Nonrivalry and
the Economics of Data

Chad Jones and Chris Tonetti

NBER Economics of Digitization
March 22, 2019
Examples of Data

- Google, Facebook
- Amazon
- Tesla, Uber, Waymo
- Medical and genetic data
- Location history
- Speech records
- Physical action data

Canonical example: data as input into machine learning algorithm. E.g. self-driving car.
Data is Nonrival

• Data is infinitely usable
  ○ Contrast with rival goods: coffee, computer, doctor
  ○ Multiple engineers/algorithms can use same data at same
time (within and across firms)

• Key ways that data enters the economy:
  ○ Nonrivalry ⇒ social gain from sharing data
  ○ Privacy
  ○ Firm: competitive advantage (“moat”)

• Social planner and consumers only care about the first two. But
  firms care a lot about the last one ⇒ inefficiency
What policies governing data use maximize welfare?

- European General Data Protection Regulation (GDPR)
  - Privacy vs. social gain from sharing
  - “The protection of natural persons in relation to the processing of personal data is a fundamental right”
  - “The right . . . must be considered in relation to its function in society and be balanced against other fundamental rights. . . ”

- The California Consumer Privacy Act of 2018
  - Allows consumers to opt out of having their data sold
Nonrivalry of Data ⇒ Increasing Returns

- Nonrivalry implies increasing returns to scale: $Y = F(D, X)$
  - Constant returns to rival inputs: $F(D, \lambda X) = \lambda F(D, X)$
  - Increasing returns to data and rival inputs: $F(\lambda D, \lambda X) > \lambda F(D, X)$

- When firms hoard data, a firm learns only from its own consumers
- But when firms share data, all firms learn from all consumers
  - Firms, fearing creative destruction, will not do this
  - But if consumers own the data, they appropriately balance data sharing and privacy
Outline

- Economic environment

- Allocations:
  - Optimal allocation
  - Firms own data
  - Consumers own data
  - Extreme privacy protection: outlaw data sharing

- Theory results and a numerical example
Basic Setup
Overview

• Representative consumer with a love for variety

• Innovation ⇒ endogenous measure of varieties

• Nonrivalry of data ⇒ increasing returns to scale

• How is data produced?
  o Learning by doing: each unit consumed ⇒ 1 unit of data
  o Alternative: separate PF (Tesla vs Google self-driving car)

• Any data equally useful in all firms ⇒ one sector of economy

• Data depreciates fully each period
The Economic Environment

Utility

\[ \int_0^\infty e^{-\rho t} L_t u(c_t, x_{it}, \tilde{x}_{it}) \, dt \]

Flow Utility

\[ u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\kappa}{2} \frac{1}{N_t} \int_0^{N_t} x_{it}^2 \, di - \frac{\tilde{\kappa}}{2} \frac{1}{N_t} \int_0^{N_t} \tilde{x}_{it}^2 \, di \]

Consumption per person

\[ c_t = \left( \int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}} \quad \text{with } \sigma > 1 \]

Data production

\[ J_{it} = c_{it} L_t \]

Variety resource constraint

\[ c_{it} = Y_{it} / L_t \]

Firm production

\[ Y_{it} = D_{it}^\eta L_{it}, \quad \eta \in (0, 1) \]

Data used by firm \( i \)

\[ D_{it} \leq \alpha x_{it} J_{it} + (1 - \alpha) B_t \quad \text{(nonrivalry)} \]

Data of firm \( i \) used by others

\[ D_{sit} \leq \tilde{x}_{it} J_{it} \]

Data bundle

\[ B_t = \left( N_t^{\frac{1}{\epsilon}} \int_0^{N_t} D_{sit}^{\frac{\epsilon-1}{\epsilon}} \, di \right)^{\frac{\epsilon}{\epsilon-1}} = N_t D_{sit} \text{ in eqm} \]

Innovation (new varieties)

\[ \dot{N}_t = \frac{1}{\chi} \cdot L_{et} \]

Labor resource constraint

\[ L_{et} + \int_0^{N_t} L_{it} \, di = L_t \]

Population growth (exogenous)

\[ L_t = L_0 e^{gLt} \]

Creative destruction

\[ \delta(\tilde{x}_{it}) = \frac{\delta_0}{2} \tilde{x}_{it}^2 \text{ (equilibrium)} \]
The Planner Problem (using symmetry of firms)

\[
\max_{\{L_{pt}, x_{it}, \tilde{x}_{it}\}} \int_0^\infty e^{-\tilde{\rho}t} L_0 u(c_t, x_{it}, \tilde{x}_{it}) \, dt, \quad \tilde{\rho} := \rho - g_L
\]

subject to

\[
c_t = Y_t/L_t
\]

\[
Y_t = N_t^{\frac{1}{\sigma - 1}} D_{it}^\eta L_{pt}
\]

\[
D_{it} = \alpha x_{it} Y_{it} + (1 - \alpha) N_t \tilde{x}_{it} Y_{it}
\]

\[
Y_{it} = D_{it}^\eta \cdot \frac{L_{pt}}{N_t}
\]

\[
\dot{N}_t = \frac{1}{\chi} (L_t - L_{pt})
\]

\[
L_t = L_0 e^{g_L t}
\]

- More sharing ⇒ negative utility cost but more consumption
- Balance labor across production and entry/innovation
Scale Effect from Sharing Data

\[ D_{it} = \alpha x_{it} J_{it} + (1 - \alpha) \left( N_t^{-\frac{1}{\epsilon}} \int_0^{N_t} (\tilde{x}_{it} J_{it})^{\frac{\epsilon - 1}{\epsilon}} \, di \right)^{\frac{\epsilon}{\epsilon - 1}} \]

\[ D_{it} = \alpha x_{it} Y_{it} + (1 - \alpha) N_t \tilde{x}_{it} Y_{it} \]

\[ = [\alpha x_{it} + (1 - \alpha) \tilde{x}_{it} N_t] Y_{it} \]

- No sharing versus sharing:
  - No sharing: Only the \( \alpha x_t \) term = no scale effect
  - Sharing: The \((1 - \alpha) \tilde{x}_t N_t \) term = extra scale effect

Source of Scale Effect: \( N_t \) scales with \( L_t \)

- Plugging into production function:

\[ Y_{it} = ([\alpha x_t + (1 - \alpha) \tilde{x}_t N_t]^{\eta} L_{it})^{\frac{1}{1-\eta}} \]
Firms Own Data
Firms Own Data: Consumer Problem

- Firms own data and choose one data policy \((x_{it}, \tilde{x}_{it})\) applied to all consumers

- Consumers just choose consumption:

\[
U_0 = \max_{\{c_{it}\}} \int_0^\infty e^{-\tilde{\rho} t} L_0 u(c_t, x_{it}, \tilde{x}_{it}) dt
\]

subject to

\[
c_t = \left( \int_0^{N_t} \frac{\sigma - 1}{\sigma} c_{it}^{\frac{1}{\sigma}} di \right)^{\frac{\sigma}{\sigma - 1}}
\]

\[
\dot{a}_t = (r_t - g_L) a_t + w_t - \int_0^{N_t} p_{it} c_{it} di
\]
Firms own Data: Data Decisions

- Firms buy $D_{bit}$ data from intermediary at given price $p_b$
- Firms sell $D_{sit}$ data to intermediary at chosen price $p_{si}$
  - Perfect competition inconsistent with nonrival data!
  - Monopolistically competitive with own data
  - See the intermediary’s downward-sloping demand curve and set price

- How much data to use / sell?
  - $x_{it}$: Use all of own data $\Rightarrow x_{it} = 1$
  - $\tilde{x}_{it}$: Trade off = selling data versus creative destruction
    $\delta(\tilde{x}_{it}) = \text{Poisson rate transferring ownership of variety}$
Firms own the Data: Incumbent Firm Problem

- Monopolistically competitive firm takes demand for variety as given (from FOC of consumer problem): \( p_{it} = \left( \frac{c_{t}}{c_{it}} \right)^{\frac{1}{\sigma}} = \left( \frac{Y_{t}}{Y_{it}} \right)^{\frac{1}{\sigma}} \)

\[
r_{t}V_{it} = \max_{L_{it},D_{bit},x_{it},\tilde{x}_{it}} \left( \frac{Y_{t}}{Y_{it}} \right)^{\frac{1}{\sigma}} Y_{it} - w_{t}L_{it} - p_{bt}D_{bit} + p_{st}\tilde{x}_{it}Y_{it} + \dot{V}_{it} - \delta(\tilde{x}_{it})V_{it}
\]

s.t. \( Y_{it} = D_{it}^{\eta}L_{it} \)

\( D_{it} = \alpha x_{it}Y_{it} + (1 - \alpha)D_{bit} \)

\( x_{it} \in [0, 1], \tilde{x}_{it} \in [0, 1] \)

\( p_{sit} = \lambda_{DIN_{t}}^{-\frac{1}{\epsilon}} \left( \frac{B_{t}}{\tilde{x}_{it}Y_{it}} \right)^{\frac{1}{\epsilon}} \)

- Data Intermediary \((p_{bt}, p_{st}, D_{bit})\) and Free Entry complete eqm.
Firms Own Data: A “No Trade” Law

• What if the government, in an attempt to protect consumers privacy, makes data sharing illegal?

• Government chooses
  
  ○ $x_{it} \in (0, 1]$
  
  ○ $\tilde{x}_{it} = 0$

• We call this the “Outlaw Sharing” allocation
Consumers Own Data
Consumers own Data: Consumer Problem

- Consumers own data, so now choose how much to share $(x_{it}, \tilde{x}_{it})$:

\[
U_0 = \max_{\{c_{it}, x_{it}, \tilde{x}_{it}\}} \int_0^\infty e^{-\tilde{\rho} t} L_0 u(c_t, x_{it}, \tilde{x}_{it}) dt
\]

s.t. \[
c_t = \left( \int_0^{N_t} \frac{c_{it}^{\sigma-1}}{\sigma} \right)^{\frac{\sigma}{\sigma-1}}
\]

\[
\dot{a}_t = (r_t - g_L)a_t + \omega_t - \int_0^{N_t} p_{it}c_{it} di + \int_0^{N_t} x_{it}p_{st}^{a}c_{it} di + \int_0^{N_t} \tilde{x}_{it}p_{st}^{b}c_{it} di
\]

- Firm problem similar to before, but now takes $x, \tilde{x}$ as given, can’t sell data, and has to buy “own” data
Key Forces: Consumers vs. Firms vs. Outlaw Sharing

- **Firms**
  - use all data on own variety, ignoring consumer privacy
  - restrict data sharing because of creative destruction

- **Consumers**
  - respect their own privacy concerns
  - sell data broadly, ignoring creative destruction

- **Outlaw sharing**
  - maximizes privacy gains
  - missing scale effect reduces consumption
Results: Comparing Allocations

1. Planner Problem
2. Firms Own Data
3. Outlaw Data Sharing
4. Consumers Own Data
Key Allocations: $alloc \in \{sp, f, c, ns\}$

- **Firm size:** $L_{i}^{alloc} = L_{pt}/N_{t} = \nu_{alloc}$

  
  $\nu_{sp} := \chi \rho \cdot \frac{\sigma - 1}{1 - \eta}$
  
  $\nu_{os} := \chi \rho \cdot \frac{\sigma - 1}{1 - \sigma \eta}$
  
  $\nu_{c} := \chi g_{L} \cdot \frac{\rho + \delta(\tilde{x}_{c})}{g_{L} + \delta(\tilde{x}_{c})} \cdot \frac{\sigma - 1}{1 - \sigma \eta}$
  
  $\nu_{f} := \chi g_{L} \cdot \frac{\rho + \delta(\tilde{x}_{f})}{g_{L} + \delta(\tilde{x}_{f})} \cdot \frac{\sigma - 1}{1 - \sigma \eta^{\frac{\epsilon - 1}{\epsilon}}}$

- **Number of firms:** $N_{t}^{alloc} = \psi_{alloc} L_{t}$

  
  $\psi_{alloc} := \frac{1}{\chi g_{L} + \nu_{alloc}}$
## Data Sharing

### Own Firm Data

\[
x_{sp} = \frac{\alpha}{1-\alpha} \frac{\bar{\kappa}}{\kappa} \left( \frac{1}{\bar{\kappa}} \cdot \frac{\eta}{1-\eta} \right)^{1/2}
\]

\[
x_f = 1
\]

\[
x_{os} \in (0, 1]
\]

\[
x_c = \frac{\alpha}{1-\alpha} \frac{\bar{\kappa}}{\kappa} \left( \frac{1}{\bar{\kappa}} \cdot \frac{\eta}{1-\eta} \cdot \frac{\sigma-1}{\sigma} \right)^{1/2}
\]

### Sharing with Other Firms

\[
\tilde{x}_{sp} = \left( \frac{1}{\bar{\kappa}} \cdot \frac{\eta}{1-\eta} \right)^{1/2}
\]

\[
\tilde{x}_f = \left( \frac{\Gamma \rho}{(2-\Gamma)\delta_0} \right)^{1/2}, \quad \Gamma := \frac{\eta(\sigma-1)}{\epsilon-1-\sigma\eta}
\]

\[
\tilde{x}_{os} = 0
\]

\[
\tilde{x}_c = \left( \frac{1}{\bar{\kappa}} \cdot \frac{\eta}{1-\eta} \cdot \frac{\sigma-1}{\sigma} \right)^{1/2}
\]

- Firms fear creative destruction and share less than planner (\(\delta_0\))
- Consumers share less than planner because of mark up
- No sharing law restricts data even more
- Firms use more own-variety data compared to consumer/planner
Output

• For $\text{alloc} \in \{sp, c, f\}$:

$$Y_{t}^{\text{alloc}} = \left[ \nu_{\text{alloc}} (1 - \alpha) \eta \tilde{x}_{\text{alloc}}^{\eta} \right]^{\frac{1}{1-\eta}} \left( \psi_{\text{alloc}} L_t \right)^{1 + \frac{1}{\sigma - 1} + \frac{n}{1-\eta}}$$

• For Outlaw Sharing:

$$Y_{t}^{\text{os}} = \left[ \nu_{\text{os}} \alpha \eta x_{\text{os}}^{\alpha} \right]^{\frac{1}{1-\eta}} \left( \psi_{\text{os}} L_t \right)^{1 + \frac{1}{\sigma - 1}}$$

• Two source of increasing returns to scale:
  
  o Standard variety effect: $\frac{\sigma}{\sigma - 1}$
  
  o Data sharing: $\frac{\eta}{1-\eta}$

• Recall $\tilde{x}_t > 0$ from data sharing $\Rightarrow$ scale effect
Consumption per person and Growth

- Consumption per person:
  
  For $\textit{alloc} \in \{\textit{sp}, \textit{c}, \textit{f}\}$:  
  
  $c_{t\text{alloc}} = \text{Const}_{\text{alloc}} \cdot L_t^{\frac{1}{\sigma-1} + \frac{\eta}{1-\eta}}$

  For outlaw sharing:  
  
  $c_{t\text{os}} = \text{Const}_{\text{os}} \cdot L_t^{\frac{1}{\sigma-1}}$

- Per capita growth:

  $g_{c\text{sp}} = g_{c\text{f}} = g_{c\text{c}} = \left( \frac{1}{\sigma - 1} + \frac{\eta}{1 - \eta} \right) g_L$

  $g_{c\text{os}} = \left( \frac{1}{\sigma - 1} \right) g_L$

  **Intuition:** No sharing means you learn from 10 workers (constant firm size), sharing means you learn from the entire population
Numerical Example: How large is $\eta$?

- Error rate is proportional to $M^{-\eta}$. Productivity = $1/($error rate$)$

![Graph showing error rate vs. number of images (millions)]

- Estimated $\eta = 0.114$
- Doubling data lowers error by 8.2 percent

- Average $\eta = 0.08$. Double data $\Rightarrow$ 6% reduction in error rate
## Numerical Example: Other Parameters

<table>
<thead>
<tr>
<th>Description</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance of data</td>
<td>$\eta$</td>
<td>0.08</td>
</tr>
<tr>
<td>Elasticity of substitution</td>
<td>$\sigma$</td>
<td>5</td>
</tr>
<tr>
<td>Weight on privacy</td>
<td>$\kappa = \tilde{\kappa}$</td>
<td>0.20</td>
</tr>
<tr>
<td>Population level</td>
<td>$L_0$</td>
<td>100</td>
</tr>
<tr>
<td>Population growth rate</td>
<td>$g_L$</td>
<td>0.02</td>
</tr>
<tr>
<td>Rate of time preference</td>
<td>$\rho$</td>
<td>0.03</td>
</tr>
<tr>
<td>Labor cost of entry</td>
<td>$\chi$</td>
<td>0.01</td>
</tr>
<tr>
<td>Creative destruction</td>
<td>$\delta_0$</td>
<td>0.4</td>
</tr>
<tr>
<td>Weight on own data</td>
<td>$\alpha$</td>
<td>1/2</td>
</tr>
<tr>
<td>Use of own data in NS</td>
<td>$\bar{x}$</td>
<td>1</td>
</tr>
</tbody>
</table>
## Allocations

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Data Sharing “own” $x$</th>
<th>Data Sharing “others” $\tilde{x}$</th>
<th>Firm size $\nu$</th>
<th>Variety $N/L = \psi$</th>
<th>Consumption $c$</th>
<th>Growth $g$</th>
<th>Creative Destruct. $\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Planner</td>
<td>0.66</td>
<td>0.66</td>
<td>1304</td>
<td>665</td>
<td>18.6</td>
<td>0.67%</td>
<td>0.0870</td>
</tr>
<tr>
<td>Consumers Own Data</td>
<td>0.59</td>
<td>0.59</td>
<td>1482</td>
<td>594</td>
<td>18.3</td>
<td>0.67%</td>
<td>0.0696</td>
</tr>
<tr>
<td>Firms Own Data</td>
<td>1</td>
<td>0.16</td>
<td>1838</td>
<td>491</td>
<td>16.0</td>
<td>0.67%</td>
<td>0.0052</td>
</tr>
<tr>
<td>Outlaw Sharing</td>
<td>1</td>
<td>0</td>
<td>2000</td>
<td>455</td>
<td>7.3</td>
<td>0.50%</td>
<td>0</td>
</tr>
</tbody>
</table>

- Firms overuse their own data and undershare with others
- Consumers share less data than planner, but not by much
- Growth rate scale effect is modest, level differences are large
## Consumption Equivalent Welfare

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Welfare $\lambda$</th>
<th>$\log \lambda$</th>
<th>Level term</th>
<th>Privacy term</th>
<th>Growth term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal Allocation</td>
<td>1</td>
<td>0</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>Consumers Own Data</td>
<td>0.9886</td>
<td>-0.0115</td>
<td>-0.0202</td>
<td>0.0087</td>
<td>0.0000</td>
</tr>
<tr>
<td>Firms Own Data</td>
<td>0.8917</td>
<td>-0.1146</td>
<td>-0.1555</td>
<td>0.0409</td>
<td>0.0000</td>
</tr>
<tr>
<td>Outlaw Sharing</td>
<td>0.3429</td>
<td>-1.0703</td>
<td>-0.9399</td>
<td>0.0435</td>
<td>-0.1739</td>
</tr>
</tbody>
</table>

- Outlaw sharing: particularly harmful law (66 percent worse!)
- Firms own data: substantially lower welfare (11 percent worse)
- Consumers own data: nearly optimal (1 or 2 percent worse)
Conclusion

- Nonrival data $\Rightarrow$ large social gain from sharing data

- If firms own data, they may:
  - privately use more data than consumers/planner would
  - share less data across firms than consumers/planner would

- Nonrivalry $\Rightarrow$ Laws that outlaw sharing could be very harmful

- Consumers owning data good at balancing privacy and sharing