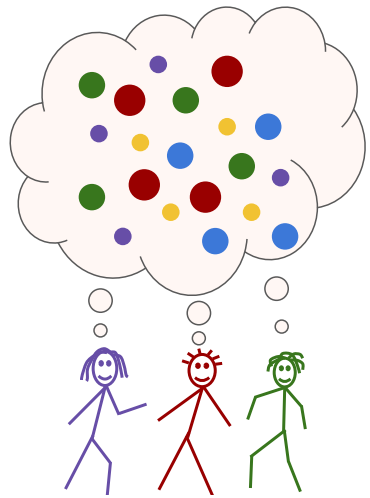
On Valuing and Procuring Personal Data



Bo Waggoner
Microsoft Research, NYC
→ U. Colorado, Boulder

Berkeley, CA
May 2019

This talk

Part 1: Background (manifesto)

- What does it mean to **own** personal data?
- How might *people* and *firms* **value** personal data?

Part 2: Research on valuing and procuring data **efficiently**

- Active-learning based (with Jake Abernethy, Yiling Chen, C.J. Ho)
- Prediction-market based (with Raf Frongillo; Jake Abernethy; Justin Harris)

Part 3: Discussion (screed)

1 What does it mean to **own, buy, sell** personal data?

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership $\approx$ *possession*

generally: ownership = power of (exclusive) control

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership $\approx$ *possession*

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

2 How do **people** value their personal data?

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

2 How do **people** value their personal data?

(1) possible harms incurred due to revelation (often difficult to assess)

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

2 How do **people** value their personal data?

(1) possible harms incurred due to revelation (often difficult to assess)

... enter differential privacy! (note: centralized model may not be credible)

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

2 How do **people** value their personal data?

(1) possible harms incurred due to revelation (often difficult to assess)

... enter differential privacy! (note: centralized model may not be credible)

(2) possible profit available from selling/renting rights

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

2 How do **people** value their personal data?

(1) possible harms incurred due to revelation (often difficult to assess)

... enter differential privacy! (note: centralized model may not be credible)

(2) possible profit available from selling/renting rights

modeled as willingness to sell accesss

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

2 How do **people** value their personal data?

(1) possible harms incurred due to revelation (often difficult to assess)

... enter differential privacy! (note: centralized model may not be credible)

(2) possible profit available from selling/renting rights

modeled as willingness to sell accesss

3 How does a **firm** value personal data?

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

2 How do **people** value their personal data?

(1) possible harms incurred due to revelation (often difficult to assess)

... enter differential privacy! (note: centralized model may not be credible)

(2) possible profit available from selling/renting rights

modeled as willingness to sell accesss

3 How does a **firm** value personal data?

in general: information derives value from improvement to decisionmaking

1 What does it mean to **own, buy, sell** personal data?

for physical objects: ownership ≈ possession

generally: ownership = power of (exclusive) control

for information: control is governed by **legal frameworks**, *e.g.* **copyright**

owning data = control over it; purchasing data = purchasing rights to use it

example: company rents the right to use data for limited purposes and durations

2 How do **people** value their personal data?

(1) possible harms incurred due to revelation (often difficult to assess)

... enter differential privacy! (note: centralized model may not be credible)

(2) possible profit available from selling/renting rights

modeled as willingness to sell accesss

3 How does a **firm** value personal data?

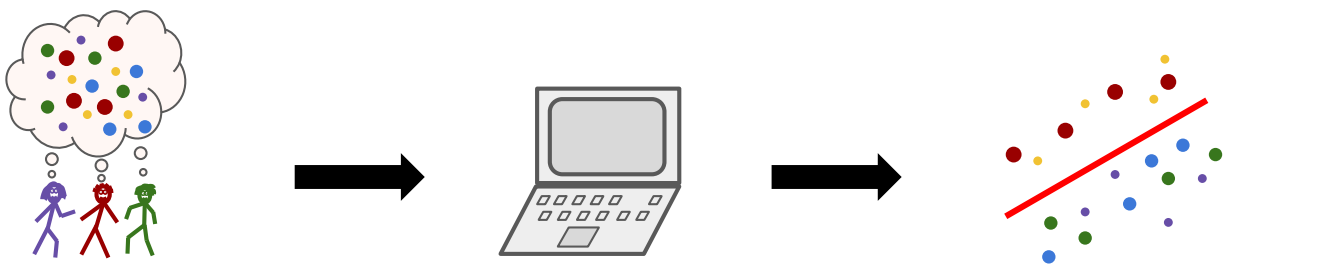
in general: information derives value from improvement to decisionmaking

proxy: loss function measures performance, data improves loss

Research #1: An Active-Learning Approach

Low-Cost Learning via Active Data Procurement.
Abernethy, Chen, Ho, **Waggoner**. EC 2015.

Problem: How to model and achieve procurement of personal data?



agents with private data and costs

algorithm with budget B

hypothesis h

Related approaches: Purchase data to estimate population statistics, especially

- Roth, Schoenebeck. EC 2012.
- Chen, Immorlica, Lucier, Syrgkanis, Ziani. EC 2018. *extends RS12*
- Ghosh, Roth. EC 2011. *studies cost for privacy*
- Ligett, Roth. WINE 2012. *ditto*

Related approaches: Purchase data to estimate population statistics, especially

- Roth, Schoenebeck. EC 2012.
- Chen, Immorlica, Lucier, Syrgkanis, Ziani. EC 2018. *extends RS12*
- Ghosh, Roth. EC 2011. *studies cost for privacy*
- Ligett, Roth. WINE 2012. *ditto*

Challenge: data can be **correlated** with **willingness to sell**

Related approaches: Purchase data to estimate population statistics, especially

- Roth, Schoenebeck. EC 2012.
- Chen, Immorlica, Lucier, Syrgkanis, Ziani. EC 2018. *extends RS12*
- Ghosh, Roth. EC 2011. *studies cost for privacy*
- Ligett, Roth. WINE 2012. *ditto*

Challenge: data can be **correlated** with **willingness to sell**

Drawbacks: specialized to statistics; not **data efficient**

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

1. Mechanism has hypothesis h_{t-1}

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

1. Mechanism has hypothesis h_{t-1}
2. Mechanism posts **menu of prices**

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

1. Mechanism has hypothesis h_{t-1}
2. Mechanism posts **menu of prices**
3. Agent t arrives with secret data

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

1. Mechanism has hypothesis h_{t-1}
2. Mechanism posts **menu of prices**
3. Agent t arrives with secret data
4. If agent agrees to sell:

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

1. Mechanism has hypothesis h_{t-1}
2. Mechanism posts **menu of prices**
3. Agent t arrives with secret data
4. If agent agrees to sell:
   - Mechanism receives data, pays menu price

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

1. Mechanism has hypothesis h_{t-1}
2. Mechanism posts **menu of prices**
3. Agent t arrives with secret data
4. If agent agrees to sell:
   - Mechanism receives data, pays menu price
5. Mechanism updates to new hypothesis h_t

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

1. Mechanism has hypothesis h_{t-1}
2. Mechanism posts **menu of prices**
3. Agent t arrives with secret data, **cost** in $[0, 1]$
4. If **cost** $\leq$ **menu**(data), agent agrees to sell:
   - Mechanism receives data, pays menu price
5. Mechanism updates to new hypothesis h_t

(assume: agents cannot fabricate data!)

For $t = 1, \ldots, T$:

1. Mechanism has hypothesis h_{t-1}
2. Mechanism posts **menu of prices**
3. Agent t arrives with secret data, **cost** in $[0, 1]$
4. If **cost** $\leq$ **menu**(data), agent agrees to sell:
   - Mechanism receives data, pays menu price
5. Mechanism updates to new hypothesis h_t

Key idea: base prices on **value of data** to the learning algorithm

Results: regret bounds $T\sqrt{\frac{\gamma}{B}}$ (online setting)
and generalization bounds $\sqrt{\frac{\gamma}{B}}$ (i.i.d. data)

Research #2: A Markets-Based Approach

A Market Framework for Eliciting Private Data.
Waggoner, Frongillo, Abernethy. NeurIPS 2015.

Goal: use a "market" to procure data **privately** and with **good incentives**!

See also:

- A Collaborative Mechanism for Crowdsourcing Prediction Problems (NeurIPS 2011). Abernethy, Frongillo.
- The Possibilities and Limitations of Private Prediction Markets (EC 2016). Cummings, Pennock, Wortman Vaughan.
- An Axiomatic Study of Scoring Rule Markets (ITCS 2018). Frongillo, **Waggoner**.
- Bounded-Loss Private Prediction Markets (NeurIPS 2018). Frongillo, **Waggoner**.

Providing Phase:

For $t = 1, \ldots, T$:

1. Mechanism has current hypothesis h_{t-1}.
2. Agent t arrives, provides data
3. Mechanism updates to hypothesis h_t

Providing Phase:

For $t = 1, \ldots, T$:

1. Mechanism has current hypothesis h_{t-1}.
2. Agent t arrives, provides data
3. Mechanism updates to hypothesis h_t

Payment Phase:

1. Mechanism reveals test dataset D
2. Each agent t receives Loss(h_{t-1}, D) - Loss(h_t, D)

Providing Phase:

For $t = 1, \ldots, T$:

1. Mechanism has current hypothesis h_{t-1}.
2. Agent t arrives, provides data
3. Mechanism updates to hypothesis h_t

Payment Phase:

1. Mechanism reveals test dataset D
2. Each agent t receives Loss(h_{t-1}, D) - Loss(h_t, D)

Key ideas:

- Aligned incentives, bounded budget
- Opt-in for users
- Pay only for **useful** data
- Can add **differential privacy**

Part 3: Discussion

(Apologies in advance)

- Why (perhaps) markets for personal data?

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control
  ethical argument: right to capture some value

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control
  ethical argument: right to capture some value
  economics argument: leads to efficient allocations/uses

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control
  ethical argument: right to capture some value
  economics argument: leads to efficient allocations/uses
  libertarian objection: exerts control over rights of others

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control
  ethical argument: right to capture some value
  economics argument: leads to efficient allocations/uses
  libertarian objection: exerts control over rights of others
  socialist objection: privatizes the commons

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control
  ethical argument: right to capture some value
  economics argument: leads to efficient allocations/uses
  libertarian objection: exerts control over rights of others
  socialist objection: privatizes the commons
  pragmatic objections: implementation, censorship, . . .

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control
  ethical argument: right to capture some value
  economics argument: leads to efficient allocations/uses
  libertarian objection: exerts control over rights of others
  socialist objection: privatizes the commons
  pragmatic objections: implementation, censorship, . . .
- Technical approaches to control; freedom-respecting software

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control
  ethical argument: right to capture some value
  economics argument: leads to efficient allocations/uses
  libertarian objection: exerts control over rights of others
  socialist objection: privatizes the commons
  pragmatic objections: implementation, censorship, . . .
- Technical approaches to control; freedom-respecting software
  principle: software should respect and empower its users . . . not exploit them!

- Why (perhaps) markets for personal data?
  route data efficiently; social welfare; possibly egalitarian
- Data is capital
  despite Arrieta-Ibarra, Goff, Hernández, Lanier, Weyl 2017
  important because: one-time fees may be **exploitative** *(even GDPR)*
- Arguments around "purchasing data" (data rights, renting, contracts)
  ethical argument: right to privacy/control
  ethical argument: right to capture some value
  economics argument: leads to efficient allocations/uses
  libertarian objection: exerts control over rights of others
  socialist objection: privatizes the commons
  pragmatic objections: implementation, censorship, . . .
- Technical approaches to control; freedom-respecting software
  principle: software should respect and empower its users . . . not exploit them!
  in particular: control over what information is revealed about you

Summary

- "Ownership:" data rights
- Value for data: "willingness to sell", loss function proxy
- Research: active-learning style, prediction-markets style
- Why markets? economic role of data
- Technical solutions? libre software

Thank you!