Undoing the Privacy Paradox with Data Styles

Eric P. S. Baumer
Computer Science & Engineering
Lehigh Univeristy
Bethlehem, PA 18015
ericpsb@lehigh.edu

Andrea Forte
College of Computing & Informatics
Drexel University
Philadelphia, PA 19104
aforte@drexel.edu

Abstract
This paper addresses the so-called privacy paradox, an apparent disconnect between stated attitudes about privacy and reported or observed privacy-related actions. Rather than conceptualizing privacy in terms of degree of concern or depth of literacy, this paper suggests that we might benefit from thinking about data styles, different sets of techniques, strategies, and attitudes in handling personal data. With this approach, rather than asking, "does this person care about privacy?" we can instead ask, "what constellation of techniques do they adopt around the management of their personal data?" This position aligns with the sensibility of the Networked Privacy 2017 workshop by stepping away from a "more privacy is better" dictum.

Not only does this formulation undo the need to explain a privacy paradox, but it also allows us to understand more accurately the actual practices people employ in negotiating their privacy.

Keywords
Privacy; data; savvy; paradox; styles.

From Privacy to Paradox
Research on privacy faces an apparent contradiction. When asked directly, people often claim to care about privacy. However, people simultaneously behave in ways that suggest they are not concerned about privacy.

This phenomenon has been referred to as the privacy paradox [2]. In some ways, it emerges from Westin's categories of approaches to privacy. The privacy fundamentalists believe their individual privacy must be protected, full stop. For example, a privacy fundamentalist might forgo entirely the use of social networking sites due to the potential for personal information to be collected and used in undesirable ways [1,19]. The privacy pragmatists see the value of individual privacy but also see the value of, say, social networking sites. A privacy pragmatist will negotiate intermediate states between full rejection of potentially privacy violating tools and full disclosure of personal information [3].

These two categories seem reasonable enough. Westin's third category, though, raises some issues. He refers to these as the unconcerned, those who simply do not care about their own privacy. These individuals

"are generally trustful of organizations collecting their personal information [and] comfortable with existing organizational procedures" [8:5].

Interestingly, most of Westin's questions pertain to the use of personal information by healthcare providers, credit card companies, tax agencies, employers, and other corporate or government organizations. While the practices Westin includes raise their own issues, more problematic is the label assigned to this third category and the concomitant assertion that these individuals do not care about their own privacy. Virtually no one, however, self-identifies as unconcerned about privacy [2]. Thus the paradox. Or rather, the apparent paradox.

From Paradox to Practices

The concept of privacy has a complex and loaded meaning, both within academic discourse [13] and in popular media [15]. The latter reveals a normative value to caring about one's privacy. Thus, rather than asking about privacy directly, the most informative work along these lines asks about specific practices related to disclosure, boundary negotiation, etc.

For instance, one might explain the apparent paradox through feelings of apathy [7]. In such situations, people do in fact care about their privacy, and many of them take great pains, for instance, configuring online privacy settings and curating different lists of contacts – friends, family, coworkers, classmates – who do or do not see certain kinds of content. However, even these individuals do not feel able to maintain their privacy. "It is precisely because they understood [Facebook's privacy settings]—and how they interacted with their social groups—that they grasped the system's limitations" [7:3747]. From these feelings of disempowerment, apathy emerges.

An alternative resolution can be seen in Nissenbaum's notion of privacy as contextual integrity [12]. Rather than access or control, she argues, privacy should be seen as adherence to normative expectations about the ways and contexts in which information is used. For instance, one might share detailed personal information on social media with the expectation that such information will only be propagated via other similar contexts. Other uses (e.g., for advertising) represent a violation of contextual integrity, and thus a breach of privacy. This account obviates the need for a paradox to explain how public sharing of personal information and a concern for privacy can coexist.

In another example, boyd [4] describes an instance where a teenager posted the lyrics of "Always Look on the Bright Side of Life" from the Monty Python film The Life of Brian. Her mother, not familiar with the reference, soon comments with enthusiastic support. Her friends, who know that the song is sung while the main character is being crucified, immediately text to ask what is wrong. boyd calls this social steganography. Similar to political dog whistles [10] and other coded speech, use of this strategy allows publicly visible content to become decipherable only by an intended audience.

Similarly, contributors to open collaboration projects, such as Wikipedia, or open source software encounter privacy-related conflicts. Although they are motivated to contribute their expertise, they may also wish to conceal their identity; these desires can be at odds. For example, individuals may be hesitant to contribute information about socially sensitive or controversial topics despite personal expertise, or may avoid contributing information about places they live (and know well) because these contributions could compromise their privacy or attract trolls. Although the Wikipedia community embraces contributions from "anonymous" individuals, in the parlance of the project, this "anonymity" requires editors to publicly disclose an IP address, Furthermore, Wikipedia generally does not allow editing from Tor, which leads to a

variety of anonymity-seeking strategies that help editors conceal potentially identifying information from other project contributors, the general public, governments, and others [6].

These issues are made even more complex by the different kinds of entities to whom data might be disclosed. Put differently, the question of "privacy from whom?" becomes quite salient. For example, many people who give up use of social networking sites, such as Facebook, cite privacy as a concern [3,16,19]. However, the term "privacy" glosses over underlying variety. Some of these non-users are concerned about social surveillance [11]. They ask: What other people whom I know personally will be able to find out about me? Others, though, are more concerned about institutional surveillance [17]. They ask: What will corporations or governments be able to find out about me? The tactics deployed in response to these concerns can vary, as well. Those concerned about social surveillance will limit their use of the site, avoiding posting certain types of content, or selectively deactivate [3]. In contrast, those concerned about institutional surveillance are more likely to cease using the site entirely [3].

From Practices to Styles

Many of these practices sit uneasily within Westin's typology [8]. One might refer to all the cases above as instances of privacy pragmatism. However, doing so glosses over important distinctions, both in the different conceptualizations of personal data evinced in these practices, and in the different strategies for negotiating access to and uses of those data. Similar concerns arise if we attempt to account for this variety in terms of degrees of literacy. Compare the individual who selectively visits webpages with and without the cover of the Tor browser to the Facebook user who carefully curates lists of Friends. One cannot easily say which practice involves more or less privacy literacy.

Instead, this paper offers the notion of data styles. Rather than asking about "privacy" concerns, one could instead inquire about risks, strategies, and results. Concerns and risks people perceive when they use networked technologies could include, e.g., losing one's employment, being embarrassed, or being the target of hate speech. The strategies they use to mitigate those risks could include, e.g., installing certain types of software, judiciously using privacy settings, or finding alternative services or platforms. The results of deploying those strategies could include, e.g., reduction in awkward encounters, feelings of increased security, or possibly a failure to prevent unwanted sharing.

One could then identify a given privacy style by determining which sets of strategies and practices tend to show up together. These styles could be gleaned through inductive analysis of, say, qualitative interview transcripts. Styles could also arise from statistical analysis, e.g., principal components of quantitative survey data. Alternatively, styles might emerge from machine learning analysis of users' privacy preference settings [18].

This approach provides at least one important upshot. As with Westin's categories, many extant instruments place people on a continuum from low to high in terms of their concerns about privacy, their privacy literacy, etc. [14]. Even multidimensional scales are still scored in terms of low and high [20]. In contrast, identifying data styles allows one to acknowledge the potential viability in a wide diversity of practices for negotiating networked privacy. This point resonates directly with the workshop's theme, that privacy should not and cannot be thought about in a monotonic fashion that simply casts more privacy as better. Doing so can allow designers, policy makers, and others to shift from encouraging more privacy to supporting different privacy styles.

From Styles to Provocations

This paper presents work in progress. As such, it concludes not with results but rather with a series of questions that might serve both to aid in the development of this work and to provoke discussion during the workshop.

What are the different classes of privacy data-related practices about which one might ask? Certain practices might be relevant only to, say, social media, such as fine tuning of friend lists. Others might apply more broadly, such as familiarity with HTTPS or email encryption. How might one determine which classes of practices are more or less relevant for a given group of study participants?

What would be the benefits and detriments to asking about specific tools (TrackMeNot, Privacy Badger, etc.) vs. broader classes of technologies (ad blockers, private browsing modes, etc.)?

What could be gained (or lost) by including offline behaviors, such as signing up for magazines under an assumed name. Others will swap grocery store loyalty cards to obfuscate their purchasing history [5,9,21]. Such questions might not pertain directly to privacy in networked environments, but they could be highly informative about the broader styles that individuals take in relationship to their personal data.

References

1. Alessandro Acquisti and Ralph Gross. 2006. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In *Proc PETS*, 36–58.
2. Susan B. Barnes. 2006. A Privacy Paradox: Social Networking in the United States. *First Monday* 11, 9.
3. Eric P. S. Baumer, Phil Adams, Vera D. Khovanskaya, Tony C. Liao, Madeline E. Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, Leaving, and (Re)Lapsing: An Exploration of Facebook Non-use Practices and Experiences. In *Proc CHI*[, 3257–3266.
4. boyd, danah. (2014). *It's Complicated*. New Haven: Yale University Press.
5. Finn Brunton and Helen Nissenbaum. 2011. Vernacular Resistance to Data Collection and Analysis: A political theory of obfuscation. *First Monday* 16, 5.
6. Andrea Forte, Nazanin Andalibi, and Rachel Greenstadt. 2017. Privacy, Anonymity, and Perceived Risk in Open Collaboration: A Study of Tor Users and Wikipedians. In *Proc CSCW*.
7. Eszter Hargittai and Alice Marwick. 2016. "What Can I Really Do?" Explaining the Privacy Paradox with Online Apathy. *International Journal of Communication* 10.
8. Ponnurangam Kumaraguru and Lf Cranor. 2005. Privacy indexes: A survey of westin's studies. Carnegie Mellon University, Pittsburgh, PA.
9. Dan Lockton. 2006. The fight back: loyalty card subversion. *Architectures*. http://architectures.danlockton.co.uk/2006/12/10/the-fight-back-loyalty-card-subversion/
10. Ian Haney López. 2014. *Dog Whistle Politics: How Coded Racial Appeals Have Reinvented Racism and Wrecked the Middle Class*. Oxford University Press.
11. Alice E. Marwick. 2012. The Public Domain: Social surveillance in everyday life. *Surveillance and Society* 9, 4: 378–393.
12. Helen Nissenbaum. 2010. *Privacy in Context*. Stanford, CA: Stanford University Press.
13. Leysia Palen and Paul Dourish. 2003. Unpacking "privacy" for a networked world. In *Proc CHI*, 129–136.
14. Yong Jin Park. 2013. Digital Literacy and Privacy Behavior Online. *Communication Research* 40, 2: 215–236.
15. Laura Portwood-Stacer. 2013. Media refusal and conspicuous non-consumption: The performative and political dimensions of Facebook abstention. *New Media & Society* 15, 7: 1041–1057.

16. Lee Rainie, Aaron Smith, and Maeve Duggan. 2013. Coming and Going on Facebook. Pew Research Center, Pew Internet and American Life Project, Washington, D.C.

17. K. Raynes-Goldie. Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday* 15, 1–4.

18. Norman Sadeh, Jason Hong, Lorrie Cranor, Ian Fette, Patrick Kelley, Madhu Prabaker, Jinghai Rao. 2009. Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing* 13(6): 401-412.

19. Stefan Stieger, Christoph Burger, Manuel Bohn, and Martin Voracek. 2013. Who commits virtual identity suicide? Differences in privacy concerns, internet addiction, and personality between facebook users and quitters. *Cyberpsychology, Behavior, and Social Networking* 16, 9: 629–34.

20. Sabine Trepte, Doris Teutsch, Philipp K. Masur, Carolin Eicher, Mona Fischer, Alisa Hennhöfer, and Fabienne Lind. 2015. Do people know about privacy and data protection strategies? Towards the "Online Privacy Literacy Scale"(OPLIS). In *Reforming European data protection law*, Serge Gutwirth, Ronald Leenes and Paul de Hert (eds.). Springer, 333–365.

21. C. Walters. 2007. Should You Be Swapping Loyalty Cards? *Consumerist*. http://consumerist.com/2007/08/07/should-you-be-swapping-loyalty-cards/