Privacy-Preserving Analytics on the Ground

Ryan Steed, Alessandro Acquisti
Carnegie Mellon University

PEPR — 09/11/2023

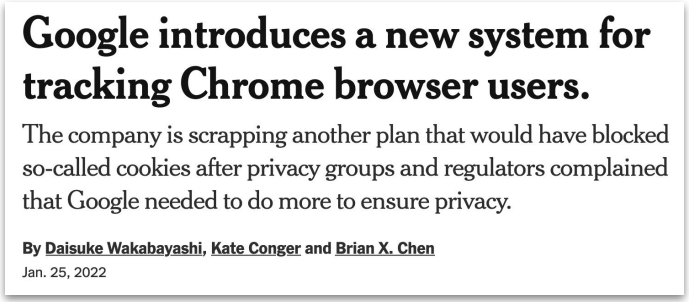
“Privacy-preserving” analytics (PPA)

- Subset of Privacy Enhancing Technology (PET)
  - Differential privacy
  - Secure multiparty computation
  - Privacy-preserving machine learning (PPML)
  - …
- Landmark adoption—and controversy
  - Differential privacy in the **2020 U.S. Decennial Census** (Abowd et al., 2022)—despite protests from stakeholders (boyd & Sarathy, 2022)
  - **Google’s Privacy Sandbox** to replace third party cookies (Goel, 2022)—while preserving targeted advertising (Cyphers, 2021)

The 2020 Census Suggests That People Live Underwater. There’s a Reason.

Technology advances forced the Census Bureau to use sweeping measures to ensure privacy for respondents. The ensuing debate goes to the heart of what a census is.

By Michael Wines
April 21, 2022



Google introduces a new system for tracking Chrome browser users.

The company is scrapping another plan that would have blocked so-called cookies after privacy groups and regulators complained that Google needed to do more to ensure privacy.

By Daisuke Wakabayashi, Kate Conger and Brian X. Chen
Jan. 25, 2022

Research questions

PPA adoption is growing, but

- Why are organizations adopting PPA techniques?
- How might PPA adoption *not* lead to better privacy online?

Research questions

PPA adoption is growing, but

- Why are organizations adopting PPA techniques?
- How might PPA adoption *not* lead to better privacy online?

Our work

➔ Grounded theory of PPA adoption

➔ Lessons from economics, sociology, STS, and law

Takeaways:

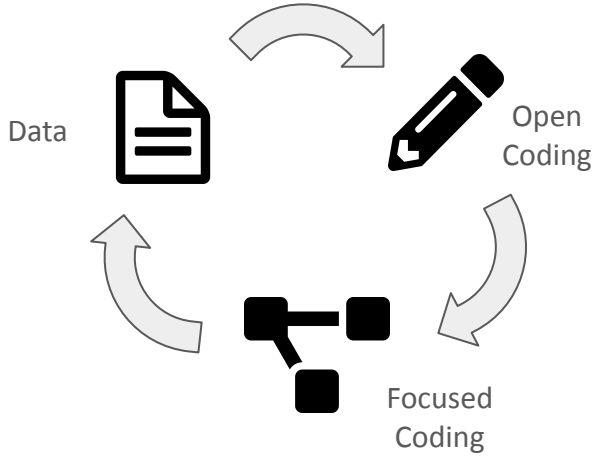
➔ The importance of interpretation

➔ Pathways to “privacy theater”

➔ Recommendations for practitioners and researchers

Our study: emergent process theory

- Method: grounded theory (Charmaz, 2014) and thematic analysis
- 1-hour semi-structured interviews, Sep. 2021–Jan. 2022 & Aug. 2023



N=28 PPA practitioners

- Execs/directors (N=11), managers (N=4), ICs (N=13) doing research, engineering, product, policy, legal
- 90% U.S.-based, 55% white, 75% cis men, 75% straight

21 organizations

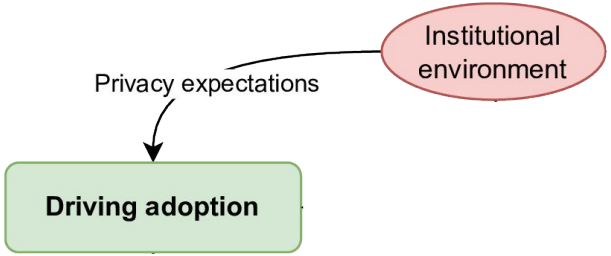
- 8 tech companies (N=13) — 6 are Fortune 500 (N=12)
- 5 privacy startups (N=6)
- 4 non-profits (N=5)
- 3 government agencies (N=4)

The literature

Why adopt socially responsible tech?

- Reduce financial risk from legal penalties; maintain social license (e.g. Carroll, 1979; Jones, 1995; Gunningham, 2004)
- Example: Execs adopted new privacy-by-design policies in response to changing privacy norms & regs (Bamberger & Mulligan, 2015)

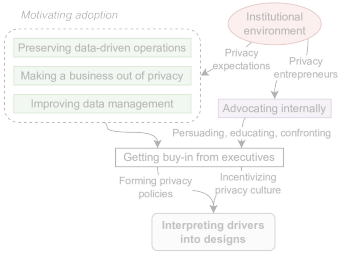
Our (emergent) process theory



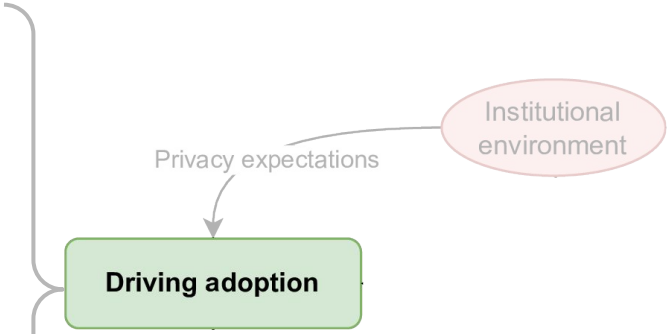
Driving adoption

Motivating adoption

- Preserving data-driven operations
- Making a business out of privacy
- Improving data management



Our (emergent) process theory



"We've learned a lot from [Facebook's privacy] controversies, but **we continue to believe personalized ads and privacy can coexist**… That's why we're investing in R&D of privacy-enhancing technologies."

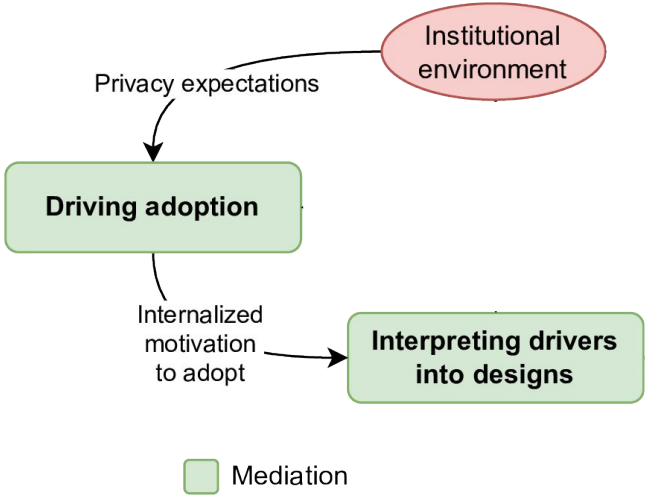
— Erin Egan, Meta Chief Privacy Officer (2020)

The literature

But, new policies do not guarantee changes to practice…

- Organizations may "decouple"—or **mediate**—policy from practice (Meyer & Rowan, 1977; Weick, 1976)
- More likely early on, or when adoption mostly due to external expectations (Bromley & Powell, 2012)
- Example: Many technologists/lawyers still didn't consider privacy in daily work (Waldman, 2017)

Our (emergent) process theory



Interpreting drivers into designs

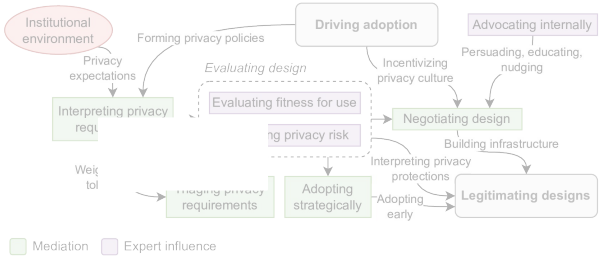
Evaluating design

Interpreting privacy requirements

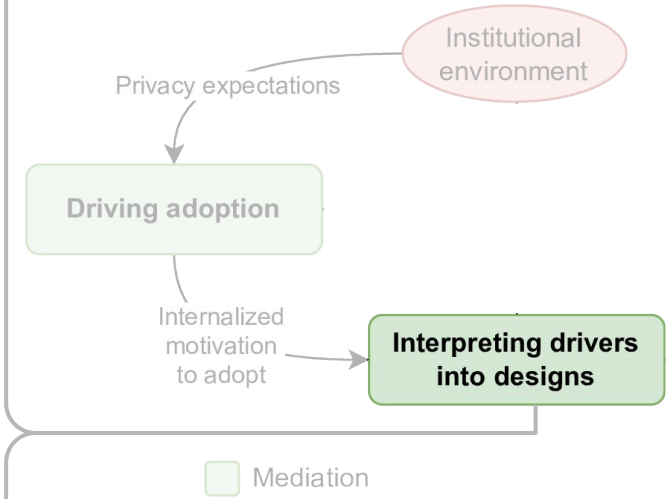
Evaluating fitness for use

Evaluating privacy risk

Negotiating design



Our (emergent) process theory

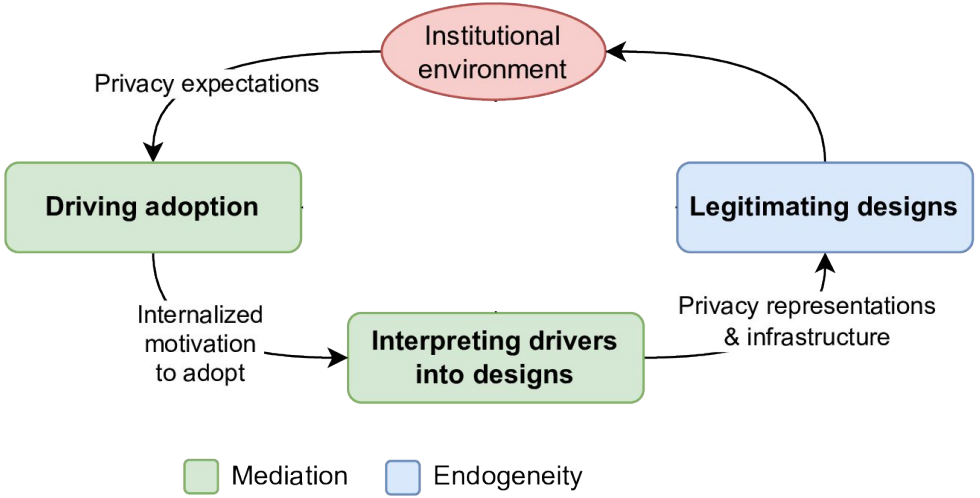


The literature

… and practice shapes expectations.

- Organizations make “educated guesses” about compliance (Edelman, 1999)
- Models are endorsed & spread—through industry networks, court decisions, sponsored research, lobbying (e.g. Wilson, 1982; Edelman, 2016; Kamieniecki, 2006)

Our (emergent) process theory



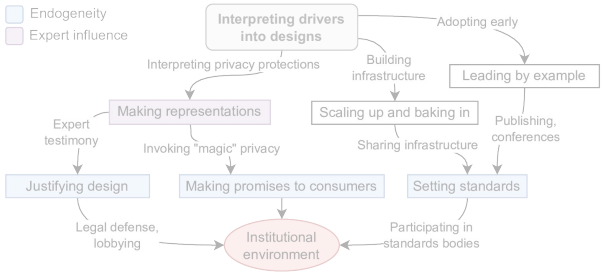
Legitimating designs

Making representations

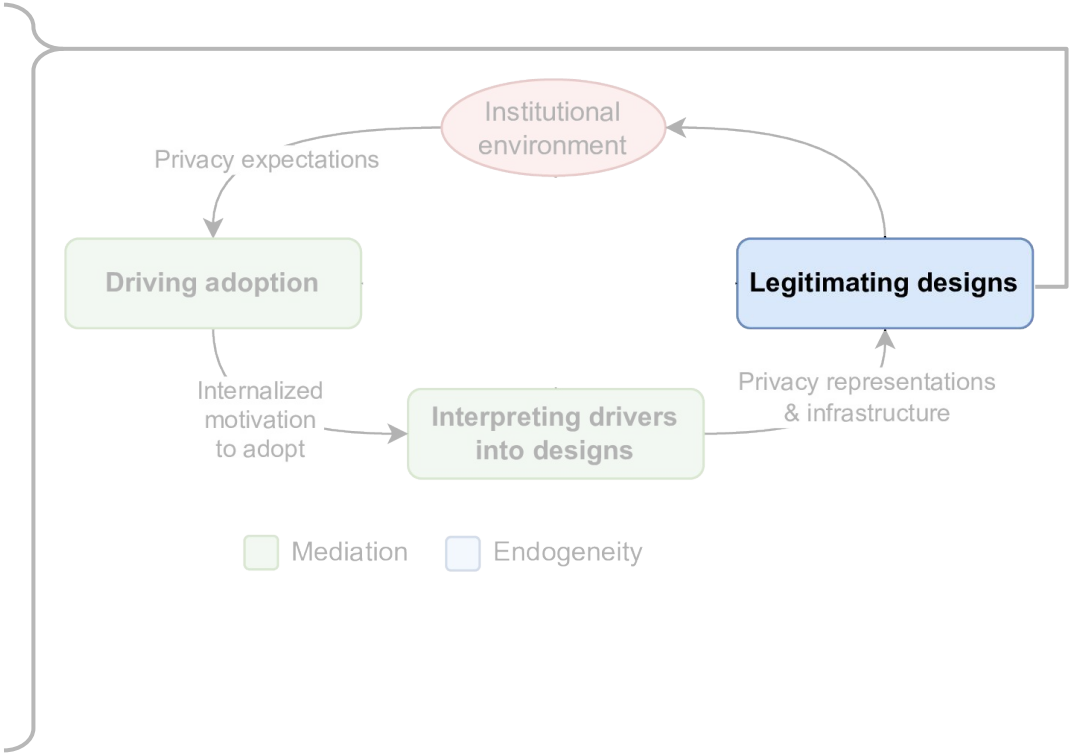
Justifying design

Making promises to consumers

Setting standards



Our (emergent) process theory

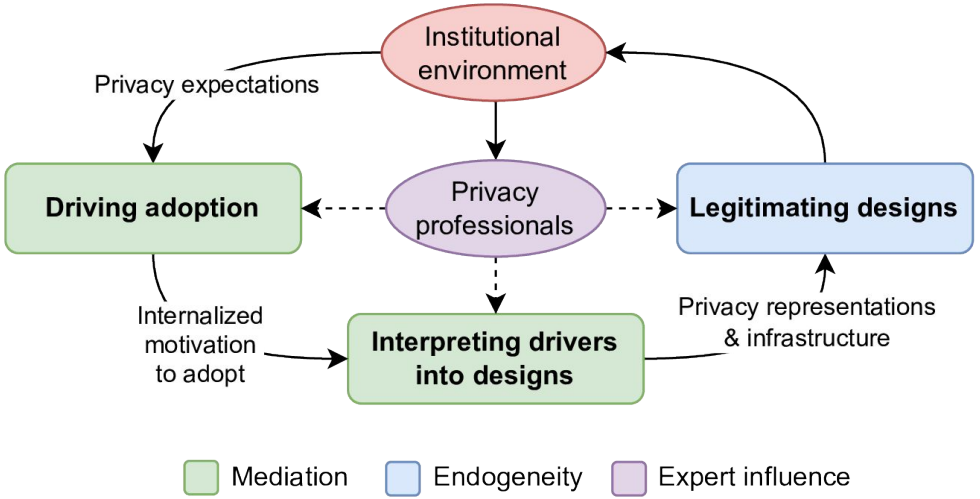


The literature

Decoupling is harder when employees are moral activists (Turco, 2012).

- "Privacy champions" evangelize privacy in daily work (Tahaei et al., 2021).
- Moral leaders can make and safeguard institutional reforms (Solinger, 2020)…
- But they may struggle in metrics-oriented, move-fast environments (Ali et al., 2023).

Our (emergent) process theory



Preserving privacy in “privacy-preserving” analytics

- How practitioners can help:
  - Establish **best practice & defaults** for communication, parameter setting early in development
  - **Share design choices** and privacy-relevant settings with independent experts and/or the public
  - Advocate internally—**build & maintain internal privacy groups & substantive standards**
  - Consider whether a given analytics practice is appropriate **regardless of PPA**
- How researchers can help:
  - **Empirically evaluate** systems after deployment
  - Consider ripple effects of adoption (e.g., encouraging more or less data collection)
  - Develop for PPA tasks that **shift power to users**—e.g., privacy-preserving auditing (Xu & Zhang, 2021)
- How policymakers can help:
  - **Deeper investigation** before affirming PPA practices; avoid blanket endorsements (see e.g. Edelman, 2016)
  - Require disclosure of key design details, or access for independent PPA auditors

Thank you!

Questions? Thoughts?
Want to read the paper?

ryansteed@cmu.edu

References

Abowd, John M., Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, et al. 2022. “The 2020 Census Disclosure Avoidance System TopDown Algorithm.” *Harvard Data Science Review*, no. Special Issue 2 (June). https://doi.org/10.1162/99608f92.529e3cb9.

Ali, Sanna J., Angèle Christin, Andrew Smart, and Riitta Katila. 2023. “Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs.” In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 217–26. FAccT ’23. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3593013.3593990.

Bamberger, Kenneth A., and Deirdre K. Mulligan. 2015. *Privacy on the Ground: Driving Corporate Behavior in the United States and Europe*. Information Policy Series. Cambridge, Massachusetts: The MIT Press.

boyd, danah, and Jayshree Sarathy. 2022. “Differential Perspectives: Epistemic Disconnects Surrounding the U.S. Census Bureau’s Use of Differential Privacy.” *Harvard Data Science Review*, no. Special Issue 2 (June). https://doi.org/10.1162/99608f92.66882f0e.

Bromley, Patricia, and Walter W. Powell. 2012. “From Smoke and Mirrors to Walking the Talk: Decoupling in the Contemporary World.” *The Academy of Management Annals* 6 (1): 483–530. https://doi.org/10.1080/19416520.2012.684462.

Carroll, Archie B. 1979. “A Three-Dimensional Conceptual Model of Corporate Performance.” *Academy of Management Review* 4 (4): 497–505. https://doi.org/10.5465/AMR.1979.4498296.

Charmaz, Kathy. 2014. *Constructing Grounded Theory*. 2nd edition. Introducing Qualitative Methods. London ; Thousand Oaks, Calif: Sage.

Cyphers, Bennett. 2021. “Google’s FLoC Is a Terrible Idea.” Electronic Frontier Foundation. March 3, 2021. https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea.

Edelman, Lauren B. 2016. *Working Law: Courts, Corporations, and Symbolic Civil Rights*. Chicago Series in Law and Society. Chicago, IL: University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/W/bo24550454.html.

Edelman, Lauren B., Christopher Uggen, and Howard S. Erlanger. 1999. “The Endogeneity of Legal Regulation: Grievance Procedures as Rational Myth.” *American Journal of Sociology* 105 (2): 406–54. https://doi.org/10.1086/210316.

Egan, Erin. 2020. “A Path Forward for Privacy and Online Advertising.” *Meta* (blog). October 2, 2020. https://about.fb.com/news/2020/10/a-path-forward-for-privacy-and-online-advertising/.

Gunningham, Neil, Robert A. Kagan, and Dorothy Thornton. 2004. “Social License and Environmental Protection: Why Businesses Go beyond Compliance.” *Law & Social Inquiry* 29 (2): 307–41.

Jones, Thomas M. 1995. “Instrumental Stakeholder Theory: A Synthesis of Ethics and Economics.” *The Academy of Management Review* 20 (2): 404–37. https://doi.org/10.2307/258852.

Kamieniecki, Sheldon. 2006. *Corporate America and Environmental Policy: How Often Does Business Get Its Way?* Stanford: Stanford University Press.

Meyer, John W., and Brian Rowan. 1977. “Institutionalized Organizations: Formal Structure as Myth and Ceremony.” *American Journal of Sociology* 83 (2): 340–63.

Solinger, Omar N., Paul G.W. Jansen, and Joep P. Cornelissen. 2020. “The Emergence of Moral Leadership.” *Academy of Management Review* 45 (3): 504–27. https://doi.org/10.5465/amr.2016.0263.

Tahaei, Mohammad, Alisa Frik, and Kami Vaniea. 2021. “Privacy Champions in Software Teams: Understanding Their Motivations, Strategies, and Challenges.” In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. CHI ’21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445768.

Turco, Catherine. 2012. “Difficult Decoupling: Employee Resistance to the Commercialization of Personal Settings.” *American Journal of Sociology* 118 (2): 380–419. https://doi.org/10.1086/666505.

Waldman, Ari Ezra. 2017. “Designing Without Privacy.” SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=2944185.

Weick, Karl E. 1976. “Educational Organizations as Loosely Coupled Systems.” *Administrative Science Quarterly* 21 (1): 1–19. https://doi.org/10.2307/2391875.

Wilson, James Q. 1982. *The Politics of Regulation*. New York: Basic Books.