#MyPrivacy: How Users Think About Social Media Privacy

Kelly Quinn

University of Illinois at Chicago
Department of Communication
Chicago, IL USA
kquinn8@uic.edu

Dmitry Epstein

University of Illinois at Chicago
Department of Communication
Chicago, IL USA
dmitry@uic.edu

ABSTRACT

This study explores privacy from the perspective of the user, leveraging a “framing in thought” approach to capture how users make sense of privacy in their social media use. Definitions of privacy collected from 608 social media users are analyzed through topic modelling and the clustering of word pairs to surface themes present in the data. Results indicate the dominance of frames related to horizontal privacy, or privacy vis-à-vis peers, over vertical privacy (i.e., that from institutions and governments). Themes relating to economic and legal frameworks had a reduced level of prominence. These findings suggest that user conceptualization of privacy reflects a cognate-based approach that emphasizes control and limits to information access.

CCS CONCEPTS

• Security and privacy → Social aspects of security and privacy • Human-centered computing → Social media • Human-centered computing → Empirical studies in collaborative and social computing

KEYWORDS

Privacy; user definitions; social media; framing

ACM Reference format:

Kelly Quinn and Dmitry Epstein. 2018. #MyPrivacy: How Users Think About Social Media Privacy. In *Proceedings of the International Conference on Social Media & Society*, Copenhagen, Denmark (SMSociety).¹ DOI:10.1145/3217804.3217945

1 INTRODUCTION

Privacy has become the de facto currency of the social media world. People routinely disclose information, which not too long ago was considered private, in exchange for digital tools and services. At the same time, the concept itself is surprisingly fluid [1], its interpretation and enactment are highly contextual [2], and these characteristics often do not align, resulting in the so-called “privacy

paradox” [3], [4]. While previous research has looked into the behavioral aspects of privacy [3], [5] or into the strategic deployment of its various meanings [6], there is limited inquiry into how social media users themselves perceive and interpret the idea of privacy. In this paper we use recently collected survey data in an attempt to unpack how users of social media frame privacy. Our goal is to help conceptualize social media privacy in a participant-centric way, thus further enhancing efforts to theorize privacy and design privacy-sensitive tools and solutions.

2 LITERATURE REVIEW

In our effort to map the frames of references to privacy used by social media users we draw on two main bodies of literature. First, we discuss framing literature, which offers an established conceptual framework for understanding the importance of definition for both behavioral and structural outcomes. Second, we discuss relevant privacy literature in order to situate our current work in earlier efforts to tease out the plurality and context-driven fluidity of privacy definitions.

2.1 Framing

Framing is a popular analytical framework, used in a variety of fields (e.g., psychology, political science, sociology, communication), across levels of analysis (e.g., individual, institutional), and with different epistemological approaches [7], [8]. Goffman has originally referred to frames as primary “schemata of interpretation” that allow “its user to locate, perceive, identify, and label a seemingly infinite number of concrete occurrences defined in its terms” and within her personal context [9]. Gamson and Modigliani explained that frames present “a central organizing idea [...] for making sense of relevant events, suggesting what is at issue” and thus giving meaning to issues and ideas [10].

Frames operate at both systemic (macro) and individual (micro) levels. At the macro level, competing social, political, and cultural actors engage in constructing, modifying, and disseminating frames

¹ Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

SMSociety '18, July 18–20, 2018, Copenhagen, Denmark

© 2018 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6334-1/18/07.

<https://doi.org/10.1145/3217804.3217945>

both intentionally and unintentionally. In doing so, the actors operate with frames in communication as they engage in competitive behavior around frames of reference that align with their values, goals, and interests [7], [8]. At the micro level, frames in thought delineate the realm of the possible or desirable by restricting and prioritizing a set of available considerations or by amending their subjective value and applicability in the eyes of an individual [8], [11]. In other words, at the fundamental level, “frames in communication” are strategically constructed and deployed, typically by powerful social actors, in an attempt to influence audiences to adapt particular interpretive “frames in thought” when making sense of complex issues [12].

The tension between frames in communication and frames in thought are at core of framing research. Gamson and Modigliani have demonstrated how the gap between these two types of frames affected policy discourse and public understanding around nuclear power, [10]. Entman has proposed a cascading activation model, where issue frame construction and diffusion are linked through mass media portrayals [13]. Framing is especially influential in areas lacking clear definitions, established frames of reference, or where mechanisms of influence are obscured from actors. Privacy is a prime example of one such area.

2.2 Privacy

Conceptualization of privacy is an ongoing debate among scholars. Westin, in one of the early, hierarchical models of privacy, viewed it as nested spaces of political, socio-cultural, and personal levels of analysis with each layer representing a different set of social structures that both constrain and enable privacy behaviors [14]. Nissenbaum has further developed this idea by referring to privacy as contextual [2]. Solove called for a pluralistic conceptualization of privacy as a “a set of protections against a plurality of distinct, but related problems” [15], thus making implicit privacy practices explicit. Such problems include issues of information collection, processing, dissemination, and invasion. Taken together, these conceptual frameworks lend themselves to speaking about a privacy space where discourse can be examined at the political, socio-cultural, or personal levels and be viewed as context-dependent [6].

Within the conceptual privacy space, one can identify a number of more concrete frames of reference to privacy. Smith, Dinev, and Xu roughly divide those approaches into value-based and cognate-based conceptualizations [5]. Central to the value-based conceptualizations is Warren and Brandies’ framing of privacy as a “right to be left alone” [16]. Critics claim, however, that this framing is rooted in the physical notion of privacy, which does not transfer well into the digital realm, where privacy is treated as a commodity [3], [17].

The cognate-based approach harbors the control and limited access paradigms of privacy. The control paradigm offers a broad

conceptual continuum of what privacy may mean. For some, control over one’s information equates with privacy itself. For others, it is viewed as a mediating factor in what constitutes “a dialectic and dynamic boundary regulation process” [18]. The limited access paradigm treats privacy just as a state of limited access to a person or her information [14], [19] and can also be viewed as a continuum, from absolute to minimal [5].

The rapid adoption of social media and the growing use of private information as de facto currency for digital (and increasingly physical) services, have pushed the conceptual work around privacy to new frontiers with an emphasis on contextual nature of privacy and the plurality of meaning placed in the idea of privacy by different actors. Some argue for privacy context collapse, where multiple audiences of an actor collapse into one [20], [21], which, in turn, requires a networked model of privacy determined through “constellation of audience dynamics, social norms, and technical functionality that affect the processes of information disclosure, concealment, obscurity, and interpretation within a networked public” [22]. Others suggest a distinction between vertical and horizontal privacy. The former relates to privacy vis-à-vis institutions (e.g., government, social media platform), while the latter refers to privacy vis-à-vis peers (e.g., co-workers, social media connections) [23]. These approaches imply a set of frames in thought about privacy. Our goal here is to address this question directly.

2.3 Framing privacy

The rich conceptual debate about privacy is a fertile ground for framing research. Somewhat surprisingly however, studies explicitly tackling the question of framing of privacy are scarce [24]. Existing research focuses primarily on frames in communication of elite actors such as policymakers, new media [6], [24], or technology designers, who often view emphasizing privacy as a barrier to adoption [25]. Fornaciari, attempted to examine frames in communication of non-elite actors. Studying privacy framing on Twitter she identified eight distinct frames ranging from privacy and technology being the most frequent to trading privacy being the least frequent frame [26].

Expanding this line of research, the current project asks to delve explicitly into frames in thought of non-elite actors. Hence, we pursue a rather direct research question: How do users of social media frame social media privacy? We explore this question with original survey data and by using topic modeling as our analytical approach. Given our explicit focus on social media, we expect the frames in thought to reify or challenge more recent conceptual developments such as a networked view of privacy or the vertical versus horizontal distinction.

Table 1: Topic identification using ConText

Topic	Topic Weight	Members
1	0.350	share, information, private, privacy, people, post, personal, thing, life, personal_information
2	0.269	friend, post, people, profile, platform_sponsor, website, family, privacy, view, public
3	0.235	information, people, privacy, platform_sponsor, access, personal_information, important, give, secure, account
4	0.149	privacy, platform_sponsor, setting, post, thing, respect, social_media, private, put, person
5	0.116	privacy, information, platform_sponsor, intrude, make, include, company, advertise, permit, personal

6	0.104	safe, protect, secure, hack, identity, personal_information, information, account, steal, feel
7	0.077	information, share, sell, company, consent, safe, address, contact, ability, person
8	0.041	information, address, photo, email, phone_number, show, secure, personal, give, number
9	0.016	business, stay, follow, follower, time, hide, people, boundary, knowledge, put
10	0.009	mine, photo, good, ability, private, word, opinion, work, privacy, limit

3 METHODOLOGY

3.1 Sample

Participants were recruited using the Qualtrics panel service, and received a small incentive by the platform sponsor. This study was approved by the Institutional Review Board at the University of Illinois at Chicago. Data were collected in a self-administered, web-based survey tool which included questions on attitudes and behaviors related to social media and privacy. Included in these was the question, “With respect to [most frequently used social media platform], what does privacy mean to you?” Participants were required to supply a definition consisting of a minimum of 135 characters. The resulting corpus included 608 individually generated definitions of privacy. The underlying sample was representative of the US population, based on 2010 US Census demographics, on characteristics of age, gender, and income. Mean age was 47.8 years (SD=16.7, range=18-90, Mdn=47.0) and gender was balanced (53.1% female, 46.2% male, .7% not reported). Racial/ethnic composition included: African-American 8.9% (n=54); Hispanic/Latino 7.6% (n=46); Asian 4.9% (n=30); Caucasian 77.0% (n=468); Multi-ethnic/Other/Undisclosed 1.2% (n=7). Participants in the study were actively engaged with social media, with 90.8% reported having two or more social media profiles and 81.1% reported accessing their favored social media site at least once/day.

3.2 Method

Topic modelling is a text-mining approach that uses statistical probabilities for discovering topics or themes in a collection of documents. It is based on the notion that documents are collections of topics that reflect a thematic structure which can be inferred by examining the probability distribution of words appearing together [27]. In each topic, different sets of terms have higher probabilities; topics can be visualized by listing these terms [28]. Topic modeling algorithms analyze the words of the unstructured texts to discover the themes that run through them and how those themes may be connected to each other [29]. In contrast to traditional content analysis, topic modelling utilizes computer algorithms to identify patterns of word co-occurrence, and thus is useful for analyzing large datasets. In addition, because it can be used without *a priori* coding structures, topic modelling offers the benefit of not requiring identification of themes in advance.

ConText [30] is an automated topic modelling tool for analyzing texts and networks that can be used to analyze the a large volume of texts. ConText leverages the “Machine Learning for Language Toolkit” [31] to perform topic modelling, which is based on the Latent Dirichlet Allocation (LDA) model [32]. LDA assumes documents are generated by drawing on fixed topic vocabularies that are composed of words with high probabilities; it then reverses this generative process to uncover the latent topics within the texts using probabilistic modelling [29].

3.3 Analysis

Initial cleaning of the data, preprocessing, was carried out in ConText. This included the removal of stopwords, or articles, prepositions, conjunctions, and transitive verbs that do not contribute to the meaning of the text (e.g., if, and, that, a, an, the, to, is, was, were). Stemming was also carried out in ConText, and adjusted tense and different forms of the same word into a unified morpheme. Finally, a codebook was applied to correct spelling, consolidate n-gram terms such as “social media,” and generalize specific categories of words, such as “Amazon” and “Facebook” into a single term of “platform_sponsor”. Topic modelling and network generation functions were then performed.

4 RESULTS

To illustrate the themes found within the privacy definitions, we specified 10 topics and 10 words per topic. Results are summarized in **Table 1**, with the words in each topic presented in order of occurrence within that topic. Topics are ordered by weight within the document corpus. As shown in **Table 1**, Topics 1 through 4 suggest themes that reflect horizontal privacy or boundary control mechanisms. They include words such as “share” and “people,” and also the types of information typically generated by social media use, such as “profiles” and “posts”. These topics have the highest weighting, which suggests their prominence in the corpus of definitions.

Themes 5 through 7 reflect concepts related to more vertical forms of privacy. Included in these topics are words relating to the use of personal information by others, including those such as “company,” “sell,” and “advertise,” as well as those related to the security of information such as “hack,” “steal,” and “secure”. Topic 7 specifically references terms that are consistent with an economic framing of privacy, though the weight of the topic suggests that it is not a major theme. Notably absent are terms related to privacy from a legal perspective, such as “right” or “law”. As indicated by the overall topic weights, these latter topics hold much less prominence in the corpus of definitions, indicating a substantially lower level of attention to conceptions of vertical privacy by the participants in this study. Topics 8 through 10 held little weight in the corpus, and were omitted from further analysis.

To provide additional clarity to the themes in the corpus of definitions, ConText was used to generate a list of co-occurring word-pairs from the definitions, using a word-distance window of seven words. These were used to map an undirected network using Gephi [33] consisting of 112 nodes and 2,946 edges. Using Gephi’s native community detection algorithm [34], three communities were detected and individually examined. To simplify, graphs were scaled to show nodes of 45 degrees or larger.

The first community (Figure 1) depicts word pairs indicative of horizontal privacy: types of information, such as “posts,” “photos,” and “profiles,” along with those whom such information might be shared with, such as “friends,” “family,” and even “strangers”. The second community (Figure 2) evidences word pairs indicative of vertical privacy awareness: “personal information,” “platform sponsors,” in addition to terms such as “secure,” “safe,” and “hack”. The final community (Figure 3) describes the types of information that may be vulnerable to exploitation, as well as of an economic framing of privacy: phone numbers, emails, sell, and company.

Of note, there is evidence in all of the clusters of cognate-based approaches to privacy. Words such as “access,” “control,” “block,” and “setting” suggest that, with respect to social media, users conceptualize privacy as a boundary-control process whereby limits on access to information is prioritized.

5 DISCUSSION

Vertical privacy, while not lost on users, seems to have lower levels of relevance across user definitions. This lack of attention to institutional privacy incursions may further reify existing power imbalances along the lines of what Braman describes as a “panspectron” society, a condition where information about an individual is collected continuously [35]. Obvious means to address this potential imbalance include for designers to make data collection more explicit and for policymakers to emphasize the importance of privacy education for users. Given the nature of work-in-progress, we acknowledge the need to further refine the methodology and to more closely interpret the findings. Limitations of this project include the limited scope of user-supplied definitions, and the varying level of detail among corpus definitions. As data collection for these definitions also included user reports of privacy activities, future work might examine these definitions in the context of reported behaviors.

REFERENCES

- [1] L. R. BeVier, "Information About Individuals in the Hands of Government: Some Reflections on Mechanisms for Privacy Protection," *William Mary Bill Rights J.*, vol. 4, no. 2, pp. 455–506, 1995.
- [2] H. Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press, 2010.
- [3] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science (80-.)*, vol. 347, no. 6221, pp. 509–515, 2015.
- [4] A. Acquisti and R. Gross, "Imagined Communities : Awareness, Information Sharing, and Privacy on the Facebook," in *Privacy Enhancing Technologies*, 2006, pp. 36–58.
- [5] H. J. Smith, T. Dinev, and H. Xu, "Information privacy research: An interdisciplinary review," *MIS Q.*, vol. 35, no. 4, pp. 989–

- 1016, 2011.
- [6] Dmitry Epstein, Merrill C. Roth, and Eric P.S. Baumer, "It's the Definition, Stupid! Framing of Online Privacy in the Internet Governance Forum Debates," *J. Inf. Policy*, vol. 4, no. 2014, p. 144, 2014.
- [7] D. Scheufele, "Framing as a theory of media effects," *J. Commun.*, vol. 49, no. 1, pp. 103–122, 1999.
- [8] D. Chong and J. N. Druckman, "Framing Theory," *Annu. Rev. Polit. Sci.*, vol. 10, no. 1, pp. 103–126, 2007.
- [9] E. Goffman, *Frame Analysis*. Cambridge, MA: Harvard University Press, 1974.
- [10] W. A. Gamson and A. Modigliani, "Media discourse and public opinion on nuclear power: A constructionist approach," *Am. J. Sociol.*, vol. 95, no. 1, pp. 1–37, Jul. 1989.
- [11] T. E. Nelson, Z. M. Oxley, and R. A. Clawson, "Toward a psychology of framing effects," *Polit. Behav.*, vol. 19, no. 3, pp. 221–246, 1997.
- [12] D. Epstein, E. C. Nisbet, and T. Gillespie, "Who's responsible for the digital divide? Public perceptions and policy implications," *Inf. Soc.*, vol. 27, no. 2, pp. 92–104, 2011.
- [13] R. M. Entman, *Projections of Power: Framing News, Public Opinion, and US Foreign Policy*. Chicago, IL: University of Chicago Press, 2004.
- [14] A. F. Westin, *Privacy and Freedom*. New York: Athenum, 1967.
- [15] D. J. Solove, *Understanding Privacy*. Cambridge, MA: Harvard University Press, 2008.
- [16] S. D. Warren and L. D. Brandeis, "The right to privacy," *Harv. Educ. Rev.*, vol. 4, no. 5, pp. 193–220, 1890.
- [17] Z. Papacharissi, "Privacy as a luxury commodity," *First Monday*, vol. 15, no. 8, 2010.
- [18] L. Palen and P. Dourish, "Unpacking 'privacy' for a networked world," *Proc. Conf. Hum. factors Comput. Syst. - CHI '03*, no. 5, p. 129, 2003.
- [19] G. S. Dhillon and T. T. Moores, "Internet Privacy," *Inf. Resour. Manag. J.*, vol. 14, no. 4, pp. 33–37, Oct. 2001.
- [20] A. E. Marwick and danah boyd, "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience," *New Media Soc.*, vol. 13, no. 1, pp. 114–133, Jul. 2011.
- [21] J. Vitak, "The Impact of Context Collapse and Privacy on Social Network Site Disclosures," *J. Broadcast. Electron. Media*, vol. 56, no. 4, pp. 451–470, Oct. 2012.
- [22] A. E. Marwick and danah boyd, "Networked privacy: How teenagers negotiate context in social media," *New Media Soc.*, vol. 16, no. 7, pp. 1051–1067, 2014.
- [23] M. Bartsch and T. Dienlin, "Control your Facebook: An analysis of online privacy literacy," *Comput. Human Behav.*, vol. 56, pp. 147–154, Mar. 2016.
- [24] F. Fornaciari, "Mapping the territories of privacy: Textual analysis of privacy frames in American mainstream news," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, pp. 1823–1832, 2014.
- [25] J. A. Obar and A. Oeldorf-Hirsch, "Clickwrap Impact : Quick-Join Options and Ignoring Privacy and Terms of Service Policies of Social Networking Services," in *#SMSociety17 Proceedings of the 8th International Conference on Social Media & Society*, 2017, p. Article 50.
- [26] F. Fornaciari, "iTweet about #privacy: Mapping privacy frames in Twitter conversation," in *ALLDATA 2017: The Third International Conference on Big Data, Small Data, Linked Data and Open Data*, 2017, pp. 70–73.
- [27] M. Steyvers and T. Griffiths, "Probabilistic topic models," in *Latent Semantic Analysis: A Road to Meaning*, T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, Eds. Hillsdale, NJ: Laurence Erlbaum, 2007.
- [28] D. M. Blei, "Topic modeling and digital humanities," *J. Digit. Humanit.*, vol. 2, no. 1, 2012.
- [29] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, p. 77, Apr. 2012.
- [30] J. Diesner, "ConText: Software for the integrated analysis of text data and network data," *Pap. Present. Soc. Semant. Networks Commun. Res. Preconference Conf. Int. Commun. Assoc.*, 2014.
- [31] A. K. McCallum, "MALLET: A machine learning for language toolkit." 2002.
- [32] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [33] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," *Third Int. AAAI Conf. Weblogs Soc. Media*, pp. 361–362, 2009.
- [34] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, vol. 2008, no. 10, p. P10008, Oct. 2008.
- [35] S. Braman, *Change of State: Information, Policy, and Power*. Cambridge, MA: MIT Press.