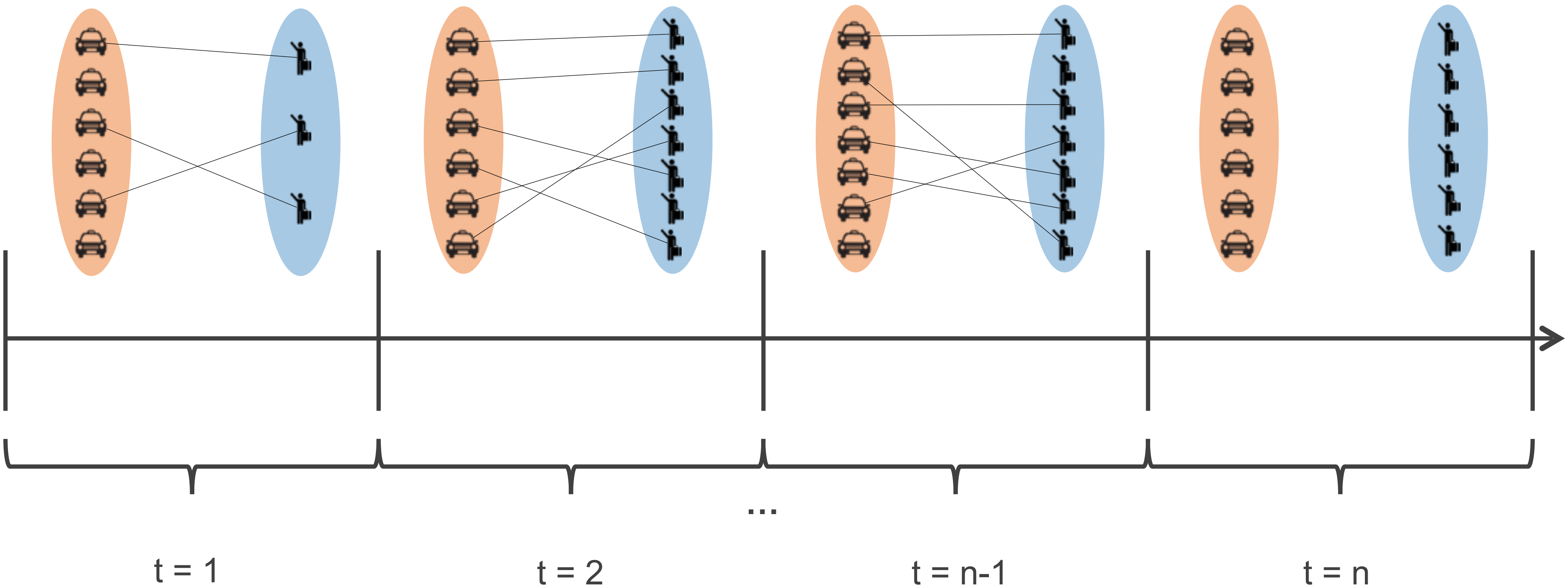
Modeling a Ride Hailing Platform



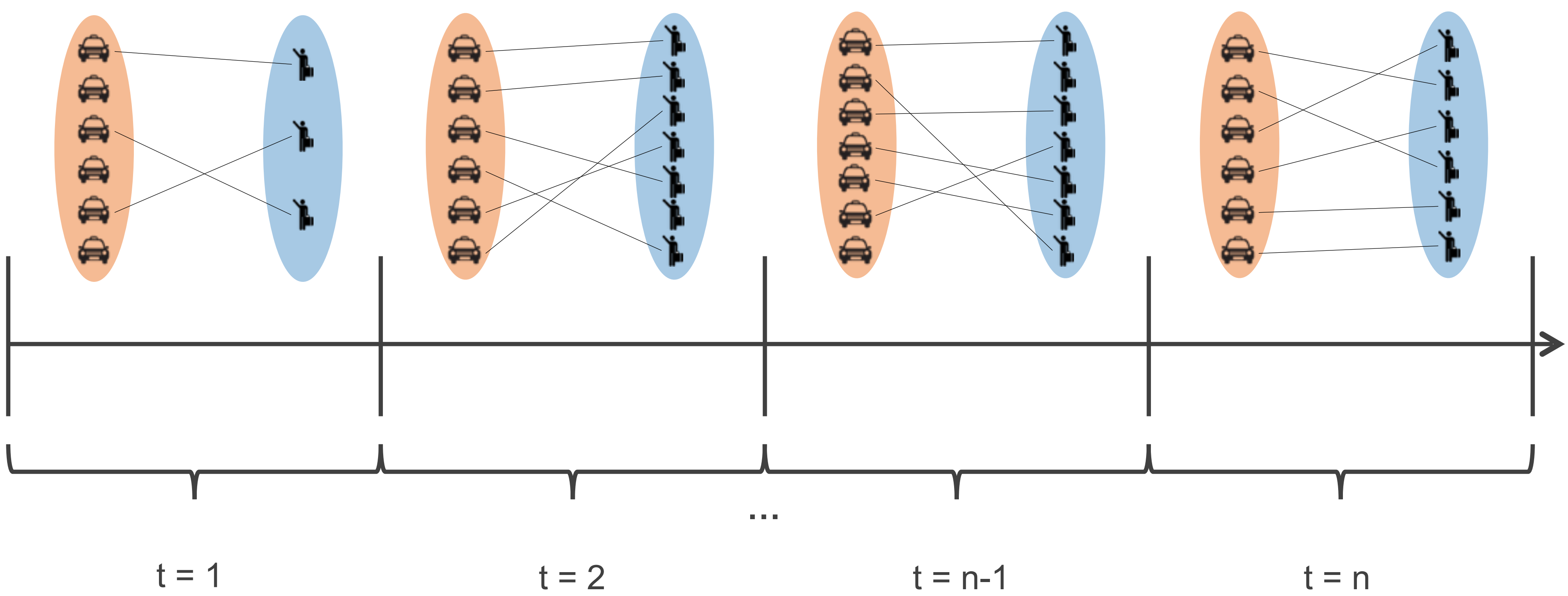
Modeling a Ride Hailing Platform



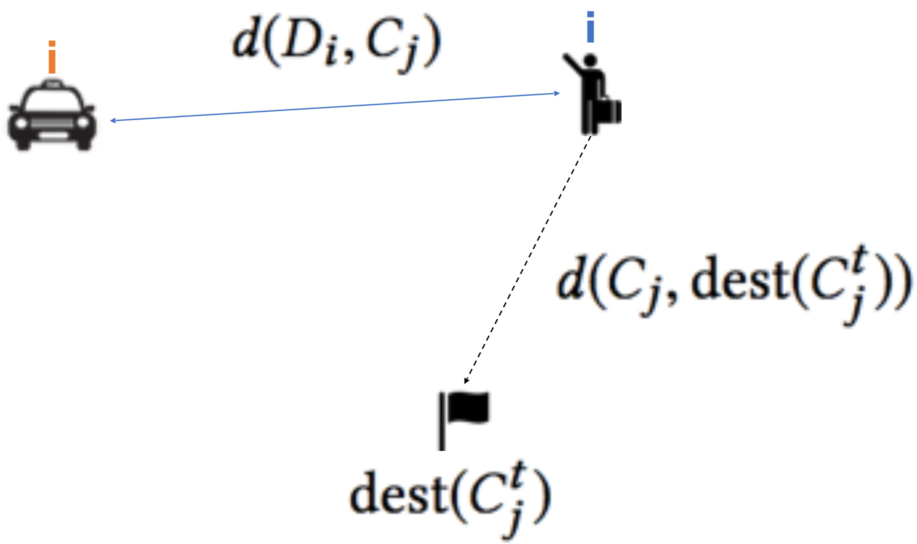
Modeling a Ride Hailing Platform



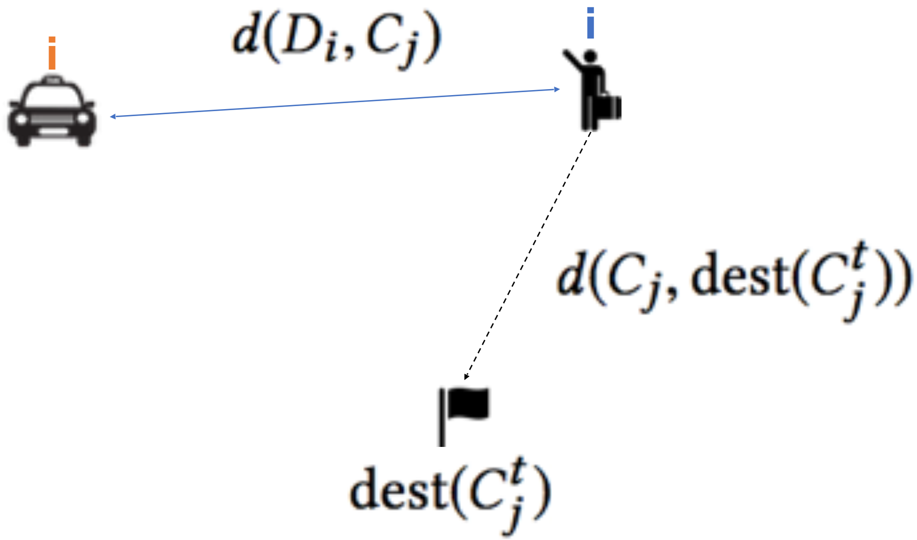
Modeling a Ride Hailing Platform



Modeling Utility for Both Sides



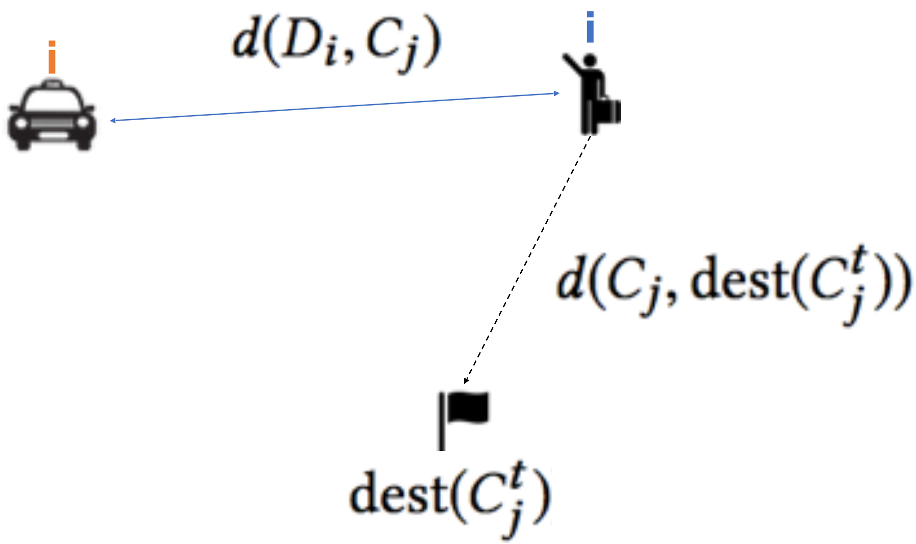
Modeling Utility for Both Sides



- Utility for the passengers: waiting time

$$U_C(i, j) = -d(D_i, C_j)$$

Modeling Utility for Both Sides



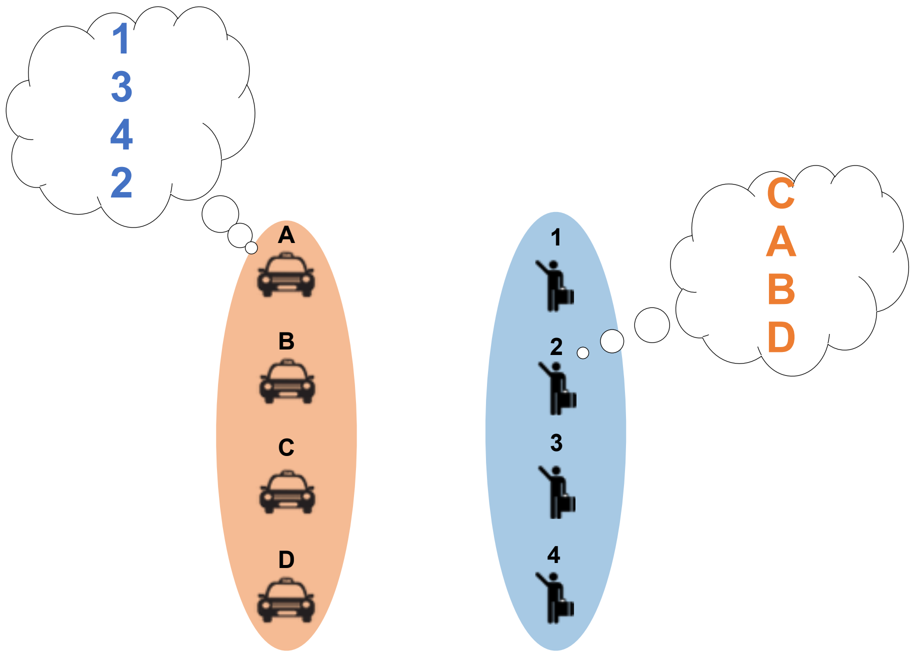
- Utility for the passengers: waiting time

$$U_C(i,j) = -d(D_i, C_j)$$

- Utility for the drivers: effective distance travelled

$$U_D(i,j) = d(C_j, \mathrm{dest}(C_j^t)) - d(D_i, C_j)$$

Naturally Manifest Preferences of Drivers & Passengers



How to fairly match preferences of both sides?

Brief History of Fair Matching

- Long lines of works on fairness in matching markets
  - School admissions
  - Hospital-doctor allocation
  - Kidney exchanges
- Nobel Prize in Economics 2012: Alvin Roth, Lloyd Shapley
- Existing works did not consider repeated matchings over time

How to think about fairness in ride hailing platforms?

Fairness of Repeated Matching

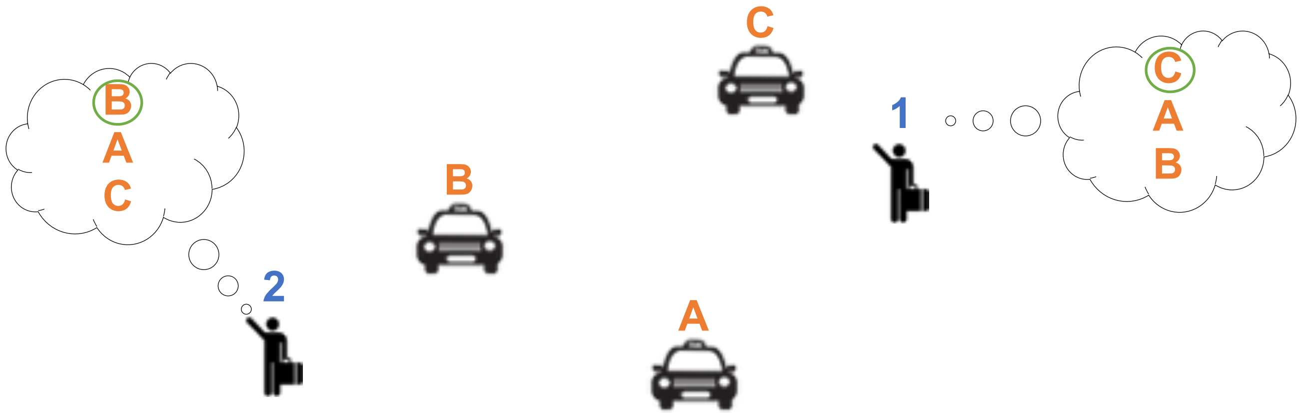
- **Amortized Parity**
  - Over time, sum of received utilities of all drivers should be equal
- **Amortized Proportionality**
  - Over time, sum of received utilities of all drivers should be proportional to the length of time they are active
  - Can also be extended to include other notions of similarity (car type, rating, …)
  - Over time, similar drivers should receive similar utility
- Similar fairness notions for the passengers as well

How good are naive matching methods?

Passenger-centric Method: NDF

Nearest Driver First (NDF)

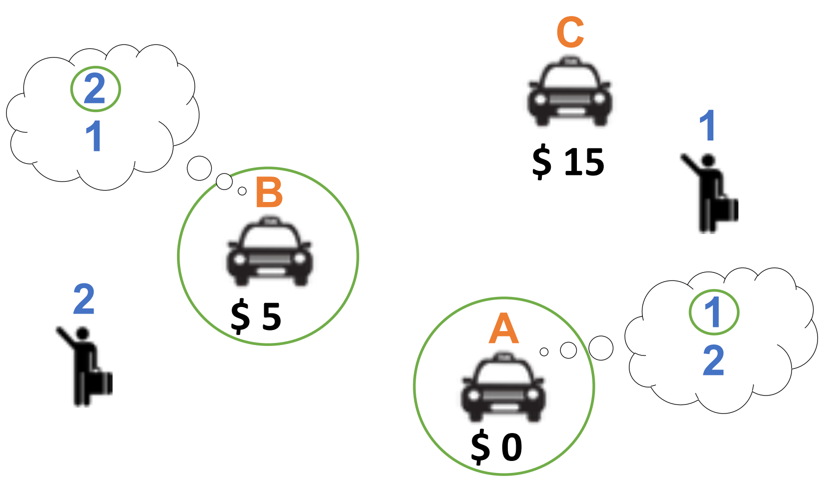
- Only consider passengers' preferences
- Match nearest driver to passenger in order to maximize utility



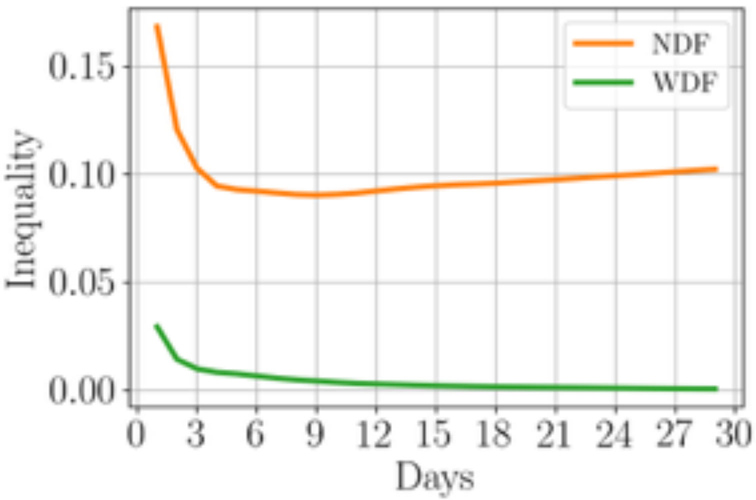
Driver-centric Method: WDF

Worst-off Driver First (WDF)

- Prioritize preferences of drivers with least accumulated utilities
- Drivers are assigned in that priority order

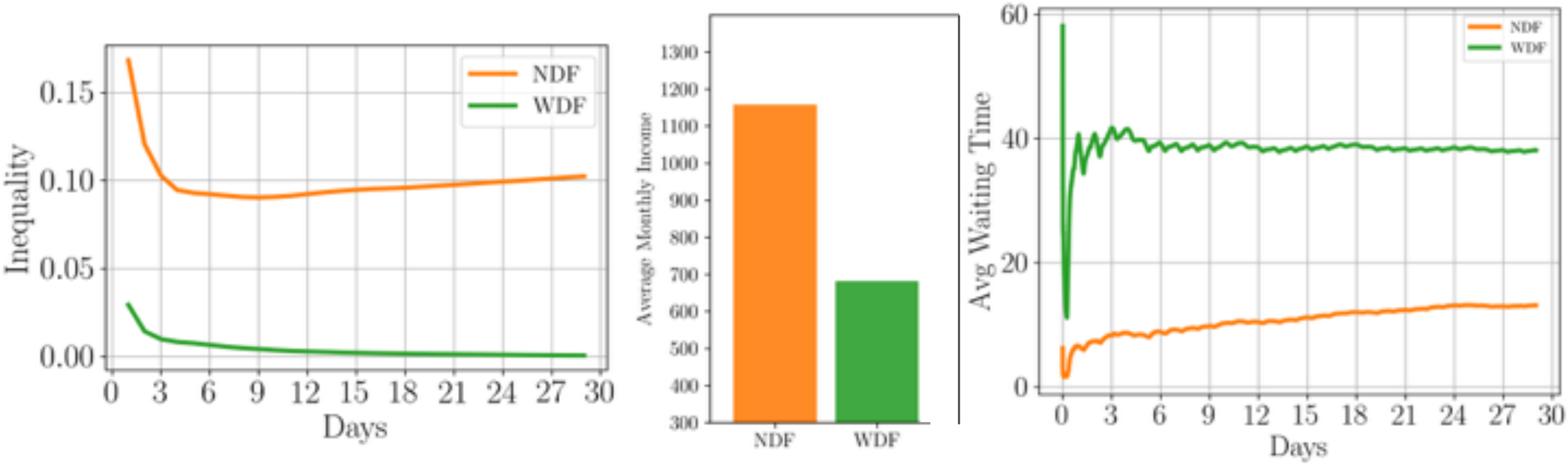


Effects of WDF and NDF



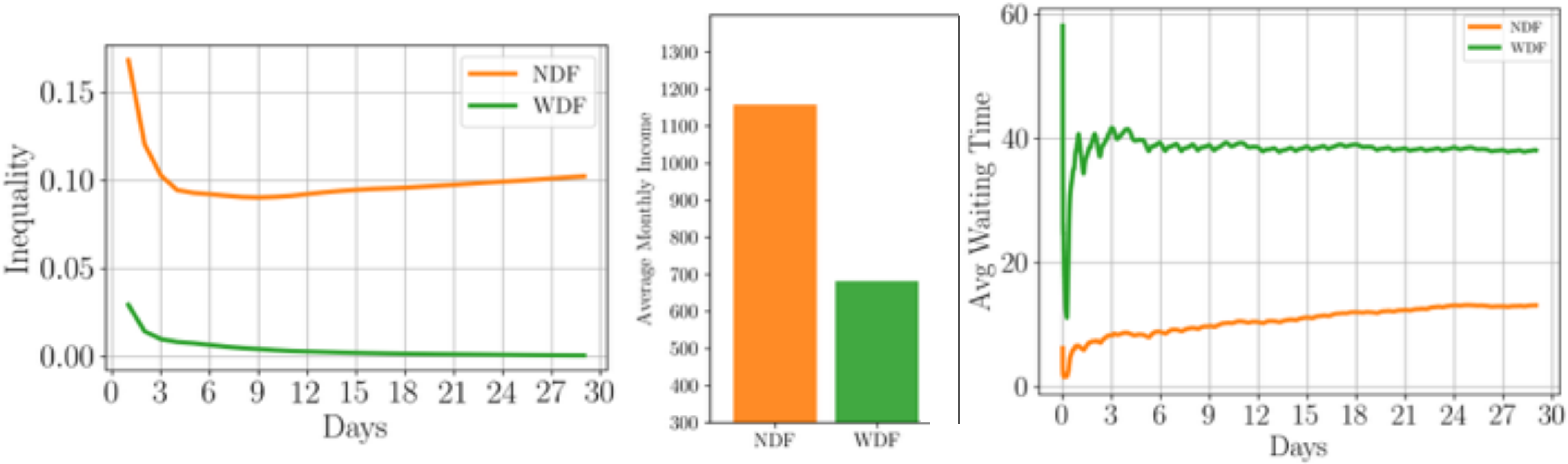
- In WDF, the inequality in driver income decreases to zero

Effects of WDF and NDF



- In WDF, the inequality in driver income decreases to zero
- Lowers average income of drivers and increases average waiting time for passengers

Effects of WDF and NDF



- In WDF, the inequality in driver income decreases to zero
- Lowers average income of drivers and increases average waiting time for passengers

Naively optimizing for one side does not help!

Our Proposal: Take Two Sides Together

$$\begin{aligned} \underset{M}{\text{minimize}} \quad & \lambda \cdot \text{inequality}_D(M) + (1 - \lambda) \cdot \text{inequality}_C(M) \\ \text{subject to} \quad & \text{constraints ensuring a correct matching.} \end{aligned}$$

Our Proposal: Take Two Sides Together

$$\begin{aligned} \underset{M}{\text{minimize}} \quad & \lambda \cdot \text{inequality}_D(M) + (1-\lambda) \cdot \text{inequality}_C(M) \\ \text{subject to} \quad & \text{constraints ensuring a correct matching.} \end{aligned}$$

- Directly minimizing inequality is NP-Hard
  - Atkinson Index [Schneckenburger et al. 2017]
  - Gini Index [Aleksandrov et al. 2018]
  - Generalized Entropy Index [Kovačević et al. 2012]

Schneckenburger et al., The Atkinson inequality index in multi-agent resource allocation. AAMAS 2017
Aleksandrov et al., Fair Division Minimizing Inequality. Arxiv 2018, EPIA 2019
Kovačević et al., On the Hardness of Entropy Minimization andRelated Problems, IEEE ITW 2012

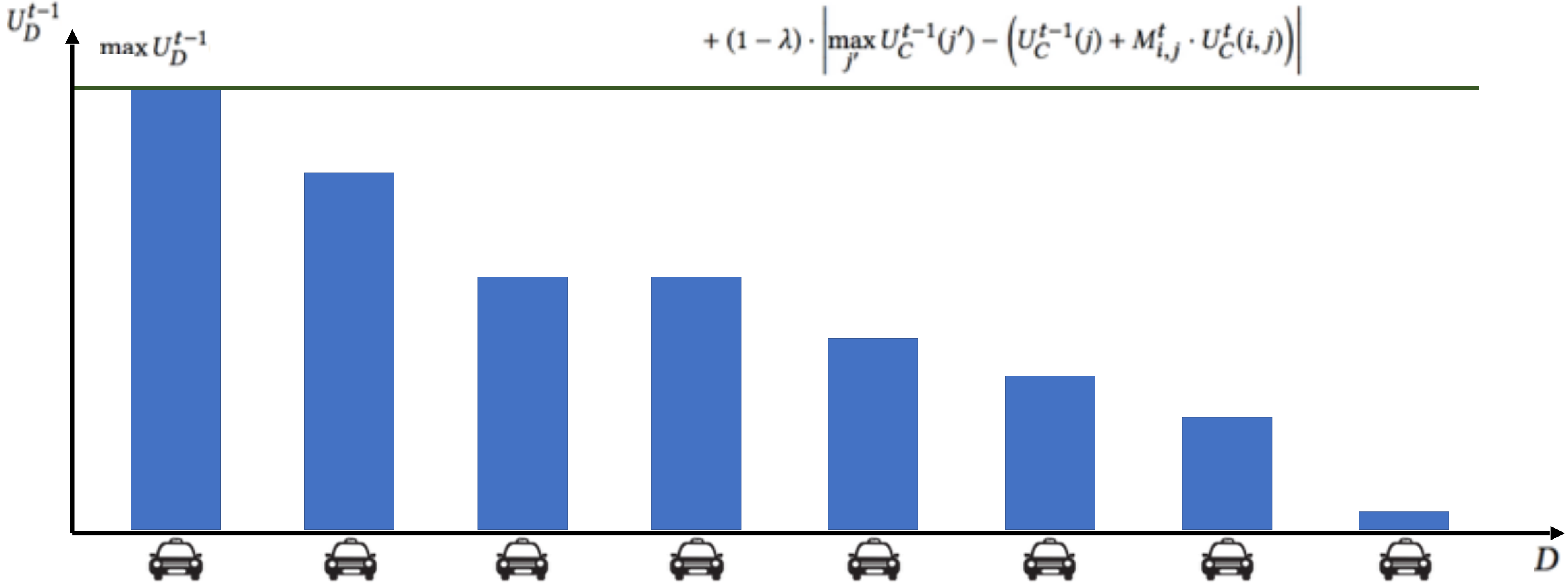
Our Proposal: Take Two Sides Together

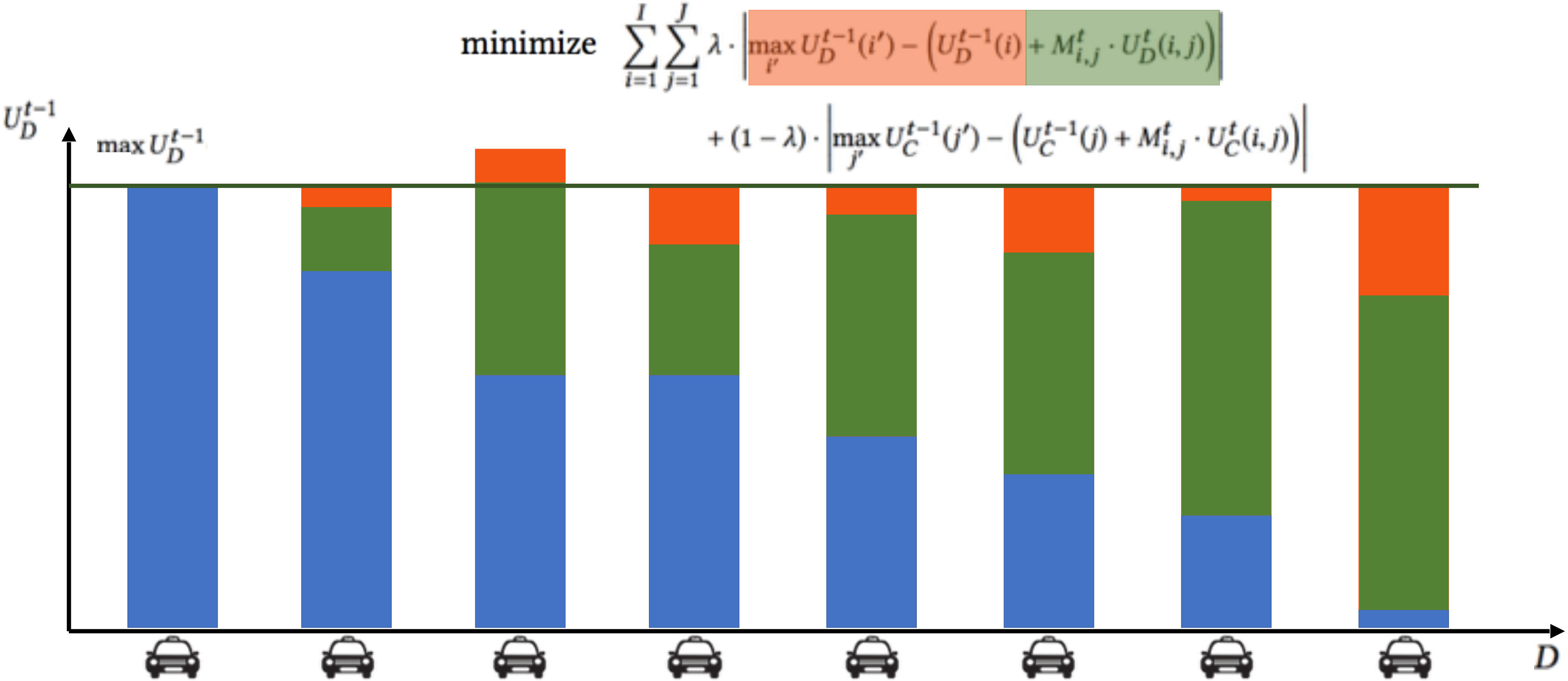
$$\begin{array}{ll} \underset{M}{\text{minimize}} & \lambda \cdot \text{inequality}_D(M) + (1-\lambda) \cdot \text{inequality}_C(M) \\ \text{subject to} & \text{constraints ensuring a correct matching.} \end{array}$$

- Directly minimizing inequality is NP-Hard
  - Atkinson Index [Schneckenburger et al. 2017]
  - Gini Index [Aleksandrov et al. 2018]
  - Generalized Entropy Index [Kovačević et al. 2012]

Minimize the difference of driver (passenger) utilities from the maximum utility gained by any driver (passenger) so far

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\lambda\cdot\left|\max_{i'}U_D^{t-1}(i') - \left(U_D^{t-1}(i) + M_{i,j}^{t}\cdot U_D^{t}(i,j)\right)\right|$$
$$+ (1-\lambda)\cdot\left|\max_{j'}U_C^{t-1}(j') - \left(U_C^{t-1}(j) + M_{i,j}^{t}\cdot U_C^{t}(i,j)\right)\right|$$

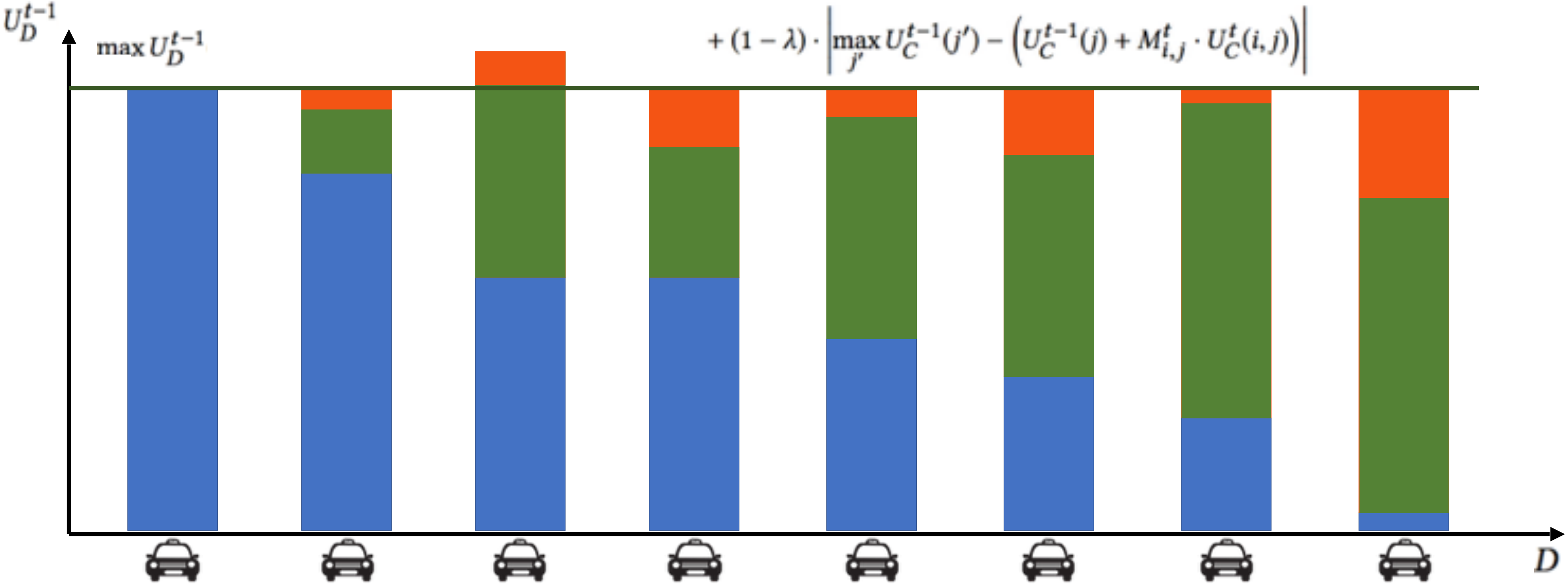




minimize $\sum_{i=1}^{I}\sum_{j=1}^{J}\lambda\cdot\left|\max_{i'}U_D^{t-1}(i')-\left(U_D^{t-1}(i)+M_{i,j}^{t}\cdot U_D^{t}(i,j)\right)\right|$
$+(1-\lambda)\cdot\left|\max_{j'}U_C^{t-1}(j')-\left(U_C^{t-1}(j)+M_{i,j}^{t}\cdot U_C^{t}(i,j)\right)\right|$
U_D^{t-1}
$\max U_D^{t-1}$
D

$$\text{minimize} \quad \sum_{i=1}^{I}\sum_{j=1}^{J} \lambda \cdot \left|\max_{i'} U_D^{t-1}(i') - \left(U_D^{t-1}(i) + M_{i,j}^{t} \cdot U_D^{t}(i,j)\right)\right|$$

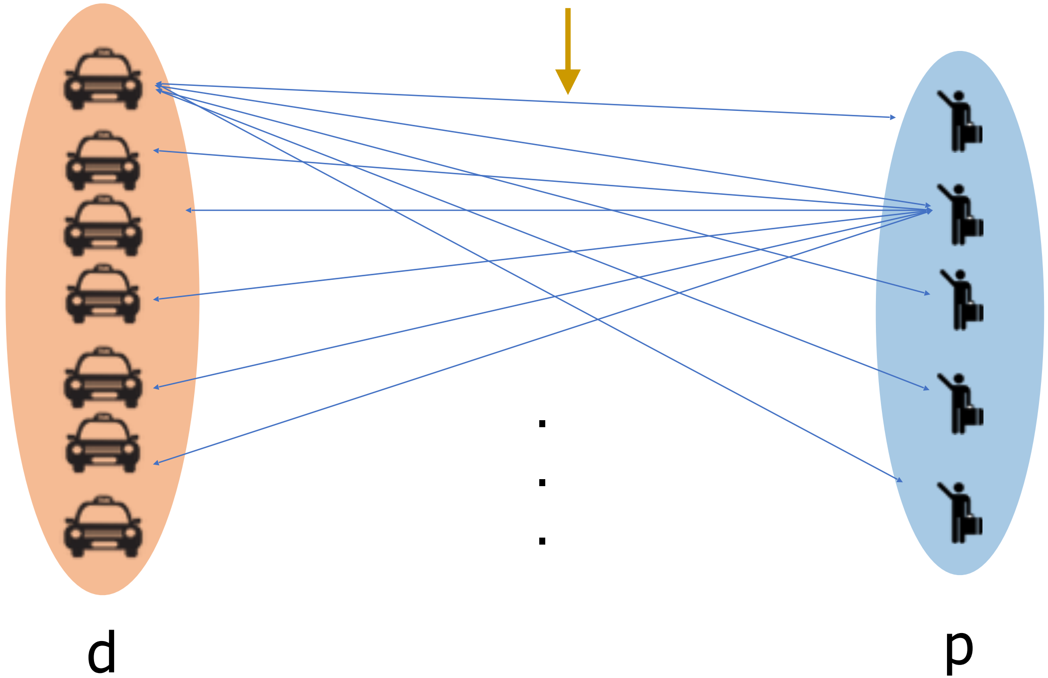
$$+ (1-\lambda) \cdot \left|\max_{j'} U_C^{t-1}(j') - \left(U_C^{t-1}(j) + M_{i,j}^{t} \cdot U_C^{t}(i,j)\right)\right|$$



The problem maps to an Unbalanced Assignment problem

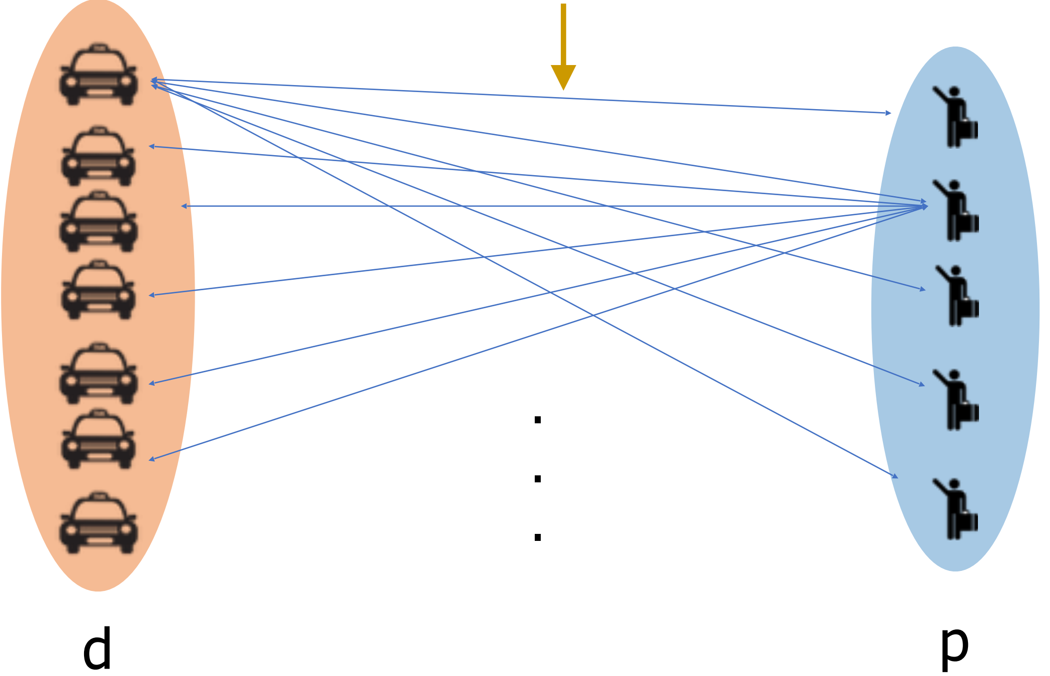
Assignment Problem

$$\lambda \cdot |U_{D_{Max}}^{t-1} - (U_D^{t-1}(i) + u_D^t(i,j))| +$$
$$(1-\lambda) \cdot |U_{C_{Max}}^{t-1} - (U_C^{t-1}(i) + u_C^t(i,j))|$$



Assignment Problem

$$\lambda \cdot |U_{D_{Max}}^{t-1} - (U_D^{t-1}(i) + u_D^t(i,j))| +$$
$$(1-\lambda) \cdot |U_{C_{Max}}^{t-1} - (U_C^{t-1}(i) + u_C^t(i,j))|$$



The goal is to get minimum-cost one-sided-perfect matching

Assignment Problem as Linear Programming

$$minimize \sum_{i=1}^{I}\sum_{j=1}^{J} \lambda \cdot \left| U_{D_{max}}^{t-1}(i') - \left(U_D^{t-1}(i) + M_{i,j}^t \cdot u_D^t(i,j)\right)\right|$$
$$+ (1-\lambda) \cdot \left| U_{C_{max}}^{t-1}(j') - \left(U_C^{t-1}(j) + M_{i,j}^t \cdot u_C^t(i,j)\right)\right|$$

subject to

$$0 \leq M_{i,j}^t \leq 1, \ \forall_{i,j}$$

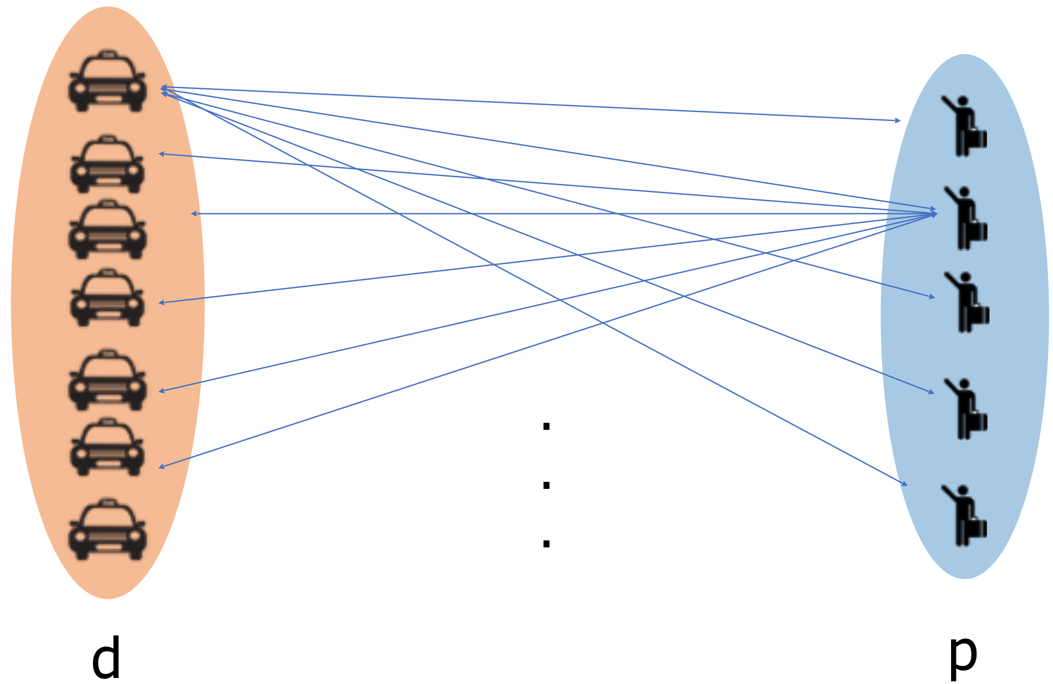
$$\sum_i M_{i,j}^t = R_j^t, \ \forall_j$$

Match only and all requesting passengers

$$\sum_j M_{i,j}^t \leq A_i^t, \ \forall_i$$

Match only available drivers

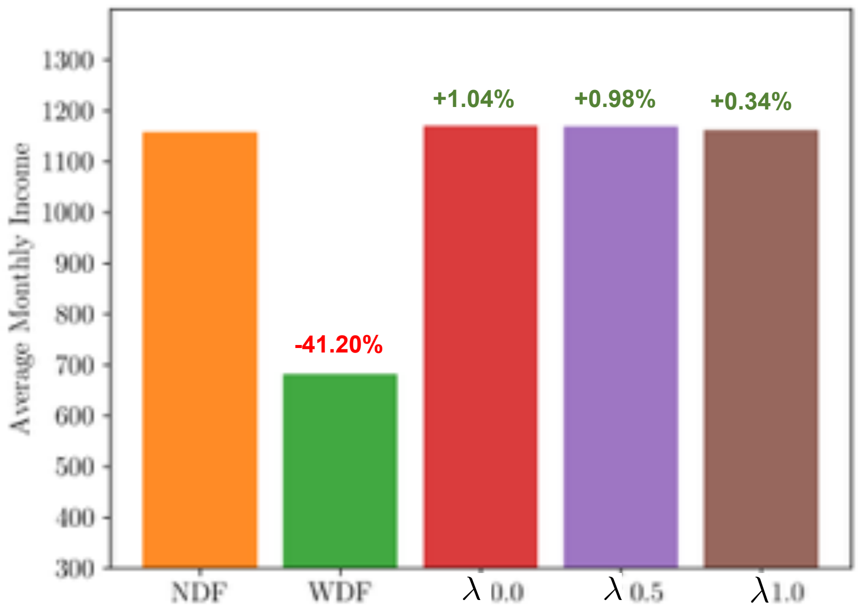
Solution with Bounded Complexity

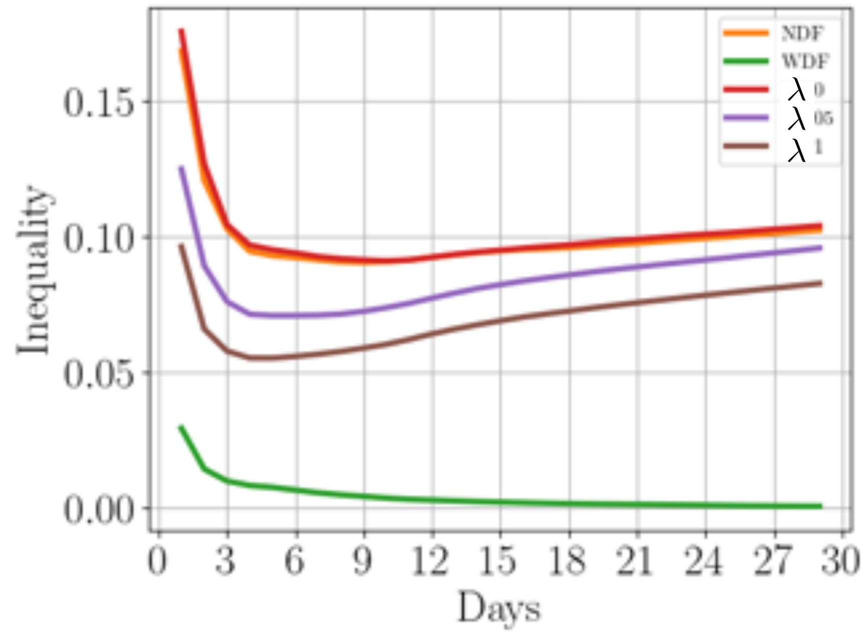


"Unbalanced" extension of Hungarian algorithm proposed in [Ramshaw and Tarjan, 2012] with time complexity $\mathcal{O}(p^2(d + logp))$

Ramshaw and Tarjan, On Minimum-Cost Assignments in Unbalanced Bipartite Graphs. HP Tech Report 2012

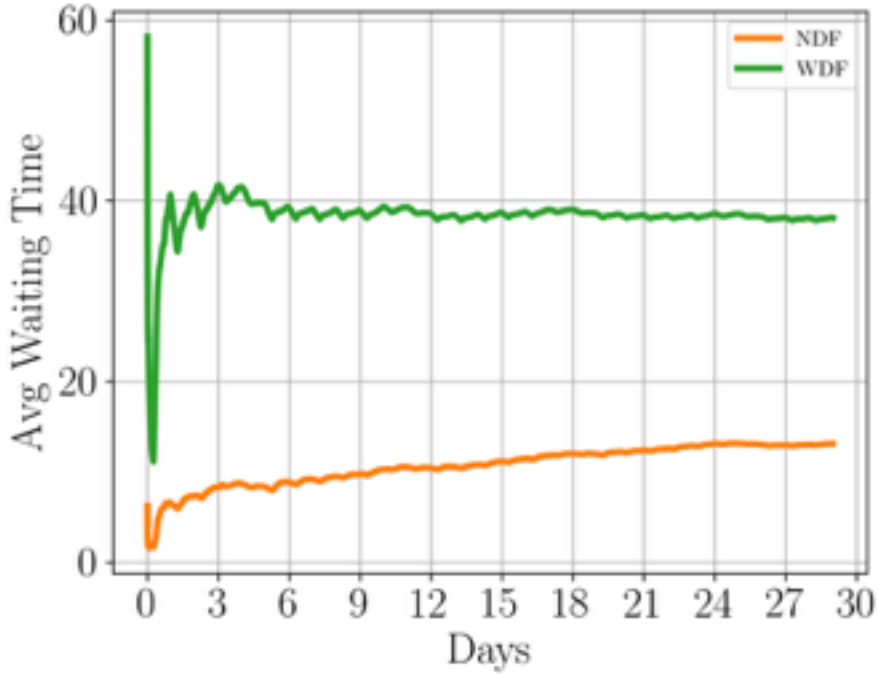
How does our Two-Sided Method Perform?



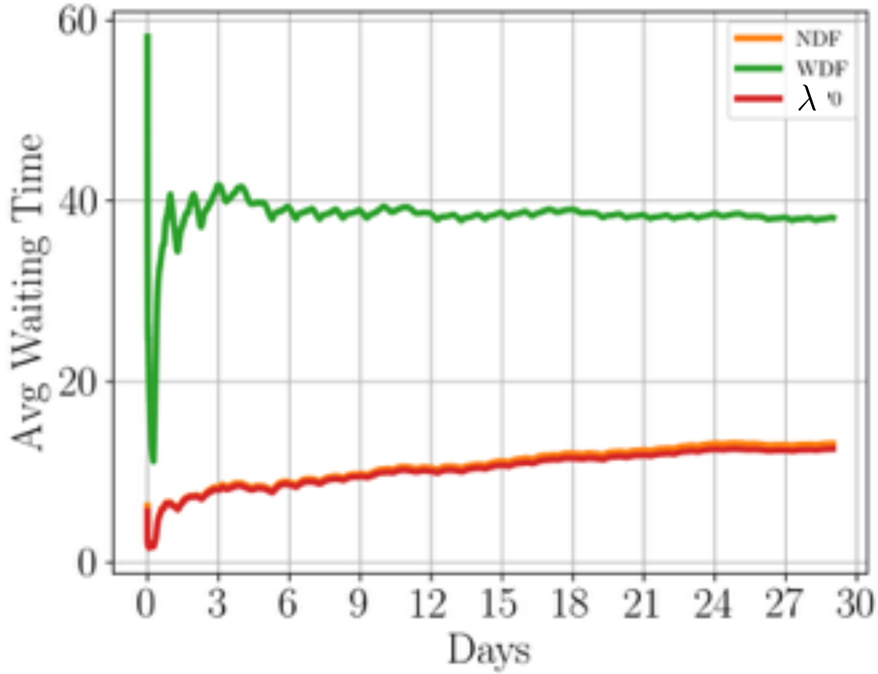


Optimizing for both sides can ensure higher average income for drivers as well as lower inequality

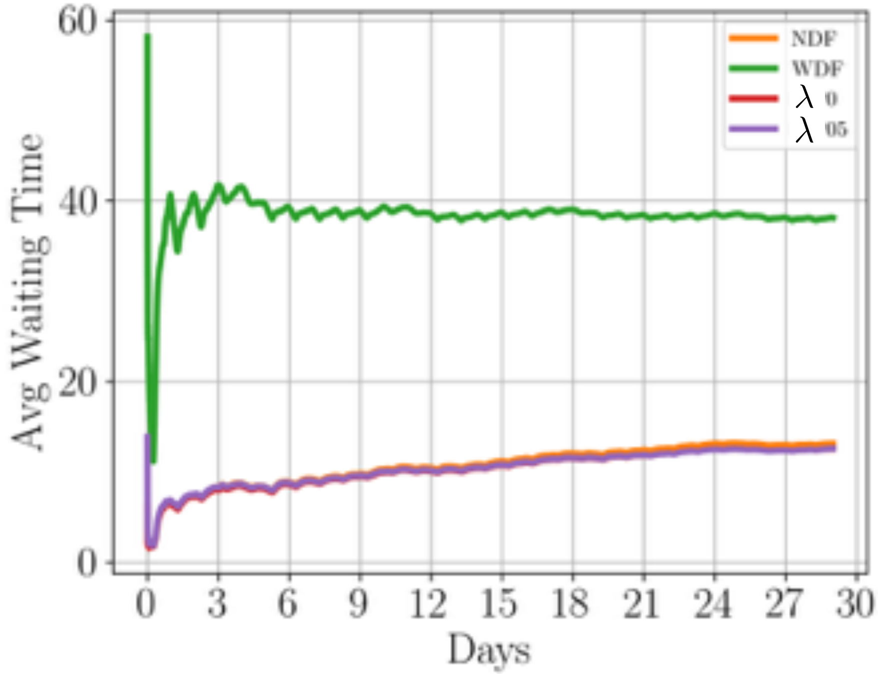
Effect on Waiting Time of Passengers



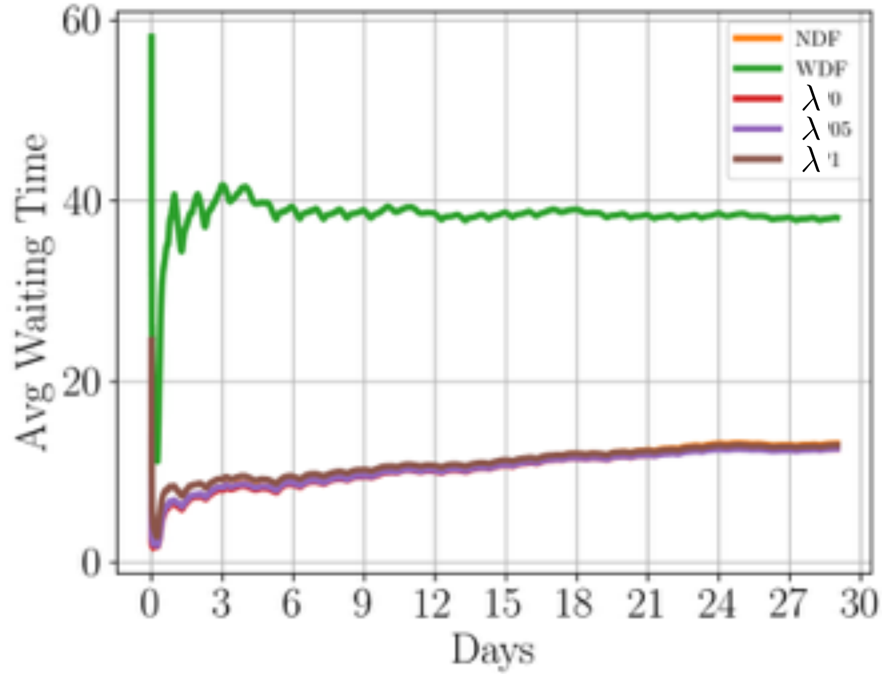
Effect on Waiting Time of Passengers



Effect on Waiting Time of Passengers



Effect on Waiting Time of Passengers



Optimizing for the amortized fairness for drivers do not increase average waiting time for the passengers

Achieving Proportionality

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\lambda\cdot\left|\max_{i'}\frac{U_D^{t-1}(i')}{T_{t-1}(i')}-\left(\frac{U_D^{t-1}(i)}{T_t(i)}+M_{i,j}^t\cdot\frac{U_D^t(i,j)}{T_t(i)}\right)\right|$$
$$+(1-\lambda)\cdot\left|\max_{j'}U_C^{t-1}(j')-\left(U_C^{t-1}(j)+M_{i,j}^t\cdot U_C^t(i,j)\right)\right|$$

- Same mechanism can be applied for utilities normalized by active time on the platform
- Only the edge-cost in the matching algorithm changes

Summary

- Identified inequality in driver income from the job assignments in real-world taxi hailing service
- Introduced notions for fair distribution of income/utilities on ride-hailing platforms
- Proposed mechanisms for matching drivers to passengers to satisfy fairness over time

More details in the paper: https://bit.ly/fair-matching

Other Notions of Fairness

- Note there can be different notions of fairness in different context

Personalized Recommendation in ecommerce platforms

Two Sides: Customers and Sellers/Producers

Recommend k items that maximize customer satisfaction

- Can lead to high inequality in exposure to sellers
- Exposure translates into sales
- Some sellers may starve to find customers and get out of business soon after they join the platform

Why Should We Care for Producers ?

Small businesses depend on the platforms for their livelihood

amazon

Small sellers fear being elbowed out in e-commerce festive sale

Business Today

Flipkart

The Economic Times

Sellers like Cloudtail and WS Retail on Amazon, Flipkart scaling up to grab top slots

Why Should We Care for Producers ?

Small businesses depend on the platforms for their livelihood

amazon

Small sellers fear being elbowed out in e-commerce festive sale

Business Today

Flipkart

The Economic Times

Sellers like Cloudtail and WS Retail on Amazon, Flipkart scaling up to grab top slots

- Legal obligation
  - National e-Commerce Policy, Government of India
- Business requirement
  - To take new producers on board
  - May be equally good as more popular ones
  - More choice for customers with higher competition

Two-Sided Fairness in Recommendation

- Fairness for Producers
  - Ensure a minimum exposure guarantee for every producer
  - Comparable to the fairness of Universal Basic Income
- Fairness for Customers
  - Resultant utility loss should be fairly distributed among customers
  - Products are allocated ensuring envy-freeness

Two-Sided Fairness in Recommendation

- Fairness for Producers
  - Ensure a minimum exposure guarantee for every producer
  - Comparable to the fairness of Universal Basic Income

- Fairness for Customers
  - Resultant utility loss should be fairly distributed among customers
  - Products are allocated ensuring envy-freeness

FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms

Gourab K Patro, Arpita Biswas, Niloy Ganguly, Krishna P. Gummadi and Abhijnan Chakraborty

WWW 2020

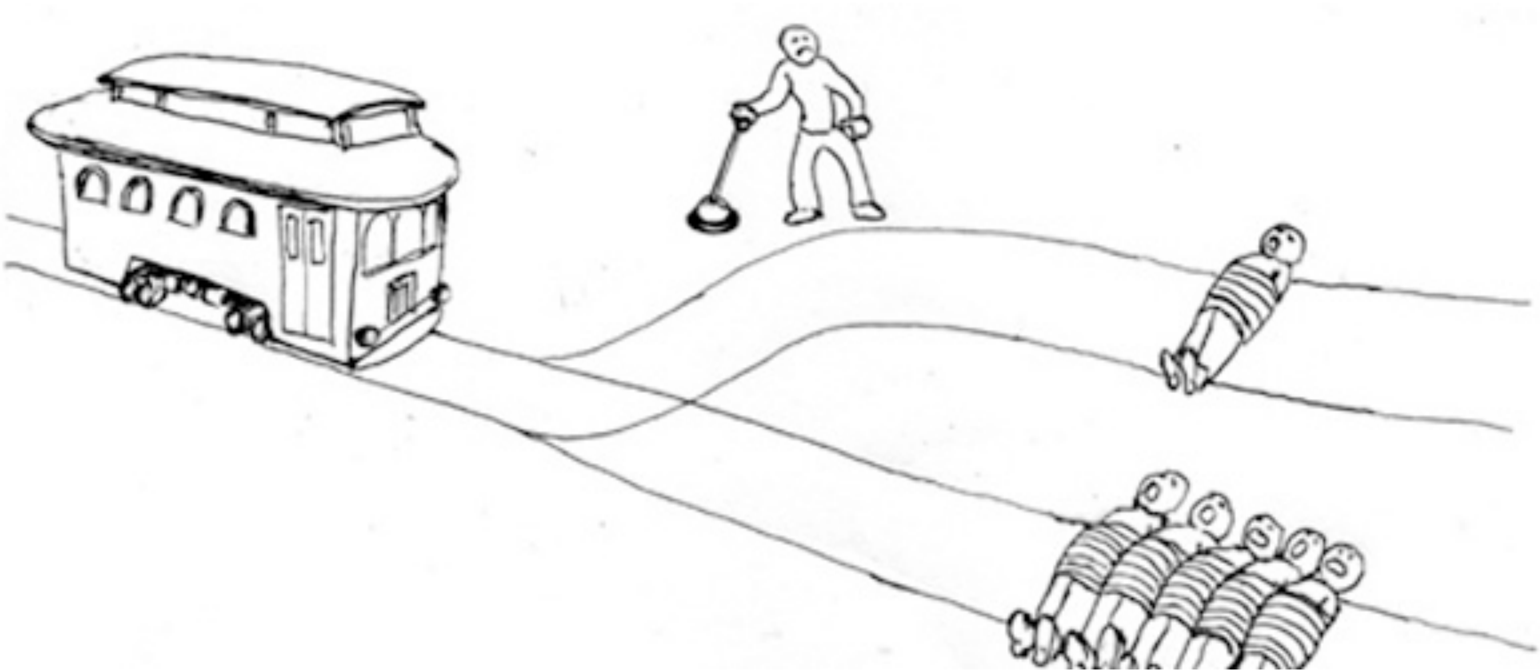
Algorithm Designer’s Dilemma

How should algorithm designers select the right fairness notion?

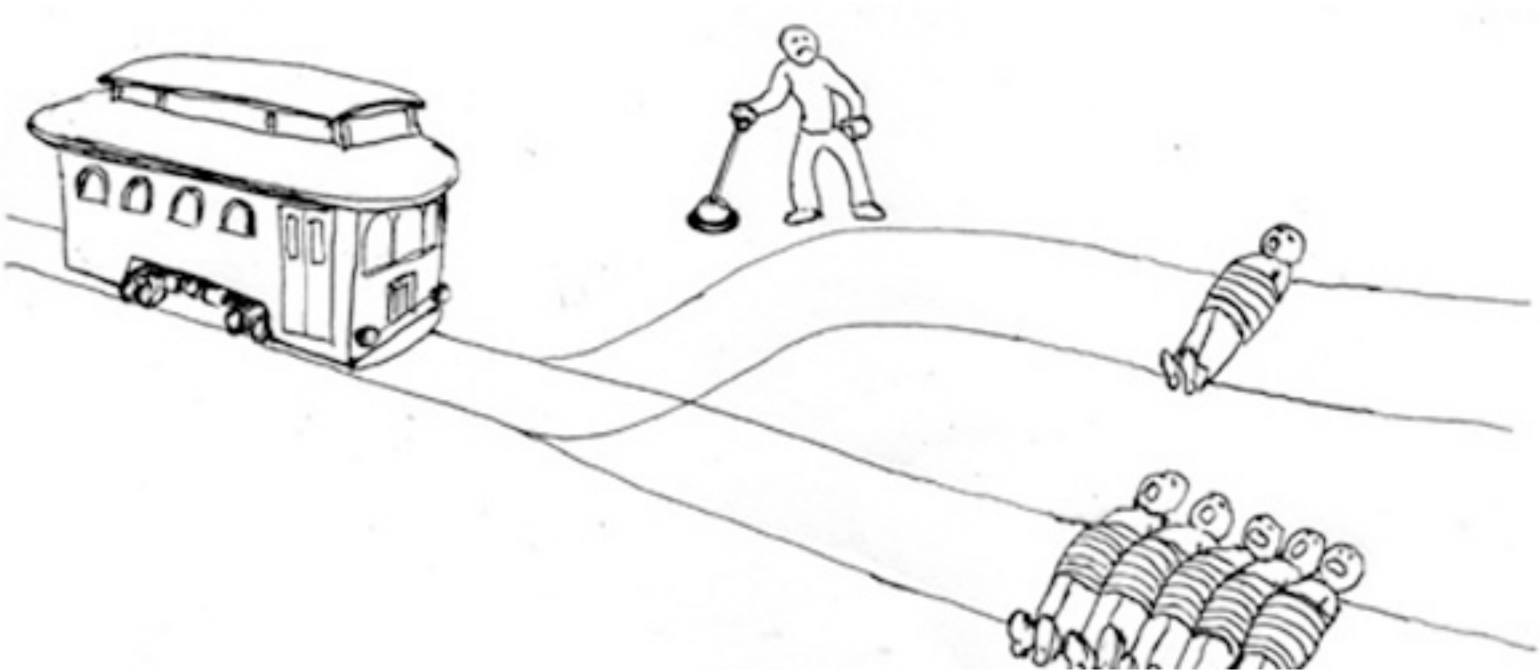
Search for the Right Fairness Notion

- Is not the job of algorithm designers
- Often dictated by (evolving) societal cultures and legal norms
- Our job is to operationalize given a normative criteria
  - Make them formally measurable
  - Design algorithms satisfying the criteria
  - Build efficient systems deployable in practice

Trolley Problem

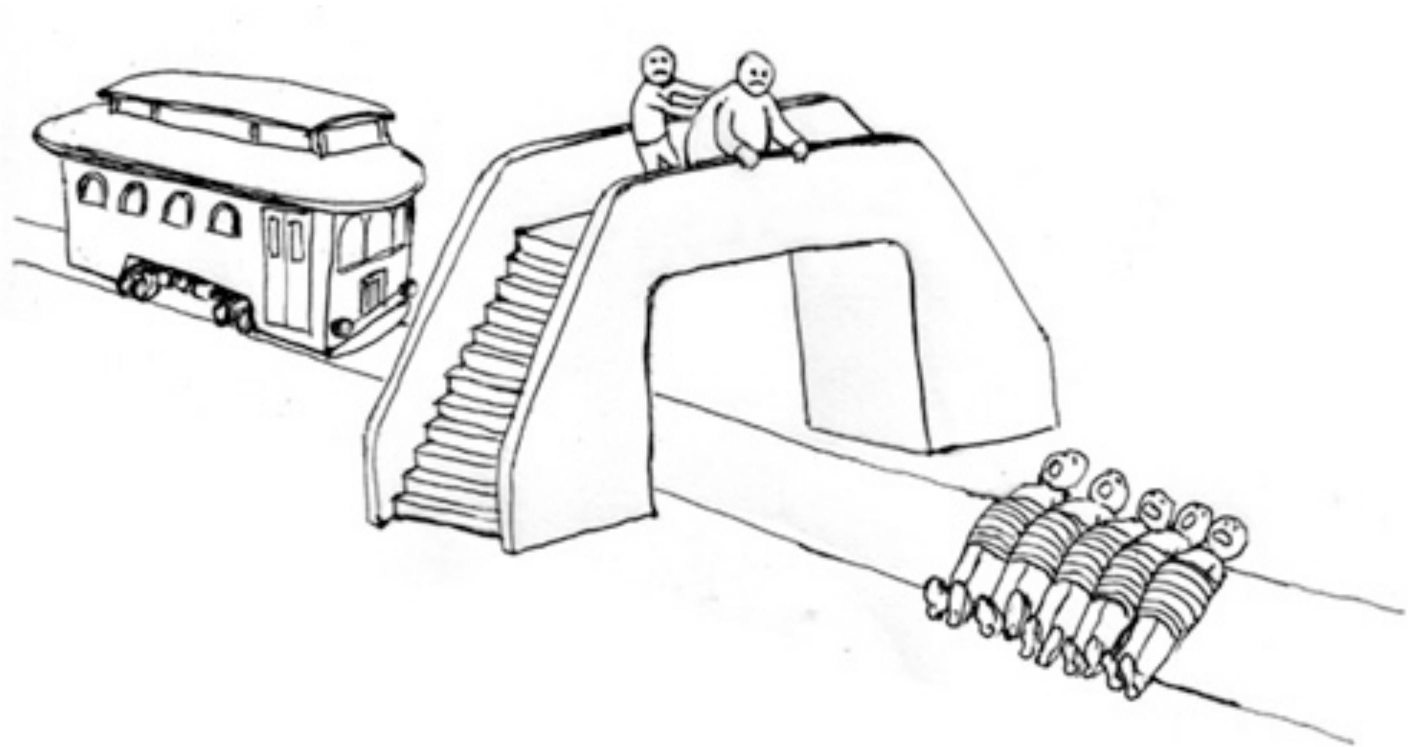


Trolley Problem



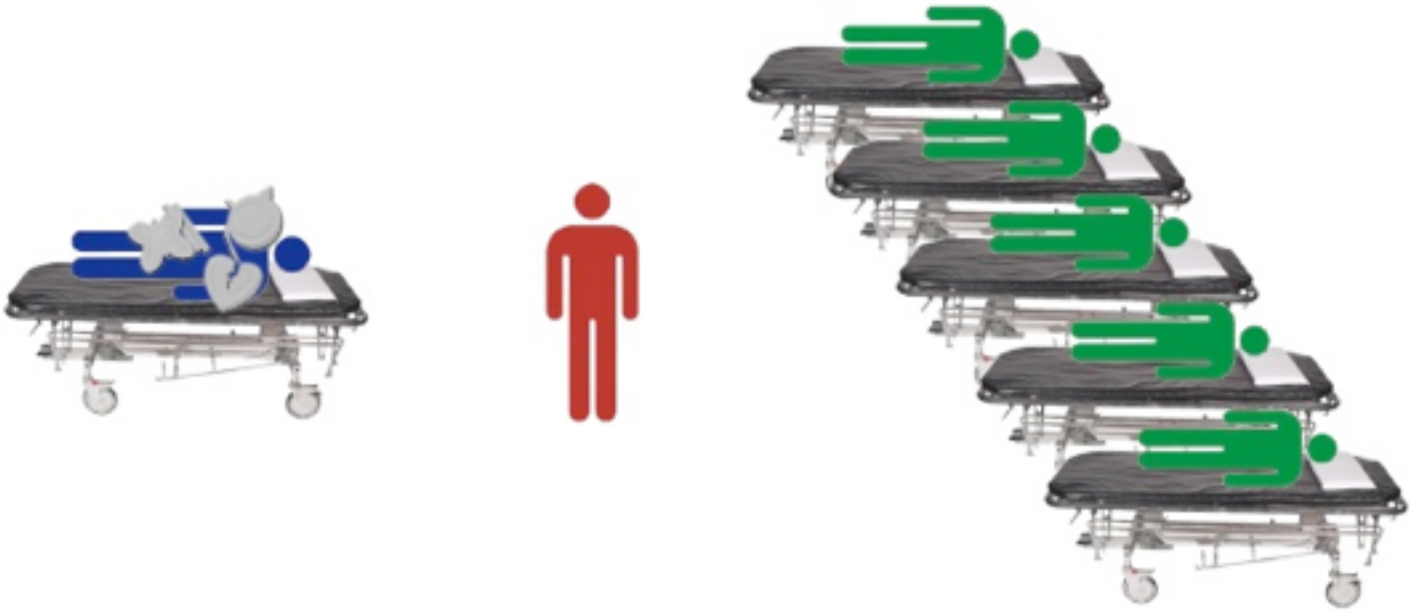
How many of you would pull the lever?

Trolley Problem



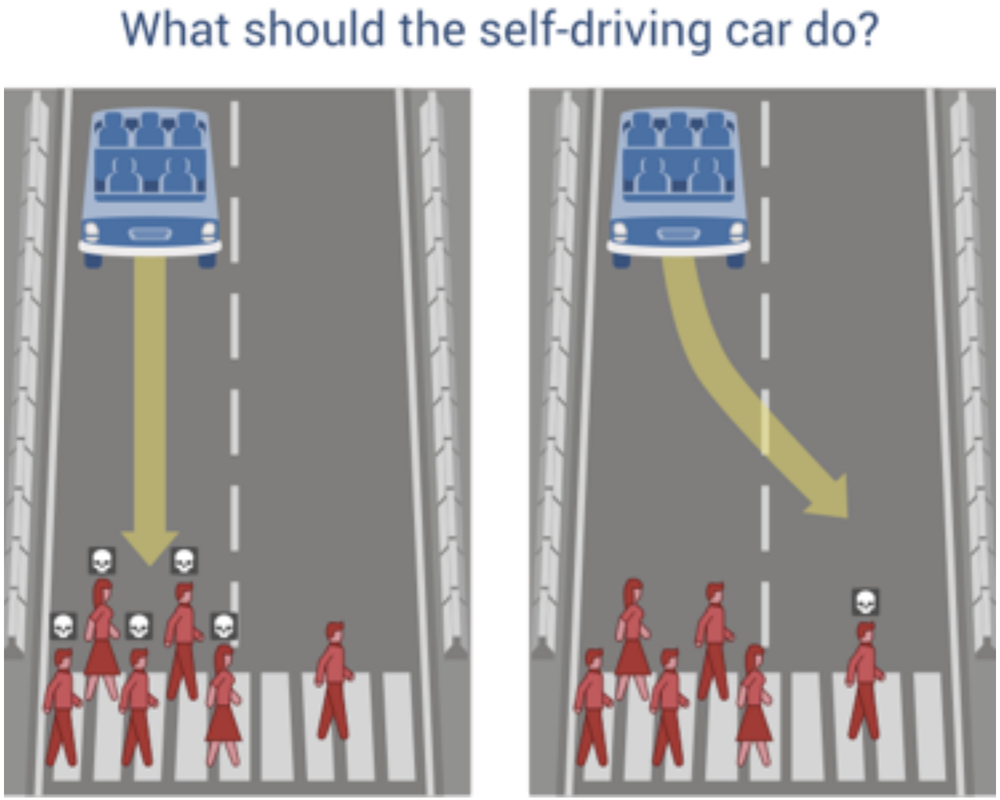
How many of you would push the fat man?

Limits of Utilitarianism



How many of you would transplant organs?

Moral Machine



Should a self-driving car prioritize

- humans over pets?
- passengers over pedestrians?
- more lives over fewer?
- women over men?
- young over old?
- fit over sickly?
- higher social status over lower?
- law-abiders over law-benders?

Should the car swerve (take action) or stay on course (inaction)?

Moral Machine

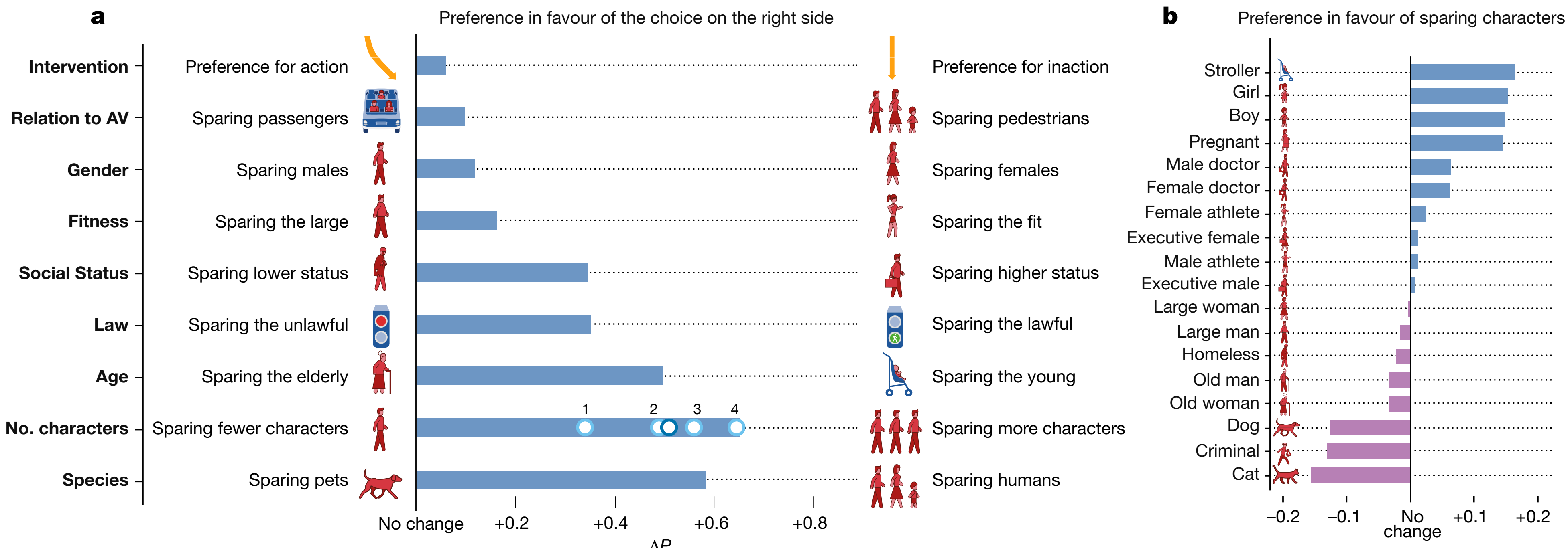


Fig. 2 | Global preferences. a, AMCE for each preference. In each row, ΔP is the difference between the probability of sparing characters possessing the attribute on the right, and the probability of sparing characters possessing the attribute on the left, aggregated over all other attributes. For example, for the attribute age, the probability of sparing young characters is 0.49 (s.e. = 0.0008) greater than the probability of sparing older characters. The 95% confidence intervals of the means are omitted owing to their insignificant width, given the sample size ($n = 35.2$ million). For the number of characters (No. characters), effect sizes are shown for each number of additional characters (1 to 4; $n_1 = 1.52$ million, $n_2 = 1.52$ million, $n_3 = 1.52$ million, $n_4 = 1.53$ million); the effect size for two additional characters overlaps with the mean effect of the attribute. AV, autonomous vehicle. **b**, Relative advantage or penalty for each character, compared to an adult man or woman. For each character, ΔP is the difference the between the probability of sparing this character (when presented alone) and the probability of sparing one adult man or woman ($n = 1$ million). For example, the probability of sparing a girl is 0.15 (s.e. = 0.003) higher than the probability of sparing an adult man or woman.

Edmond Awad et al., *The Moral Machine Experiment*

Takeaway

- Algorithm designers should not aim to find the right fairness notion
- Often dictated by (evolving) societal cultures and legal norms
- Our job is to operationalize given a normative criteria
  - Make them formally measurable
  - Design algorithms satisfying the criteria
  - Build efficient systems deployable in practice

Recommended Reading

- Awad, Edmond, et al. "The moral machine experiment." Nature 563.7729 (2018): 59-64.
- Zafar, Muhammad Bilal, et al. "Fairness constraints: Mechanisms for fair classification." Artificial Intelligence and Statistics. PMLR, 2017.
- Zafar, Muhammad Bilal, et al. "Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment." WWW 2017.
- Mehrabi, Ninareh, et al. "A survey on bias and fairness in machine learning." arXiv:1908.09635
- Sühr, Tom, et al. "Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform." ACM KDD 2019.
- Biega, Asia, et al. "Equity of attention: Amortizing individual fairness in rankings." ACM SIGIR 2018.
- Patro, Gourab, et al. "FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms." WWW 2020.


Thank you