The Translation Tutorial will begin at 4:00 UTC

Data Externalities between users exist if the social and individual value of data sharing diverges. That is for example the case on a social network. If one user shares information about themselves, this gives away some information about others, which might influence their bargaining power. This “leakage” of information, among others, reduces the bargaining power of users over their data. We identify three important features of the data economy that induce data externalities—using mathematical modeling:

- individual contracting
- once-and-for-all data transactions
- substitutable data

Interventions changing either of these characteristics can change the nature of data externalities.

How can the FAccT community help find interventions
that improve Fairness, Accountability and Transparency?

Data Externalities

Dirk Bergemann (Yale), Mihaela Curmei (Berkeley), Yixin Wang (Berkeley), Yuan Cui (Northwestern), Andreas Haupt (MIT)

A Translation Tutorial

The Economist '17

Data is unlike oil. In several respects.

Data is social.

Three Case Studies

45min	Setting you up for the Discussion Groups	1. Models 2. Observations 3. Assumptions
10 min		
25 min	Setting you up for the Discussion Groups	4. Interventions 5. Discussions
10 min	Wrap-Up	6. Beginnings

Incentivized Experiments with Undergraduate Students: The Digital Privacy Paradox

"Consumers say they care about privacy, but at multiple points in the process end up making choices that are inconsistent with their stated preferences."

Athey+ '17

1. Models

“Economics [is] the science of thinking in terms of models and the art of choosing models which are relevant to the contemporary world.”

Keynes ‘23

The Players

- Users A 😊 and B 😊 individually decide whether to join Platform 🤖.
- Platform proposes a data use agreement: Users join or decline.

- We collect information about how you use our Products. For example, we log [...] what posts, videos and other content you view on our Products.

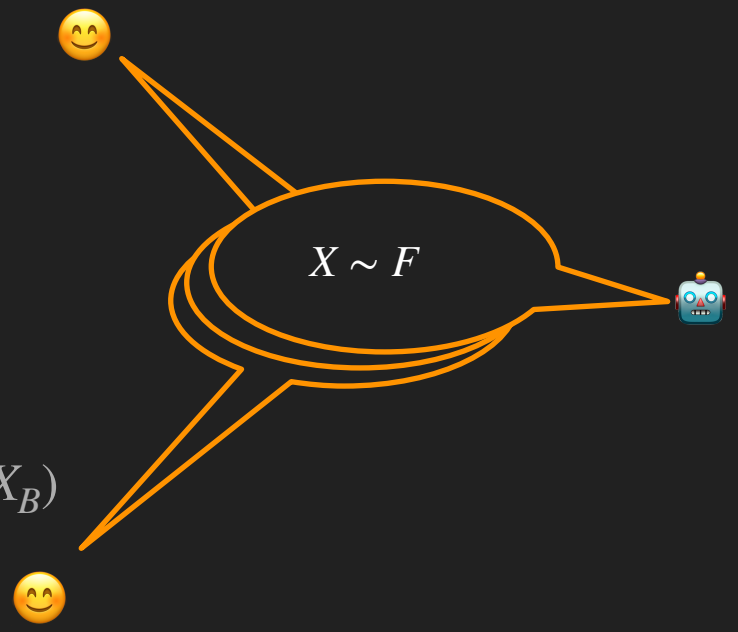
- We use the information we have to [...] personalize features and content and make suggestions for you on and off our Products.

- We store data until it is no longer necessary to provide [...] Facebook Products, or until your account is deleted - whichever comes first.

Facebook, '21

Modelling Information

- We want to model data sharing
- Need: A Model of Information
- Common assumption in Economics:
  - A piece of Information is a realizations of a random quantity
  - All 😊 and 🤖 agree on a prior/the odds
- A sees the realization of X_A
- B sees the realization of X_B
- A, B with Platform on the distribution of (X_A, X_B)



(Compare Aumann '76)

Who decides When

- Users and Platform are strategic
- Agents act in a way that they alone could not improve for themselves
  - Platform makes a proposal
  - Users join or not, doing their selfish best given 🤖's agreement, and who of 😊 joins

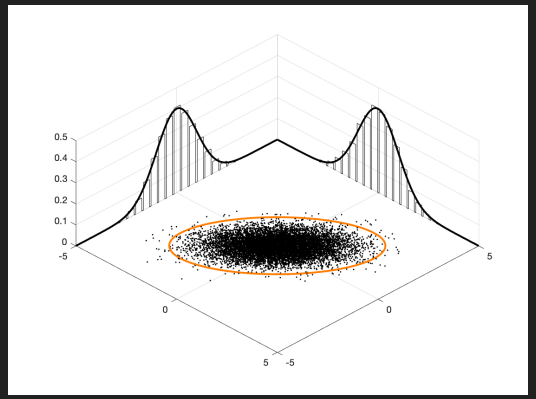
🤖 Here is a user agreement:
reveal your X, get b

😊 As B joins,
I (i.e. A) join

😊 As A join,
I (i.e. B) join

Model

- Actors: Users 😊A, 😊B, Platform 🤖
- All agree that $X_A, X_B \sim N(0,1)$,
  not necessarily independent
- Platform moves first,
  offers service/payment b_A, b_B to each user
- If A and/or B join, 🤖 observes X_A and/or X_B

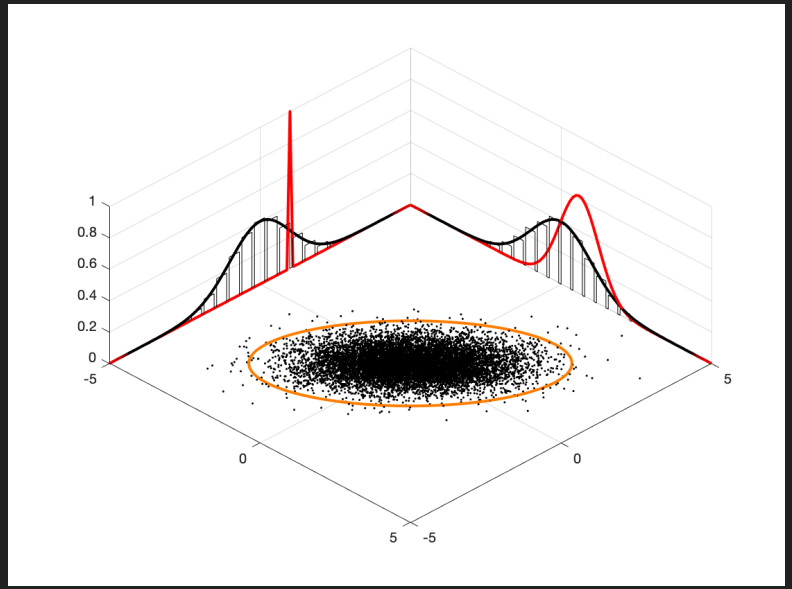


Objectives

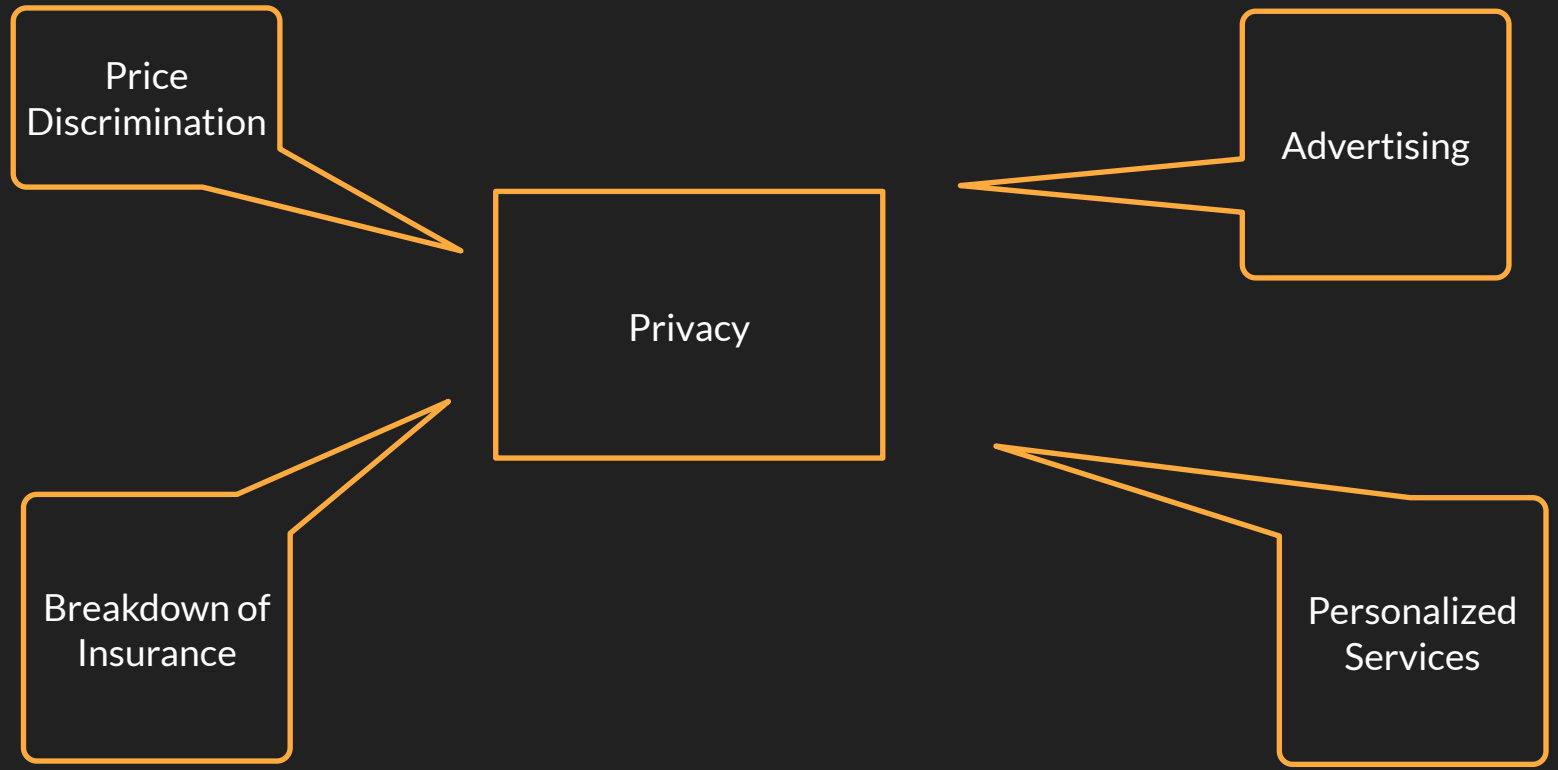
- In Economic Models: Actors are optimizers
- Objectives involve estimation loss
  - $\mathrm{Acc} = \{A, B\}, \{A\}, \{B\}, \emptyset$: joining users
  - $L_A(\mathrm{Acc})$: The smallest loss of an estimate of X_A when users in Acc join
- Platform's objective:

$$\min L_A(\mathrm{Acc}) + L_B(\mathrm{Acc}) + b_A + b_B$$
$$= \max - L_A(\mathrm{Acc}) - L_B(\mathrm{Acc}) - b_A - b_B$$

- Users' objective: Maximize weighted loss
- $\max v_A L_A(\mathrm{Acc}) + b_A$ resp. $\max v_B L_B(\mathrm{Acc}) + b_B$
- Users and Platform have a conflicting interests



Where to look for Values of Privacy



(Compare Bergemann+ ’20)

X_A, X_B: Information
v_A, v_B: Privacy Awareness
L_A, L_B: estimation loss
b_A, b_B: Service level

A Data Externality

- Recall:
  Platform, users A and B agree on a distribution
  $X_A, X_B \sim N(0,1), \mathrm{cov}(X_A, X_B) \approx 1$
- $v_A = 1/2$ (not very privacy-aware) $v_B > 1$ (privacy-aware)
- For any $\mathrm{cov}(X_A, X_B)$, A will join for $b_A \geq 1/2$

A does not join:
$v_A L_A \leq 1/2$

A joins:
$v_A L_A + 1/2 \geq 1/2$

Acemoglu+ '21, (AEJ: Micro, forthcoming)

A Data Externality

$X_A, X_B \sim N(0,1)$
$\mathrm{cov}(X_A, X_B) \approx 1$
$v_A = 1/2$
$v_B \gg 0$ (but B!)

- If A joins, B's data is almost fully known to 🤖
- Hence, B will join for $b_B \approx 0$ if A joins
- But if B joins ($b_B \approx 0$), A's data is almost fully known to 🤖
- Hence, A will join for $b_A \approx 0$
- Platform can offer $b_A, b_B \approx 0$
- Each user's sharing decision has a negative Data Externality on the other user

Acemoglu+ '21, (AEJ: Micro, forthcoming)

2. Observations

Summary of Observations

Data Sharing is Excessive

With respect to what?

Reimbursement is Low

With respect to what?

Glossary

X_A, X_B: Information

v_A, v_B: Privacy Awareness

L_A, L_B: estimation loss

b_A, b_B: Service level

Data Sharing is Excessive

- Benchmark: Maximize the sum of objectives

$$\frac{\begin{array}{rr} -\quad L_A(\mathrm{Acc}) - & L_B(\mathrm{Acc}) - b_A - b_B \\ +v_A L_A(\mathrm{Acc}) + & v_B L_B(\mathrm{Acc}) + b_A + b_B \end{array}}{(\nu_A - 1)L_A(\mathrm{Acc}) + (\nu_B - 1)L_B(\mathrm{Acc})}$$

- In Benchmark A should join depending on the correlation with B and vice versa
- But! Lemma: All users with $\nu_u - 1 < 0$ share their data.
- Theorem: Data sharing is excessive.

Acemoglu+ '21, (AEJ: Micro, forthcoming)

Reimbursements are Low

- Benchmark: If no-one else joins, i.e. B does not join

  In Benchmark, A would join if offered at least

  $b_A = v_a(L_A(\emptyset) - L_A(\{A\}))$

- If B is joining, A joins for $b_A \overset{\geq}{=} v_A(L_A(\{B\}) - L_A(\{A, B\}))$

- Corollary: The reimbursements are depressed

A does not join:

$v_A L_A(\emptyset)$

A joins:

$$v_A L_A(A) + b_A$$
$$\geq v_A L_A(\{A\}) + v_A(L_A(\emptyset) - v_A L_A(\{A\}))$$
$$= v_A L_A(\emptyset)$$

Acemoglu+ '21, (AEJ: Micro, forthcoming)

3. Assumptions

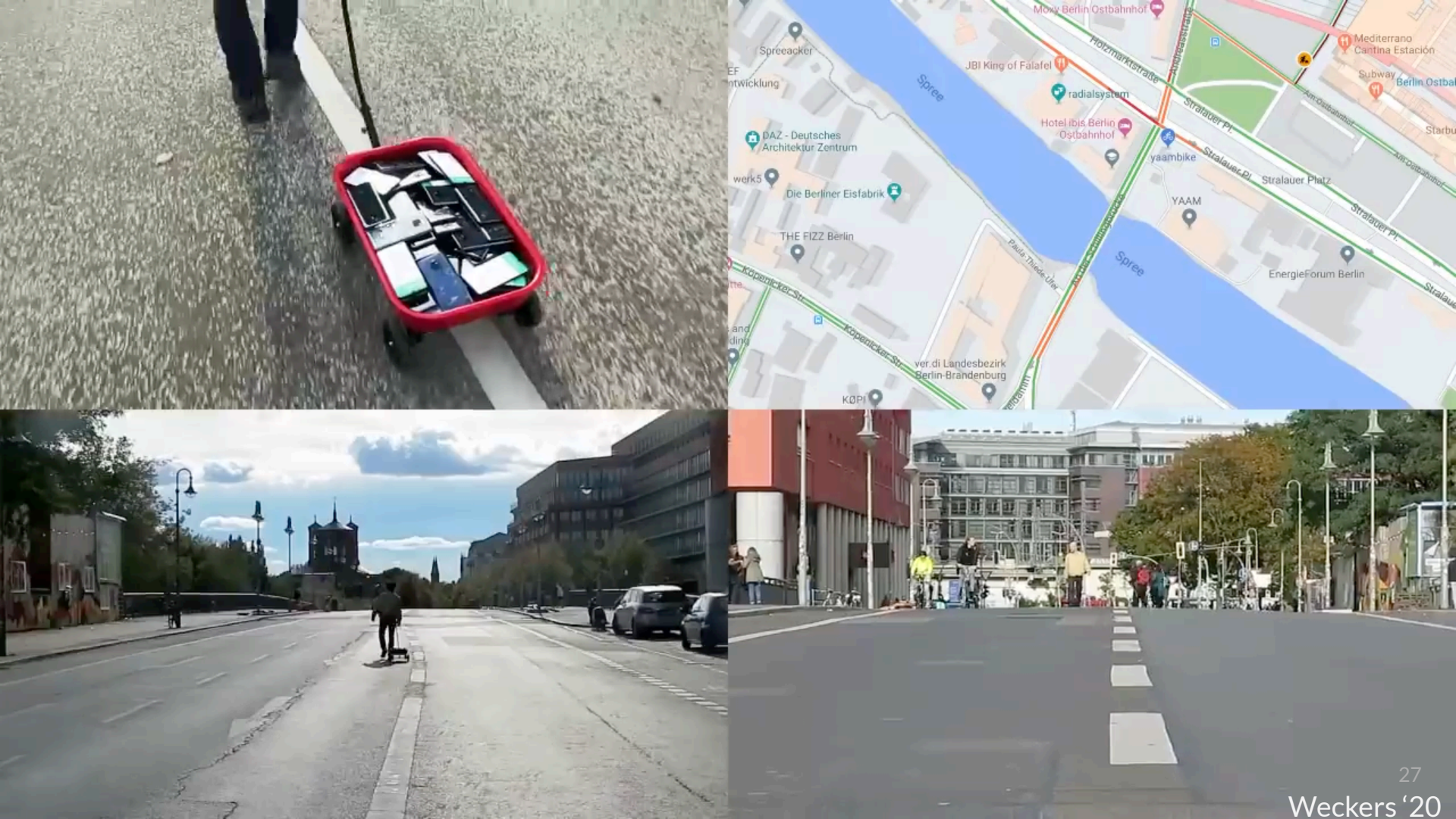
“An assumption is critical if its modification in an arguably more realistic direction would produce a substantive difference in the conclusion produced by the model.”

Rodrik ‘15

The Critical Assumptions

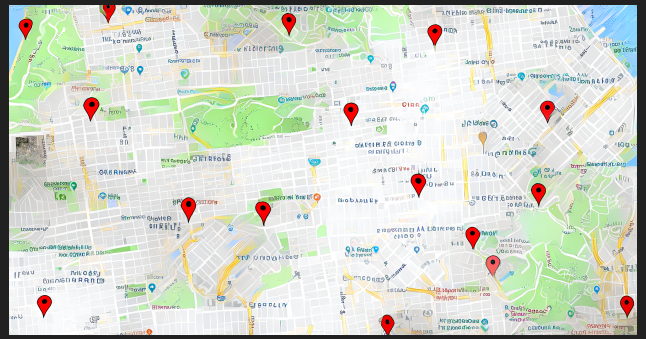
- Data is Substitutable
- Contracts are signed once
- Users Contract Individually

Data might not be substitutable


Spreeacker
JBI King of Falafel
radialsystem
Hotel ibis Berlin Ostbahnhof
DAZ - Deutsches Architektur Zentrum
werk5
Die Berliner Eisfabrik
THE FIZZ Berlin
yaambike
YAAM
Stralauer Platz
EnergieForum Berlin
ver.di Landesbezirk Berlin-Brandenburg
KØPI
Spree
Spree
Mediterrano Cantina Estación
Subway
Holzmarktstraße
Stralauer Pl.
Köpenicker Str.
Paula-Thiede-Ufer

Weckers ‘20

Data is Substitutable



- Platform and users care about state of the world S
- X_A substitutable: X_B alone predicts S
  - $S = (X_A, X_B)/S$ derived using DP algorithms
- X_A complementary: Without X_A, X_B uninformative
- Theorem: If (X_A, X_B) is substitutable, data sharing is excessive, prices are depressed
- Theorem: If (X_A, X_B) is complementary, data sharing is low, prices are high

We know this!

That's new! Benchmarks?

One-Off Contracting

- The model does not capture such dynamics Agents give up their data once-and-for-all, no repeated interaction
- If users can delete data, threat to leave is higher when platform changes policies
- User's bargaining power might increase

🤖 • Here is a user agreement, reveal X you get b

😊 As B joins, I join

😊 As A joins, I join

🤖 • Facebook and WhatsApp now merge databases

😊 As B leaves, I leave

😊 As A leaves, I leave

(Compare Grossman+ '86)

Individual Contracting

- What if users can negotiate together?
- Users and platform agree that $X_A, X_B \sim N(0,1), \mathrm{cov}(X_A, X_B) \approx 1$
- A is mildly privacy aware: $\nu_A = 1/2$, B a lot: $\nu_B > 1$
- Assume that user B can give user A a "reimbursement" of at least $1/2$
- Then no data would be shared, which matches our benchmark

🤖 • Here is a user agreement, reveal X you get b

🤗 Let's see how we split this reimbursement

Explaining the Digital Privacy Paradox

"Consumers say they care about privacy, but at multiple points in the process end up making choices that are inconsistent with their stated preferences."

Athey+ '17

4. Interventions

Discussion Group 1: Complementary and Substitutable Data

- Excessive sharing if data is substitutable
- Little sharing if data is complementary
- What are environments where data is complementary or substitutable?
- Which interventions make substitutable data complementary?

Ishihashi ’21, (compare Hörner+ JET ’13, Shamir ’79)

Discussion Group 2: Making Algorithms Forget

- If platforms cannot keep insights gained from users when they leave, users have more bargaining power
- One tool to give in the hands of users: Effective Algorithmic Forgetting
- Which algorithms to use for algorithmic forgetting?
- How do privacy and data externalities interact?

🤖

- Facebook and WhatsApp now merge databases

😊 As B leaves, I leave; Please delete my data

😊 As A leaves, I leave; Please delete my data

(Compare Ishihashi '20)

Discussion Group 3: Enabling Collective Bargaining over Data

- Collectively, escaping from a data externality is possible
- Writing the rules of decision making is hard
- How to select data processors to share data with?
- How to internally make decisions on how to split reimbursements?

- Here is a user agreement, reveal X you get b

Let's see how we split this reimbursement

(Compare Pentland+ '20)

5. Discussions

Discussion Group Facilitators

Discussion Group 1: Making Algorithms Forget

Mihaela Curmei

Discussion Group 2: Complementary and Substitutable Data

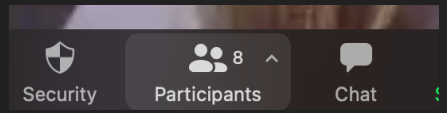
Charles Cui

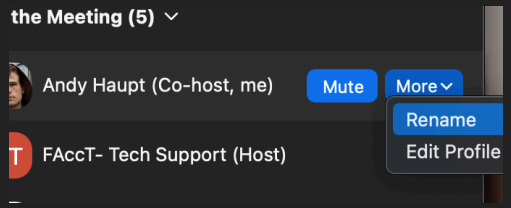
Discussion Group 3: Collective Bargaining over Data

Dirk Bergemann, Yixin Wang

f.l.t.r: Mihaela Curmei, Yuan Cui, Dirk Bergemann, Yixin Wang

Breakout Group Logistics





1. Before the break: Put in the beginning of your name XYZ, where X, Y, Z = 1, 2, 3
2. X is your first preferences for a breakout room, Y your second, Z your third
   a. E.g. "312" means: The person would like most to discuss about Collective Bargaining; the second most preferred would be Making Algorithms Forget; the third most preferred would be to discuss about the Sign of Data Externalities
3. We check whether groups are balanced; if so, you can self-assign.

1. Making Algorithms Forget

2. Complementary and Substitutable Data

3. Collective Bargaining over Data

The Discussion Groups will begin at 4:55 UTC

Data Externalities are differences in the social and individual value of data sharing.
Data Externalities reduce the bargaining power of data subjects over their data if:

- individual contracting and
- once-and-for-all data transactions shape the interaction and
- data from different users are substitutable.

Interventions changing either of these characteristics can change the nature of data externalities.

How can the FAccT community contribute?

Discussion Groups

- We will put you into discussion groups
- PM Charles Cui if you need to be put into a room

6. Beginnings

Reports from the Discussion Groups

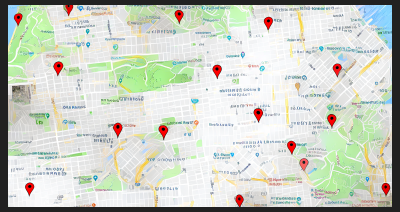
Making Algorithms Forget
Mihaela Curmei

😊 As B leaves, I leave; Please delete my data

Collective Bargaining
Yixin Wang

Let's see how we split this reimbursement

Data is Substitutable
Charles Cui



Thank you, we hope to hear from you!

Facilitators:
Dirk Bergemann dirk.bergemann@yale.edu
Charles Cui charlescui@u.northwestern.edu
Mihaela Curmei mcurmei@berkeley.edu
Andreas Haupt haupt@mit.edu
Yixin Wang yixin.wang.sh@gmail.com
(Thanks to: Rediet Abebe riediettes@gmail.com)

Report on 3/10/2021 on https://www.md4sg.com/workshop/faact21/faact21tutorial