Privacy Substitutes

Jonathan Mayer Arvind Narayanan

Introduction

Debates over information privacy are often framed as an inescapable conflict between competing interests: a lucrative or beneficial technology, as against privacy risks to consumers. Policy remedies traditionally take the rigid form of either a complete ban, no regulation, or an intermediate zone of modest notice and choice mechanisms.

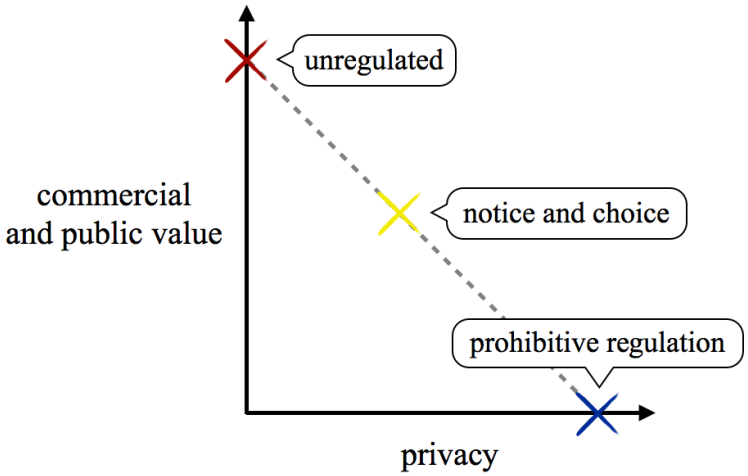
We believe these approaches are unnecessarily constrained: there is often a spectrum of technology alternatives that trade off functionality and profit for consumer privacy. We term these alternatives "privacy substitutes," and in this article we argue that public policy on information privacy issues can and should be a careful exercise in both selecting among and providing incentives for privacy substitutes.

Disconnected Policy and Computer Science Perspectives

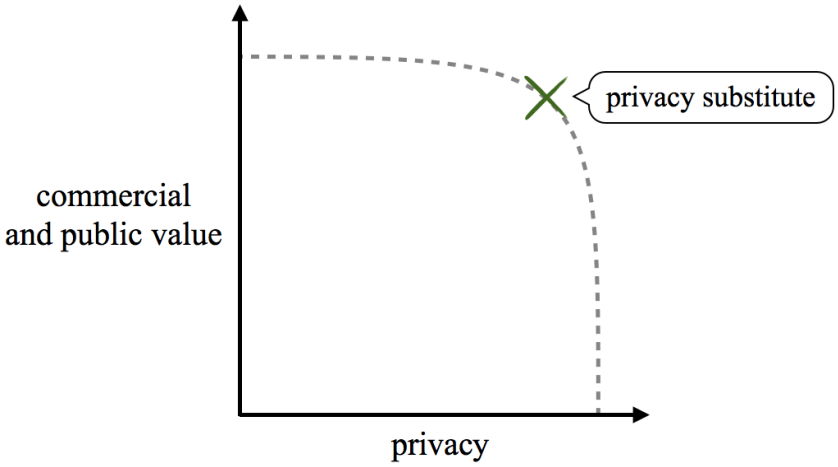
Policy stakeholders frequently approach information privacy through a rote balancing. Consumer privacy interests rest on one side of the scales, and commercial and social benefits sit atop the other.[1] Where privacy substantially tips the balance, a practice warrants prohibition; where privacy is significantly outweighed, no restrictions are appropriate. When the scales near

[1] *See, e.g.*, *Balancing Privacy and Innovation: Does the President's Proposal Tip the Scale?*, 112th Cong. 1 (2012); (statement of Rep. Mary Bono Mack, Chairman, H. Subcomm. on Commerce, Mfg., & Trade) ("When it comes to the Internet, how do we—as Congress, as the administration, and as Americans—balance the need to remain innovative with the need to protect privacy?"); FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE 36 (2012) ("Establishing consumer choice as a baseline requirement for companies that collect and use consumer data, while also identifying certain practices where choice is unnecessary, is an appropriately balanced model.").

equipoise, practices merit some (questionably effective) measure of mandatory disclosure or consumer control.[2]



Computer science researchers, however, have long recognized that technology can enable tradeoffs between privacy and other interests. For most areas of technology application, there exist a spectrum of possible designs that vary in their privacy and functionality[3] characteristics. Cast in economic terms, technology enables a robust production-possibility frontier between privacy and profit, public benefit, and other values.



[2] Recent scholarship has challenged the efficacy of notice and choice models for technology privacy. *E.g.*, Yang Wang et al., *Privacy Nudges for Social Media: An Exploratory Facebook Study*, PROC. 22D INT'L CONF. ON WORLD WIDE WEB (2012); Pedro Giovanni Leon et al., *What Do Online Behavioral Advertising Privacy Disclosures Communicate to Users?*, PROC. 2012 ACM WORKSHOP ON PRIVACY ELECTRONIC SOC'Y (2012).

[3] Including speed, accuracy, usability, cost, technical difficulty, security, and more.

The precise contours of the production-possibility frontier vary by technology application area. In many areas, privacy substitutes afford a potential Pareto improvement relative to naïve or status quo designs. In some application areas, privacy substitutes even offer a strict Pareto improvement: privacy-preserving designs can provide the *exact same* functionality as intrusive alternatives. The following subparts review example designs for web advertising, online identity, and transportation payment to illustrate how clever engineering can counterintuitively enable privacy tradeoffs.

Web advertising. In the course of serving an advertisement, dozens of third-party websites may set or receive a unique identifier cookie.[4] The technical design is roughly akin to labeling a user's web browser with a virtual barcode, then scanning the code with every page view. All advertising operations—from selecting which ad to display through billing—can then occur on advertising company backend services. Policymakers and privacy advocates have criticized this status quo approach as invasive since it incorporates collection of a user's browsing history.[5] Privacy researchers have responded with a range of technical designs for advertising functionality.[6]

Proposal	**Design Elements**	**Information Shared**
PrivAd	• Anonymizing intermediary service	None

[4] *See* Jonathan R. Mayer & John C. Mitchell, *Third-Party Web Tracking: Policy and Technology*, PROC. 2012 IEEE SYMP. ON SECURITY & PRIVACY 3-5 (2012).
[5] *Id.* at 4-5.
[6] *E.g.*, Michael Backes et al., *ObliviAd: Provably Secure and Practical Online Behavioral Advertising*, PROC. 2012 IEEE SYMP. ON SECURITY & PRIVACY (2012); Mikhail Bilenko et al., *Targeted, Not Tracked: Client-side Solutions for Privacy-friendly Behavioral Advertising*, PROC. 11TH PRIVACY ENHANCING TECHS. SYMP. (2011); Matthew Fredrikson & Ben Livshits, *RePriv: Re-envisioning In-browser Privacy*, PROC. 2011 IEEE SYMP. ON SECURITY & PRIVACY (2011); Saikat Guha et al., *Privad: Practical Privacy in Online Advertising*, PROC. 8TH USENIX CONF. ON NETWORKED SYSTEMS DESIGN & IMPLEMENTATION (2011); Vincent Toubiana et al., *Adnostic: Privacy Preserving Targeted Advertising*, PROC. 17TH NETWORK & DISTRIBUTED SYSTEMS SYMP. (2010); Jonathan Mayer & Arvind Narayanan, *Tracking Not Required: Frequency Capping*, WEB POLICY (Apr. 23, 2012), http://webpolicy.org/2012/04/23/tracking-not-required-frequency-capping/; Arvind Narayanan et al., *Tracking Not Required: Behavioral Advertising*, 33 BITS OF ENTROPY (June 11, 2012), http://33bits.org/2012/06/11/tracking-not-required-behavioral-targeting/; Jonathan Mayer & Arvind Narayanan, *Tracking Not Required: Advertising Measurement*, WEB POLICY (July 24, 2012), http://webpolicy.org/2012/07/24/tracking-not-required-advertising-measurement/.

	• Computation in the web browser • Storage in the web browser	
ObliviAd	• Cryptography • Secure hardware	• Protocol information
Adnostic	• Computation in the web browser • Cryptography • Storage in the web browser • Trusted intermediary service	• Protocol information
Tracking Not Required	• Coarsened information • Computation in the web browser • Pre-computed information • Storage in the web browser	• Protocol information • Low entropy information
RePriv	• Computation in the web browser • Storage in the web browser	• Protocol information • Interest segments
Targeted, Not Tracked	• Storage in the web browser	• Protocol information • Recent browsing activity

The designs above represent points in the spectrum of possible tradeoffs between privacy—here, the information shared with advertising companies—and other commercial and public values. Moving from top to bottom, proposals become easier to deploy, faster in delivery, and more accurate in advertisement selection and reporting—in exchange for diminished privacy guarantees.

Online identity. Centralized online identity management benefits consumers through both convenience and increased security.[7] Popular implementations of these "single sign-on" or "federated identity" systems include a sharp privacy drawback, however: the identity provider learns about the consumer's activities. By way of rough analogy: imagine going to a bar, where the bouncer phones the state DMV to check the authenticity of your driver's license. The bouncer gets confirmation of your identity, but the DMV learns where you are. Drawing on computer security research, Mozilla has deployed a privacy-preserving alternative, dubbed Persona. Through the use of cryptographic attestation, Persona provides centralized identity

[7] *See Why Persona?*, MOZILLA DEVELOPER NETWORK, https://developer.mozilla.org/en-US/docs/Mozilla/Persona/Why_Persona (last visited June 29, 2013).

management without Mozilla learning the consumer's online activity. Extending the bar analogy: instead of calling the DMV, the bouncer instead carefully checks the driver's license for official and difficult-to-forge markings. The bouncer can still be sure of your identity, but the DMV does not learn of your drinking habits.

Transportation payment. Transportation fare cards and toll tags commonly embed unique identifiers, facilitating intrusive tracking of a consumer's movements. Intuitively, the alternative privacy-preserving design would be to store the consumer's balance on the device—but this approach is vulnerable to cards being hacked for free transportation.[8] An area of cryptography called secure multiparty computation provides a solution, allowing two parties to transact while only learning as much about each other as is strictly mathematically necessary to complete the transaction.[9] A secure multiparty computation approach would enable the transportation provider to reliably add and deduct credits from a card or tag—without knowing the precise device or value stored.

Non-Adoption of Privacy Substitutes

Technology organizations have rarely deployed privacy substitutes, despite their promise. A variety of factors have effectively undercut commercial implementation.

[8] *See, e.g.*, Loek Essers, *Android NFC Hack Enables Travelers To Ride Subways for Free, Researchers Say*, COMPUTERWORLD (Sept. 20, 2012), https://www.computerworld.com/s/article/9231500/Android_NFC_hack_enables_travelers_to_ride_subways_for_free_researchers_say.

[9] Secure multiparty computation has been implemented in various well-known protocols. The area traces its roots to Andrew Yao's "garbled circuit construction," a piece of "crypto magic" dating to the early 1980s. Researchers have used secure multiparty computation to demonstrate privacy-preserving designs in myriad domains—voting, electronic health systems and personal genetics, and location-based services, to name just a few. The payment model we suggest is based on David Chaum's "e-cash." His company DigiCash offered essentially such a system (not just for transportation, but for all sorts of payments) in the 1990s, but it went out of business by 1998. *See generally How DigiCash Blew Everything*, NEXT! MAG., Jan. 1999, *available at* http://cryptome.org/jya/digicrash.htm.

Engineering conventions. Information technology design traditionally emphasizes principles including simplicity, readability, modifiability, maintainability, robustness, and data hygiene. More recently, overcollection has become a common practice—designers gather information wherever feasible, since it might be handy later. Privacy substitutes often counterintuitively turn these norms on their head. Consider, for example, "differential privacy" techniques for protecting information within a dataset.[10] The notion is to intentionally introduce (tolerable) errors into data, a practice that cuts deeply against design intuition.[11]

Information asymmetries. Technology organizations may not understand the privacy properties of the systems they deploy. For example, participants in the online advertising frequently claim that their practices are anonymous—despite substantial computer science research to the contrary.[12] Firms may also lack the expertise to be aware of privacy substitutes; as the previous part showed, privacy substitutes often challenge intuitions and assumptions about technical design.

Implementation and switching costs. The investments of labor, time, and capital associated with researching and deploying a privacy substitute may be significant. Startups may be particularly resource constrained, while mature firms face path-dependent switching costs owing to past engineering decisions.

Diminished private utility. Intrusive systems often outperform privacy substitutes (e.g. in speed, accuracy, and other aspects of functionality), in some cases resulting in higher private utility. Moreover, the potential for presently unknown future uses of data counsels in favor of overcollection wherever possible.

---

[10] *See generally* Cynthia Dwork, *Differential Privacy: A Survey of Results*, PROC. 5TH INT'L CONF. ON THEORY & APPLICATIONS OF MODELS OF COMPUTATION (2008).

[11] Most production systems have data errors, in fact, but they are subtle and underappreciated. Differential privacy is ordinarily a matter of kind and degree of error, not whether error exists at all.

[12] *See*, *e.g.*, Mayer & Mitchell, *supra* note 3, at 3-4. Some of these misstatements may, of course, be strategic.

Inability to internalize. In theory, consumers or business partners might compensate a firm for adopting privacy substitutes. In practice, however, internalizing the value of pro-privacy practices has proven challenging. Consumers are frequently unaware of the systems that they interact with, let alone the privacy properties of those systems; informing users sufficiently to exercise market pressure may be impracticable.[13] Moreover, even if a sizeable share of consumers were aware, it may be prohibitively burdensome to differentiate those consumers who are willing and able to pay for privacy. And even if those users could be identified, it may not be feasible to transfer small amounts of capital from those consumers. As for business partners, they too may have information asymmetries and reflect (indirectly) lack of consumer pressure. Coordination failures compound the difficulty of monetizing privacy: without clear guidance on privacy best practices, users, businesses, and policymakers have no standard of conduct to request adherence to.

Organizational divides. To the extent technology firms do perceive pressure to adopt privacy substitutes, it is often through government relations, policy, and law staff. In some industries the motivation will be one further step removed, filtering through trade associations and lobbying groups. These non-technical representatives often lack the expertise to propose privacy alternatives themselves or adequately solicit engineering input.[14]

Competition barriers. Some technology sectors reflect monopolistic or oligopolistic structures. Even if users and businesses demanded improved privacy, there may be little competitive pressure to respond.

---

[13] In theory, uniform privacy signaling mechanism or trust intermediaries might assist in informing users. In practice, both approaches have had limited value. *See, e.g.*, Benjamin Edelman, *Adverse Selection in Online "Trust" Certifications*, Proc. Int'l Conf. on E-Commerce (2009) (studying efficacy of website certification providers); Adrienne Porter Felt et al., *Android Permissions: User Attention, Comprehension, and Behavior*, PROC. SYMP. ON USABLE SECURITY & PRIVACY (2012) (exploring usability of the Android device permissions model).

[14] We have observed the difficulty imposed by organizational divides firsthand in the World Wide Web Consortium's process to standardize Do Not Track: participants from the online advertising industry have largely been unable to engage on privacy alternatives owing to organizational divides.

Policy Prescriptions

Our lead recommendation for policymakers is straightforward: understand and encourage privacy substitutes using ordinary regulatory practices. When approaching a consumer privacy problem, policymakers should begin by exploring not only the relevant privacy risks and competing values, but also the space of possible privacy substitutes and their associated tradeoffs. If policymakers are sufficiently certain that socially beneficial privacy substitutes exist,[15] they should turn to conventional regulatory tools to incentivize deployment of those technologies.[16]

Policymakers should also target the market failures that lead to non-adoption of privacy substitutes. Engaging directly with industry engineers, for example, may overcome organizational divides and information asymmetries. We are skeptical of the efficacy of consumer education efforts,[17] but informing business partners could alter incentives.

Finally, policymakers should press the envelope of privacy substitutes. Grants and competitions, for example, could drive research innovations in both academia and industry.

Conclusion

This brief article is intended to begin reshaping policy debates on information privacy, from stark and unavoidable conflicts to creative and nuanced tradeoffs. Much more remains to be said: Can privacy substitutes also reconcile individual privacy with government intrusions (e.g.

[15] Sometimes a rigorously vetted privacy substitute will be ready for deployment. Frequently, to be sure, the space of privacy substitutes will include gaps and ambiguities. But policymakers are no strangers to decisions under uncertainty and relying on the best available science.

[16] We caution against requiring particular technical designs: in future, better designs may become available, or deficiencies in present designs may be uncovered. Cast in more traditional terms of regulatory discourse, this is very much an area for targeting ends, not means.

[17] *See supra* note 2.

for law enforcement or intelligence)?[18] How can policymakers recognize privacy substitute pseudoscience?[19] We leave these and many more questions for another day, and part ways on this note: Pundits often cavalierly posit that information technology has sounded the death knell for individual privacy. We could not disagree more. Information technology is poised to protect individual privacy—if policymakers get the incentives right.

[18] The congressional response to Transportation Security Administration full-body scanners might be considered an instance of a privacy substitute. Congress allowed the TSA to retain the scanners, but required a software update that eliminated intrusive imaging. 49 U.S.C. § 44901(*l*) (2012).

[19] For example, some technology companies are lobbying for European Union law to exempt pseudonymous data from privacy protections. *See* CTR. DEMOCRACY & TECH, CDT POSITION PAPER ON THE TREATMENT OF PSEUDONYMOUS DATA UNDER THE PROPOSED DATA PROTECTION REGULATION (2013), *available at* https://www.cdt.org/files/pdfs/CDT-Pseudonymous-Data-DPR.pdf. Information privacy researchers have, however, long recognized that pseudonymous data can often be linked to an individual. *See, e.g.*, Mayer & Mitchell, *supra* note 3, at 3-4.