Consumer Value of Privacy: Evidence from an Online Retailer

Mimansa Bairathi **Ankit Sisodia**
Mayur Choudhary

Third Federal Trade Commission
Conference on Marketing and Public Policy

Privacy Paradox

Pew Research Center

RESEARCH TOPICS PUBLICATIONS OUR METHODS SHORT READS

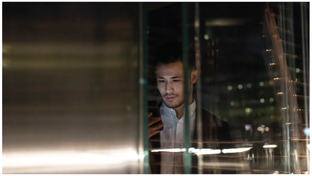
Home > Research Topics > Internet & Technology > Technology Policy Issues > Online Privacy & Sec

REPORT | NOVEMBER 15, 2019 SHARE

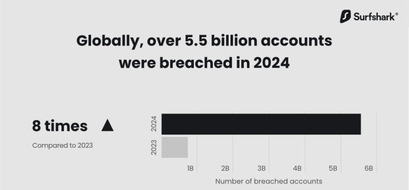
Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information

Majorities think their personal data is less secure now, that data collection poses more risks than benefits, and believe it is not possible to go through daily life without being tracked

BY BROOKE AUXIER, LEE RAINIE, MONICA ANDERSON, ANDREW PERRIN, MADHU KUMAR AND ERICA TURNER



(Garage Island Crew/Getty Images)



USD 4.88M

The global average cost of a data breach in 2024—a 10% increase over last year and the highest total ever.

Regulation Regarding Collection of Consumer Data

Regulation Regarding Collection of Consumer Data

- Consumers share data freely on social media, app permissions
- Why do consumers say they care about privacy but act as if they don't?

Related Literature

- Culnan and Bies (2003) explain this using privacy calculus
- Nominal incentives lead to data sharing (Athey et al., 2017)
- Value consumers place on personal data -
  - → Lin (2022) and Adjerid et al. (2019) use Qualtrics experiments
  - → John et al. (2011) use controlled lab experiments
  - → Acquisti et al. (2013) use artefactual field experiments
- Privacy is context dependent - Nissenbaum (2004) and Nissenbaum (2011)
- This paper - field setting
  - → Consumers interact with an e-commerce platform
  - → Consumer face consequential choices over repeated interactions

Research Questions

1. Does disclosing personal data carry a measurable economic cost?
   → What is the valuation of personal data?
2. Does the cost differ across different types of personal information?
3. Does the institutional context shape privacy valuations?

Empirical Setting - Online Fashion Retailer

- Data from a major online fashion retailer in India
- Tiered loyalty program
  - $\rightarrow$ Tier 1 - 0 - 499 points
  - $\rightarrow$ Tier 2 - $>=$ 500 points
- Earn points
  - $\rightarrow$ Purchase - 10 points per ₹100 (or $1.2).
  - $\rightarrow$ Engagement on platform via activities

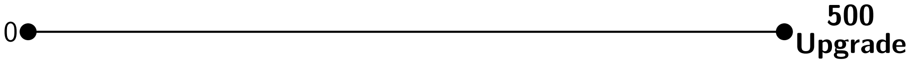
Campaigns Incentivizing Consumer Data Disclosure

- Consumers offered loyalty points in exchange for personal information
- Time-period
  - → December 15-19, 2018
  - → June 12-18, 2019
- Personal data
  - → Birthday - 10 points (June only)
  - → Mobile number - 50 points (December) and 30 points (June).
  - → Credit card number - 50 points (December) and 20 points (June)

Identification

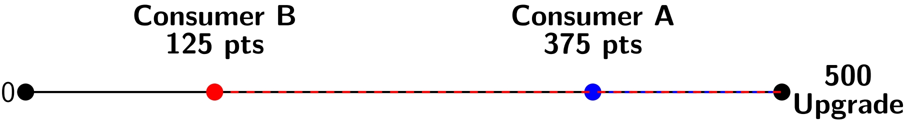
- Absolute number of loyalty points offered in exchange for a piece of personal information remains constant
- Challenge - Need exogenous variation in payoff for information disclosure
- We exploit the tiered structure of the loyalty program
- Consumers' distance (in points) to the upgrade threshold differs across consumers and across time
- Effective marginal benefit from a fixed number of loyalty points differs based on consumers' proximity to upgrade threshold (Chung et al., 2014; Hull, 1932; Kivetz et al., 2006; Misra and Nair, 2011)

Stylized Example

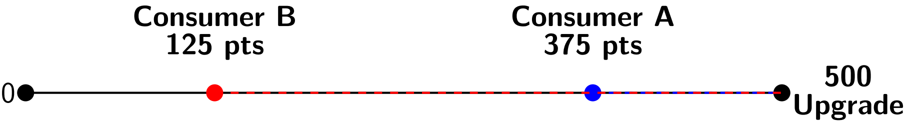


Stylized Example

Stylized Example

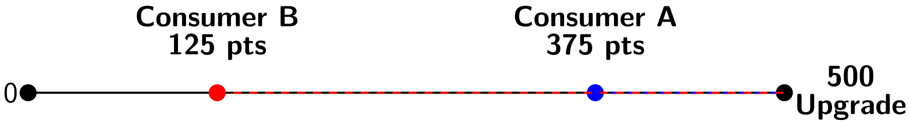


Stylized Example

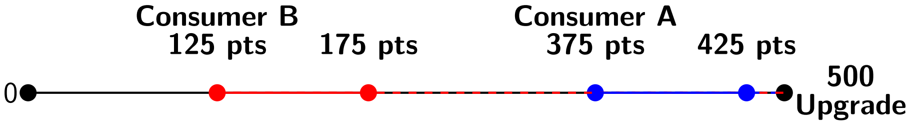


Retailer offers a consumer 50 points in exchange for mobile number

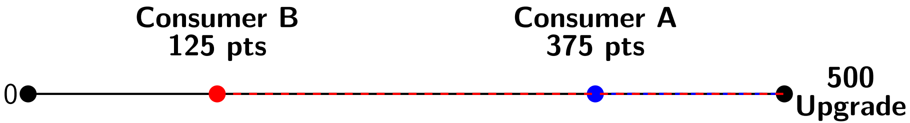
Stylized Example



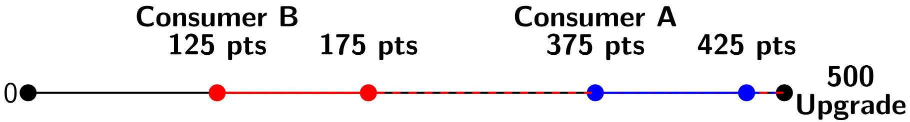
Retailer offers a consumer 50 points in exchange for mobile number



Stylized Example

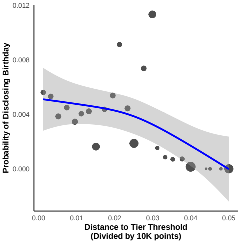


Retailer offers a consumer 50 points in exchange for mobile number

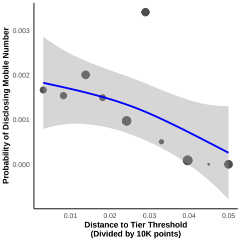


Effective marginal payoff for A $>$ B, since A closer to upgrade

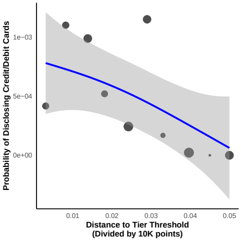
Model-Free Evidence



Birthday



Mobile number



Payment card details

Empirical Specification

For consumer i at time t,

$$\mathbb{1}\{\text{If disclose}\}_{it} = \beta_0 + \beta_1 \text{Distance (in points)}_{it} + \alpha_i + \epsilon_{it} \quad (1)$$

α_i controls for time-invariant differences in consumers inherent preference for privacy

Likelihood to Disclose Personal Data Decreases with Proximity

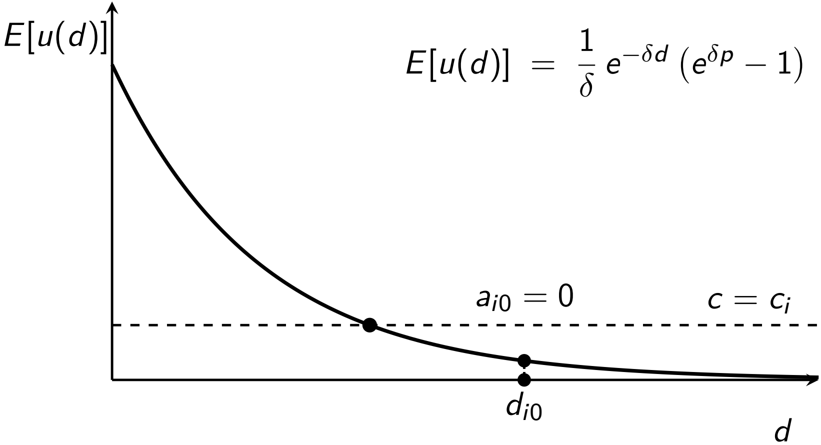
	(1) Date of birth	(2) Mobile number	(3) Payment card details
Distance to higher tier (points)	-1.212***	-0.116***	-0.043***
	(0.049)	(0.008)	(0.004)
Customer F.E.	Yes	Yes	Yes
Observations	110,708	233,398	233,735

Notes: Clustered standard errors in parentheses. Distance to higher tier (points) divided by 10,000.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

How Much Do Consumers Value Privacy?

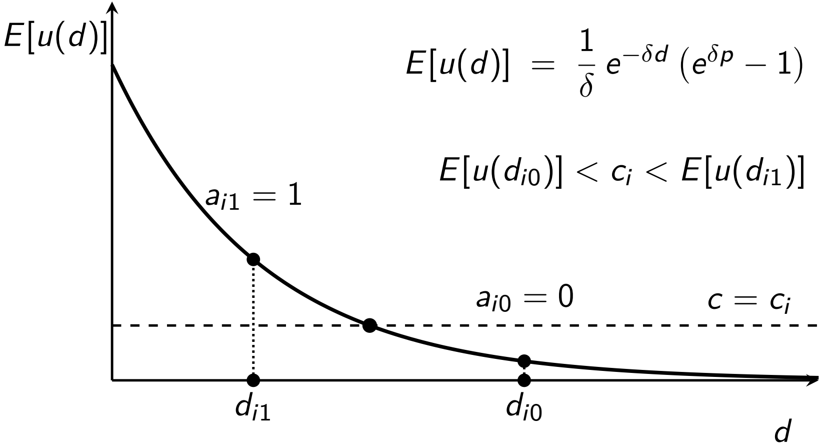
- Distance from upgrade threshold is independent of consumers' valuation of personal information
- Use this feature to back out the distribution of valuation using consumers' binary choice
- Assume consumers churn at a Poisson rate δ

Distance and Expected Utility



$$a_i(d) = \begin{cases} 1 & c_i \leq E[u(d)] = f(d) \\ 0 & \text{otherwise} \end{cases}$$

Distance and Expected Utility



$$a_i(d) = \begin{cases} 1 & c_i \leq E[u(d)] = f(d) \\ 0 & \text{otherwise} \end{cases}$$

Empirical Distribution of Private Valuation

- The identifying assumption is that the distance is independent of private valuation, i.e.,

$$d_i \perp\!\!\!\perp c_i \text{ for all } i$$

- The estimator for the distribution conditional on disclosure, i.e., $I = \{i : a_i = 1\}$

$$\hat{F}_C(c) = \hat{P}r(C < c = f(d)|i \in I) = \frac{\hat{P}r(C < f(d), i \in I)}{\hat{P}r(i \in I)}$$

- Since d_i is assigned independent of C, [1] $\hat{P}r(C < f(d), i \in I) = \frac{\sum_{i=1}^{N} \mathbb{1}\{d_i > f^{-1}(c), a_i = 1\}}{N}$, so that,

$$\hat{F}_C(c) = \frac{\hat{P}r(C < c, i \in I)}{\hat{P}r(i \in I)} = \frac{\sum_{i=1}^{N} \mathbb{1}\{d_i > f^{-1}(c), a_i = 1\}}{\sum_{i=1}^{N} \mathbb{1}\{a_i = 1\}}$$

[1] $\hat{P}r(i \in I) = \frac{\sum_{i=1}^{N} \mathbb{1}\{a_i = 1\}}{N}$

Valuation of Personal Data

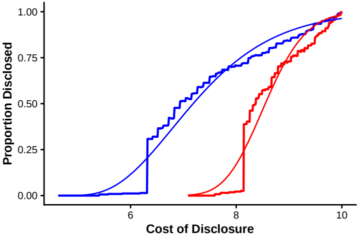
$$\boxed{\hat{\bar{c}} = E(c) = \int_0^\infty (1 - \hat{F}_C(c))dc}$$

	\$ Valuation of personal data	
	Mean	p-value
Birthday	\$0.10	< 0.01
Mobile number	\$0.52	< 0.01
Payment card details	\$0.54	< 0.01

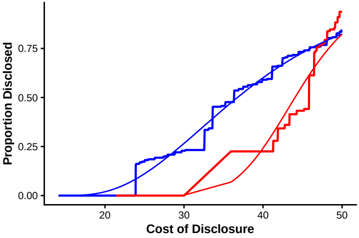
Does Institutional Context Shape Privacy Valuations?

- Explore heterogeneity in disclosure costs based on consumers' ability to earn loyalty points through other means
  1. Purchase
  2. Activities on platform e.g. games, campaigns

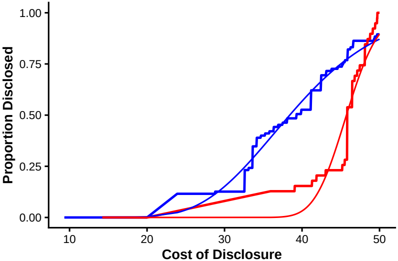
Consumers Spending More Value Personal Information More



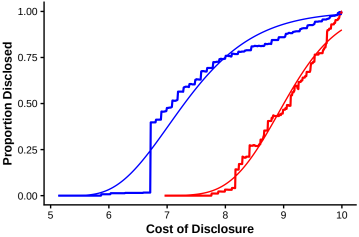
(a) Birthday



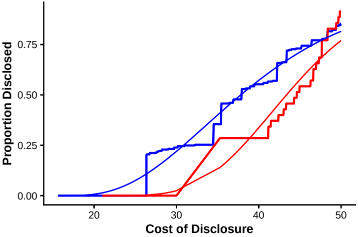
(b) Mobile number



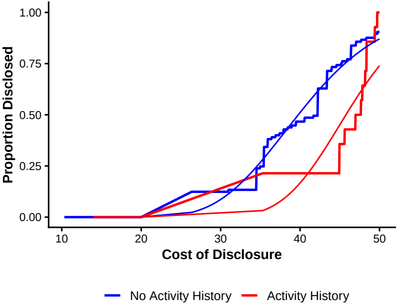
More Active Consumers Value Personal Information More



(a) Birthday



(b) Mobile number



Consumers with Alternate Means to Payoffs Value Personal Information More

	Past spend			Past activity		
	(1)	(2)	(3)	(4)	(5)	(6)
	Low Spenders	High Spenders	Δ	Less Active	More Active	Δ
Date of birth	0.10	0.11	0.01***	0.10	0.12	0.02***
Mobile number	0.51	0.58	0.07***	0.52	0.57	0.05***
Payment card details	0.51	0.60	0.09***	0.53	0.60	0.07***

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Implications

Managers

- Data collection is most effective when incentives align with
  - $\rightarrow$ Sensitivity of information
  - $\rightarrow$ Institutional structure
- Consumers with alternate pathways have a higher valuation

Policymakers

- Limitations of notice-and-consent frameworks
- Effective regulation needs to account for context and behavioral influences

THANK YOU