2015

Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding

Joel R. Reidenberg
Fordham University School of Law, JREIDENBERG@law.fordham.edu

Travis Breaux
Carnegie Mellon University, breaux@cs.cmu.edu

Lorrie F. Cranor
Carnegie Mellon University, lorrie@cmu.edu

Brian M. French
Carnegie Mellon University

ABSTRACT

Privacy policies are verbose, difficult to understand, take too long to read, and may be the least-read items on most websites even as users express growing concerns about information collection practices. For all their faults, though, privacy policies remain the single most important source of information for users to attempt to learn how companies collect, use, and share data. Likewise, these policies form the basis for the self-regulatory notice and choice framework that is designed and promoted as a replacement for regulation. The underlying value and legitimacy of notice and choice depends, however, on the ability of users to understand privacy policies.

This paper investigates the differences in interpretation among expert, knowledgeable, and typical users and explores whether those groups can understand the practices described in privacy policies at a level sufficient to support rational decision-making. The paper seeks to fill an important gap in the understanding of privacy policies through primary research on user interpretation and to inform the development of technologies combining natural language processing, machine learning and crowdsourcing for policy interpretation and summarization.

For this research, we recruited a group of law and public policy graduate students at Fordham University, Carnegie Mellon University, and the University of Pittsburgh ("knowledgeable users") and presented these law and policy researchers with a set of privacy policies from companies in the e-commerce and news & entertainment industries. We asked them nine basic questions about the policies' statements regarding data collection, data use, and retention. We then presented the same set of policies to a group of privacy experts and to a group of non-expert users.

The findings show areas of common understanding across all groups for certain data collection and deletion practices, but also demonstrate very important discrepancies in the interpretation of privacy policy language, particularly with respect to data sharing. The discordant interpretations arose both within groups and between the experts and the two other groups.

The presence of these significant discrepancies has critical implications. First, the common understandings of some attributes of described data practices mean that semi-automated extraction of meaning from website privacy policies may be able to assist typical users and improve the effectiveness of notice by conveying the true meaning to users. However, the disagreements among experts and disagreement between experts and the other groups reflect that ambiguous wording in typical privacy policies undermines the ability of privacy policies to effectively convey notice of data practices to the general public.

The results of this research will, consequently, have significant policy implications for the construction of the notice and choice framework and for the US reliance on this approach. The gap in interpretation indicates that privacy policies may be misleading the general public and that those policies could be considered legally unfair and deceptive. And, where websites are not effectively conveying privacy policies to consumers in a way that a "reasonable person" could, in fact, understand the policies, "notice and choice" fails as a framework. Such a failure has broad international implications since websites extend their reach beyond the United States.

TABLE OF CONTENTS

I. INTRODUCTION

Privacy policies are verbose, difficult to understand, take too long to read, and may be the least-read items on most websites even as users express growing concerns about information collection practices. But, for all their faults, privacy policies remain the single most important source of information for users to attempt to learn how companies collect, use, and share data. The reason that privacy policies are so important is that the United States takes a "notice and choice" approach to Internet privacy. The idea is that companies post their privacy policies, users read and understand policies, and users follow a rational decision-making process to engage with companies offering an acceptable level of privacy. This structure is designed and promoted as a replacement for regulation. The underlying value and legitimacy of notice and choice thus depends on the ability of users to understand privacy policies.

This paper investigates whether expert, knowledgeable, and typical users can understand the practices described in privacy policies at a level sufficient to support rational decision-making. The paper seeks to fill an important gap in the understanding of privacy policies through primary research on user interpretation and to inform the development of natural language processing and crowdsourcing for policy interpretation and summarization.[1] Part II of the paper discusses the existing landscape for notice and choice policies and the gaps in prior research on user understanding. Part III then defines the methodology for the research. Part IV presents the results and reports on discrepancies in the interpretation of the language in the privacy policies among three different groups: privacy experts, law and policy graduate students, and non-expert users. These results reveal significant discrepancies across the groups. Part V analyzes the critical implications of these discrepancies.

The results of this research will, consequently, have significant policy implications for the construction of the notice and choice framework and for the US reliance on this approach. The implications also expand beyond the United States since websites extend their reach globally.

[1] *See* Norman Sadeh, Alessandro Acquisti, Travis D. Breaux, Lorrie Faith Cranor, Aleecia M. McDonald, Joel Reidenberg, Noah A. Smith, Fei Liu, N. Cameron Russell, Florian Schaub, Shomir Wilson, James T. Graves, Pedro Giovanni Leon, Rohan Ramanath, Ashwini Rao, "Towards Usable Privacy Policies: Semi-automatically Extracting Data Practices From Websites' Privacy Policies (Poster)," ACM Symposium on Usable Security and Privacy (SOUPS 2014); N. Sadeh, A. Acquisti, T. Breaux, L. Cranor, A. McDonals, J. Reidenberg, N. Smith, F. Liu, C. Russel, F. Schaub, and S.Wilson, "The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi-Automatically Answer Those Privacy Questions Users Care About,". Carnegie Mellon University, School of Computer Science, Institute for Software Research Technical Report, CMU-ISR-13-119 (2013) http://reports-archive.adm.cs.cmu.edu/anon/isr2013/abstracts/13-119.html; Steve Bellovin and Sebastian Ziemeck, *Machine Learning Analysis of Privacy Policies*, [UNIV. OF MICHIGAN TELECOMM. L. REV.] (forthcoming); Sebastian Zimmeck and Steven M. Bellovin, "Privee: An Architecture for Automatically Analyzing Web Privacy Policies" In Proceedings of 23rd USENIX Security Smposium, August 2014, USENIX Association https://www.usenix.org/conference/usenixsecurity14/technicalsessions/presentation/zimmeck.

II. THE LANDSCAPE

This Part will first explain how and why notice and choice is used as a mechanism to address privacy protection. In the United States, notice and choice has become the principal means to address privacy online. While more extensive regulation exists in Europe,[2] notice and choice on an international scale plays important roles in the implementation of privacy rights and in the assurance of international data flows.

For notice and choice to work effectively, notice must be meaningful for users. This Part will also address prior research into the usability of privacy policies, describe usability problems, and, thus, reveal the gap to be filled by this research.

A. The Notice and Choice Framework

Since the 1970s, the United States promoted fair information practice standards as the guidepost for the protection of privacy.[3] These principles appear in US law, but the US legal system shies away from comprehensive privacy regulation.[4] Historically, the United States has addressed discrete privacy issues in narrow statutes targeted to specific problems and focused on specific actors.[5] Over the years, the White House, Congress, and the Federal Trade Commission have encouraged private sector responses to privacy challenges in lieu of new regulation.[6]

Notice and choice are the critical elements for self-regulation of fair information practices. "Notice" is generally described in terms of transparency of the information practices. The FTC has stated the principle as giving:

> "consumers . . . ... notice of an entity's information practices before any personal information is collected from them.... . . . . [N]otice of some or all of the following have been recognized as essential to ensuring that consumers are properly informed before divulging personal information:
>
> - identification of the entity collecting the data;
> - identification of the uses to which the data will be put;

---

[2] Directive 95/46/EC.

[3] *See* ROBERT GELLMAN, A SHORT HISTORY OF FIPS (2014).

[4] *See, e.g.,* PAUL M. SCHWARTZ & JOEL R. REIDENBERG, DATA PRIVACY LAW: A STUDY OF US DATA PROTECTION (1996).

[5] *See, e.g.,* Joel R. Reidenberg, *Privacy in the Information Economy: A Fortress or Frontier for Individual Rights*, 44 FED. COMM. L.J. 195 (1992); Schwartz & Reidenberg, *supra* note 4.

[6] *See, e.g.*, White House, Consumer Privacy Bill of Rights, http://www.whitehouse.gov/sites/default/files/privacy-final.pdf (Feb. 23, 2012) (voluntary approach); U.S. Privacy Protection Study Comm'n, Personal Privacy in an Information Society: Report to the President (July 1977); Federal Trade Commission, Privacy Online: A Report to Congress 7 (1998), http://www.ftc.gov/sites/default/files/documents/public_events/exploring-privacy-roundtable-series/priv-23a_0.pdf [hereinafter "Privacy Online"]; U.S. Dep't Comm., Privacy Self-Regulation in the Information Age (June 1997), http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf at viii.

- identification of any potential recipients of the data;
- the nature of the data collected and the means by which it is collected if not obvious (passively, by means of electronic monitoring, or actively, by asking the consumer to provide the information);
-whether the provision of the requested data is voluntary or required, and the consequences of a refusal to provide the requested information; and
-the steps taken by the data collector to ensure the confidentiality, integrity and quality of the data.[7]

Adequate and meaningful notice is necessary for users to be able to make informed decisions about their privacy choices.

"Choice" is typically defined in terms of consent. As the FTC articulates:

> "At its simplest, choice means giving consumers options as to how any personal information collected from them may be used. Specifically, choice relates to secondary uses of information — i.e., uses beyond those necessary to complete the contemplated transaction.[8]

Combined, notice and choice are used as a fundamental aspect of privacy protection in the private sector.

Internationally, notice and choice is also an important part of the international framework for transborder data flows. In 2000, the European Union and the United States adopted the Safe Harbor agreement to facilitate international data flows.[9] Under the voluntary agreement, US companies would agree to seven principles that were designed to assure the privacy of their EU origin data. The Safe Harbor agreement specifically included "notice" and "choice" as two essential principles.[10]

The "notice" principle required website operators to "inform individuals about the purposes for which it collects and uses information about them" in "clear and conspicuous language."[11] The "choice" principle added that those collecting personal information were required to "offer individuals the opportunity to choose (opt out [of]) whether their personal information [was] (a) to be disclosed to a third party or (b) to be used for a purpose that is incompatible with the purpose(s) for which it was originally collected."[12] Like the notice principle, "choice" demanded that companies construe their privacy agreements with clarity,

---

[7] *See* Privacy Online, *supra* note 6, at 7-8.
[8] Id, at 8.
[9] *See* U.S. Dep't Of Commerce, *Safe Harbor Privacy Principles* (Jul. 21, 2000), http://export.gov/safeharbor/eu/eg_main_018475.asp; Joel R. Reidenberg, E-Commerce and Trans-Atlantic Privacy, 38 Hous. L. Rev. 717 (2001)
[10] *Id.*
[11] *Id.*
[12] *Id.*

stating that "[i]ndividuals must be provided with clear and conspicuous, readily available, and affordable mechanisms to exercise choice."[13]

Similarly, the Article 29 Working Party of European data protection commissioners has looked to notice and choice in a number of initiatives to protect personal data. In 2013, for example, the Working Party released a guidance document for website operators on obtaining website user's consent for the use of tracking cookies.[14] The Working Party specified that to provide sufficient notice, websites must provide users with specific information about how and why they used cookies.[15] Attaining users "blanket consent" without first supplying exact facts would not suffice.[16] The Working Party suggested that website operators configure browsers so as to require users to actively signify their consent, so that there was no doubt of users' subjective intent.[17] Moreover, the Working Party emphasized that users be offered free choices regarding tracking cookies, and that users retain the option to browse a website while declining cookies.[18] The Italian Data Protection Authority internalized the Working Party's guidance in May of 2014.[19] Among its resolutions was that website banners should contain clear and visible notice and consent requests for users.[20]

Though many of the Working Party's initiatives have adopted the notice and choice principles, the data protection authorities also recognize the limitations of notice and choice. At the Safe Harbor Conference in 2009, Dutch Data Protection Authority chairman, Jacob Kohnstamm stated in his introductory remarks that enforcement tool (e.g. fines) may be a superior means of ensuring the protection of website users' personal data.[21] Kohnstamm stated that "[d]ue to new technological applications transparency alone (notice and choice) is no longer sufficient to guarantee that individuals can oversee the consequences of data processing activities…independent oversight is necessary. It is necessary to ensure a level playing field. To ensure that all are abiding to the same rules."[22]

B. Research on Usability and Technical Tools

Prior research has shown, however, that the terms contained in policies are frequently unfamiliar to users and the level of education necessary to understand the policies is high.[23]

---

[13] *Id.*
[14] *See Working Document Providing Guidance on Obtaining Consent for Cookies, the European Commission Article 29 Working Party* (adopted Oct. 2, 2013), http://ec.europa.eu/justice/data-protection/index_en.htm.
[15] *Id.* at 3.
[16] *Id.*
[17] *Id.*
[18] *Id.* at 5.
[19] *See* Simplified Arrangements to Provide Information and Obtain Consent Regarding Cookies, The Italian Data Protection Authority (May 8, 2014), http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3167654.
[20] *Id.* at 3.
[21] Kohnstamm, Jacob, Chairmna, Dutch Data Protection Authority, *Introductory Speech at the Safe Harbor Conference* (2009), *available at* http://www.dutchdpa.nl/downloads_int/20091118_speech_jko_washington.pdf.
[22] *Id.* at 8.
[23] *See* M. Hochhauser, Lost in the fine print: Readability of financial privacy notices, July 2001. http://www.privacyrights.org/ar/GLB-Reading.htm; M. A. Graber, D. M. D'Alessandro, and J. Johnson-West, Reading level of privacy policies on internet health web sites. Journal of Family Practice (July 2002).

Similarly, research has also shown that notice of privacy policies may not be effective and that some notices are designed to nudge users into disclosing larger quantities of personal information than necessary for the interaction.[24] Privacy technologists have developed tools to facilitate notice and choice for online users, but they have achieved only limited success.

1. Usability

Previous work has shown that while users have difficulty finding and using privacy policy information, they remain interested in this information and when this information is made salient it can impact users' online purchase decisions.[25] Research has also demonstrated that users are interested in several different pieces of information found in privacy policies.[26] These research results suggest that the information in privacy policies could be helpful if presented in a usable way.

Prior research also found that expecting users to read privacy policies places an unreasonably high burden on them because policies take so long to read.[27] As a result, there have been several approaches to improving usability. One approach to making privacy policies more accessible is a privacy "nutrition label" that summarizes key points from a privacy policy in a succinct and standard form. While this approach has shown promise in research studies, it has not yet been widely adopted.[28]

Layered privacy notices are another approach. Layered notices present a website's privacy policy to users in multiple "layers," with each describing elements of the policy in

---

http://www.jfponline.com/the-publication/past-issue-single-view/reading-level-of-privacy-policies-on-internet-health-web-sites/7aec24c0f23755624911626 35f50b929.html.

[24] *See* Wang Yang, Pedro Giovanni Leon, Xiaoxuan Chen, Saranga Komanduri, Gregory Norcie, Kevin Scott, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. "The Second Wave of Global Privacy Protection: From Facebook Regrets to Facebook Privacy Nudges." 74 Ohio St. L.J. 1307 (2013).

[25] J. Tsai, S. Egelman, L. Cranor, and A. Acquisti. The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. Information Systems Research (Feb. 2010) http://www.heinz.cmu.edu/~acquisti/papers/acquisti-onlinepurchasing-privacy.pdf. For the use of using default settings to prompt users to disclose/share personal information, see I., Dinner, E.J. Johnson, D.G. Goldstein, and K. Liu, Partitioning default effects: Why people choose not to choose. Journal of Experimental Psychology-Applied 17, 4, 332. (2011); D.G. Goldstein, E.J. Johnson, A. Herrmann, and M. Heitmann, Nudge your customers toward better choices. Harvard Business Review 86, 12, 99–105 (2008). For a discussion of changes to Facebook's interface that promote sharing, *see* F. Stutzman, R. Gross, and A. Acquisti, Silent listeners: The evolution of privacy and disclosureon facebook. Journal of Privacy and Confidentiality 4, 2, 2 (2013).

[26] *See* P.G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, L.F. Cranor. What Matters to Users? Factors that Affect Users' Willingness to Share Information with Online Advertisers. In Proceedings of the Eighth Symposium On Usable Privacy and Security (SOUPS '13), Newcastle, United Kingdom, 2013 http://cups.cs.cmu.edu/soups/2013/proceedings/a7_Leon.pdf ; J. Lin, B. Liu, N. Sadeh, and J.I. Hong, Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings , 2014 ACM Symposium on Usable Security and Privacy (SOUPS 2014), July 2014 (quantifying users's willingness to disclose/grant different mobile app privacy permissions – "different pieces of information found in privacy policies").

[27] McDonald, Aleecia M., and Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S J. L. & POL'Y 543 (2008).

[28] P.G. Kelley, L.J. Cesca, J. Bresee, and L.F. Cranor. Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach. CHI2010 http://www.cylab.cmu.edu/research/techreports/2009/tr_cylab09014.html

greater levels of detail and specificity.[29] Typically, these notices consist of a "short notice in a common template format" coupled with a "longer complete notice."[30]

Proponents of this approach argue that layered notices "easily build consumer trust" and "increase public understanding of privacy and data protection" because the notices are "easy to read and understand."[31] One study revealed, however, that though layered notices enabled study participants to make decisions more quickly, the participants often responded inaccurately to questions about terms they had read in the notices.[32] Furthermore, the results suggested that participants rarely probed beyond the initial layer, thus leaving them with "incorrect impressions" of the privacy practices endorsed by the more-complete policy.[33]

2. Technical Tools

Privacy technologists have also developed a variety of tools for users to express privacy preferences and for users to opt out of receiving targeted ads. However, the evaluation and deployment of these technologies over the years shows that the challenges to building effective tools have not been overcome. These challenges include the imposition of burdens on users to capture complex and diverse privacy preferences[34] and general usability features such as mechanisms to opt out in the context of online behavioral advertising.[35]

---

[29] *See* Center for Information Policy Leadership, *Ten steps to develop a multilayered privacy notice* at 1, *available at* www.informationpolicycentre.com/files/Uploads/Documents/Centre/Ten_Steps_whitepaper.pdf.

[30] *Id.* This guide proposes that notices contain three layers:

> Layer 1 – The short notice: the very minimum, for example, when space is very limited, providing only the identity of the data controller, contact details, and the purposes of processing.
>
> Layer 2 – The condensed notice: covering the basics in less than a page, ideally using subheadings, and covering Scope; Personal information collected; Uses and sharing; Choices (including any access options); Important information; How to contact us.
>
> Layer 3 – The full notice.

Id.

[31] *Id.* at 2, 3.

[32] *See generally* Aleecia M, McDonald et al., *A Comparative Study of Online Privacy Policies and Formats*, *available at* http://robreeder.com/pubs/PETS2009.pdf.

[33] *Id.* at 15.

[34] *See e.g.* M. Benisch, P.G. Kelley, N. Sadeh,and L.F. Cranor, "Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs", Journal of Personal and Ubiquitous Computing. 15 :7 (Oct. 2011) available at http://www.normsadeh.com/file_download/142

[35] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor, "Why Johnny Can't Opt Out: a Usability Evaluation of Tools to Limit Online Behavioral Advertising," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012, http://dl.acm.org/citation.cfm?doid=2207676.2207759; Pedro Giovanni Leon, Justin Cranshaw, Lorrie Faith Cranor, Jim Graves, Manoj Hastak, Blase Ur, and Guzi Xu. 2012. What do online behavioral advertising privacy disclosures communicate to users?. In Proceedings of the 2012 ACM workshop on Privacy in the electronic society (WPES '12). ACM, New York, NY, USA, 19-30. DOI=10.1145/2381966.2381970 http://doi.acm.org/10.1145/2381966.2381970.

a. P3P

Growing concern by Congress and threats from the FTC to regulate online privacy gave rise to the Platform for Privacy Preferences (P3P)– a web standard that enables web browsers to read website privacy policies automatically and compare them with user-specified privacy preferences.[36] Essentially, P3P would enable users to avoid websites whose practices did not meet their privacy preferences.[37] P3P specification 1.0 was launched in 2002.[38] Though a more-developed specification 1.1 working draft was later produced, it was never finalized, as the P3P working group "closed . . . due to lack of industry participation" in 2006.[39]

While some popular web browsers have integrated P3P tools,[40] others have not.[41] Furthermore, the users of browsers that have integrated P3P are reportedly unaware of the tool.[42] In addition, thousands of websites that adopted P3P appear to have used P3P codes to circumvent browser cookie blocking, without making accurate computer-readable statements about their privacy policies.[43] Thus, P3P policies have become an unreliable source of privacy policy information.

b. Do Not Track

In 2007, privacy advocates began discussing the creation of a mechanism that would enable users to register their opposition to being tracked online. The mechanism would be similar to the "Do Not Call" list for opting out of telemarketing solicitations.[44] Over time, this idea developed into a technical mechanism that would allow user agents – including web browsers, cell phones, email clients, and anti-malware packages – to send a "do not track" signal

---

[36] *See* Lorrie Faith Cranor, *Necessary But Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice*, 10 J. TELECOMM. & HIGH TECH. L. 273, 279 (2012).

[37] *Id. See also* Kimberly Rose Goldberg, Note, *Platform for Privacy Preferences ("P3P"): Finding Consumer Assent to Electronic Privacy Policies*, 14 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 255, 2003.

[38] *See* Cranor, *supra* note 36.

[39] Cranor, *supra* note 36, at 280.

[40] Namely, Microsoft Internet Explorer 6, 7, 8, and 9. *See* Cranor, *supra* note 36, at 280.

[41] Neither Firefox, Safari, nor Chrome have integrated P3P, though "a number of prototype plug-ins and extensions," "authoring tools," and "prototype P3P user agents" have been developed. *See* Cranor, *supra* note 36, at 280-81.

[42] Cranor asserts that "While I know of no formal studies, my informal polls of hundreds of audience members at talks I have given suggests that outside of groups of privacy experts, almost nobody has heard of P3P . . . ." Cranor, *supra* note 1, at fn. 38.

[43] Pedro Giovanni et al., *Token Attempt: The Misrepresentation of Website Privacy Policies through the Misuse of P3P Compact Policy Tokens 4*, Workshop on Privacy in the Electronic Society (WPES) (Oct. 2010), *available at* http://www.cylab.cmu.edu/ files/pdfs/tech_reports/CMUCyLab10014.pdf.

[44] Soghoian, C. The History of the Do Not Track Header. (January 21, 2011), http://paranoia.dubfire.net/2011/01/history-of-do-not-track-header.html.

on the user's behalf.[45] In 2010, a Federal Trade Commission report requested comments on the idea of Do Not Track (DNT).[46]

DNT similarly became a popular topic with legislators. Multiple bills at both the state and federal levels would have made DNT a legal requirement, but only one passed into law: California's AB 370.[47] Under AB 370, companies with customers who are California citizens must disclose how, if at all, they respond to an incoming DNT request. In practice, this disclosure requirement is a de facto national (and international) standard, since most English language websites will likely have at least one visitor from California.[48]

By 2012, all major web browsers had implemented an interface for users to send a DNT request. However, the implementation of DNT remains elusive. While DNT is a promising idea, there are three major barriers to wide adoption. First, there is no agreement on the treatment of a DNT request by a website, i.e., as to how the website should respond.[49] Second, only a few prominent companies such as Mozilla, Twitter, and AP News have publicly encouraged DNT and the list of implementers is quite modest.[50] Lastly, if there were a new standard that called for all companies to perform a minimum set of actions upon receipt of a DNT signal, companies might simply refuse to allow access to users requesting DNT.

3. Research On Automated Understanding of Privacy Policies

Researchers have also considered whether automated processing of privacy policies will be able to provide users with meaningful information for notice and choice.[51] One recent study

---

[45] Mayer, J. and Narayanan, A. Do Not Track: Universal Web Tracking Opt Out <donottrack.us>; Mayer, J., Narayanan, A. and Stamm, S. Do Not Track: A Universal Third-Party Web Tracking Opt Out (March 7, 2011) http://tools.ietf.org/html/draft-mayer-do-not-track-00.

[46] Fed. Trade Comm'n, *A Preliminary FTC Staff Report on Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers* (December 1, 2010), http://www.ftc.gov/os/2010/12/101201privacyreport.pdf.

[47] Assembly Bill No. 370; CHAPTER 390: An act to amend Section 22575 of the Business and Professions Code, relating to consumers, *available at* http://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140AB370.

[48] Compliance with AB 370 seems to circumvent the purpose of DNT. An informal survey shows that most of the companies complying with AB 370 offer a vague statement saying that they ignore DNT. These notices are usually contained somewhere in the privacy policy or in a file linked from the privacy policy.

[49] The World Wide Web Consortium (W3C) successfully published a late-stage draft of the technical mechanisms to send and receive DNT signals. *See* Tracking Preference Expression (DNT) W3C Last Call Working Draft, http://www.w3.org/TR/tracking-dnt/.

[50] Mozilla published an implementation guide with example source code. *See* Mozilla, *The Do Not Track Field Guide*. But, there is no consensus on the treatment of the signal. Several companies have announced they honor DNT. *See* Do Not Track Us. Do Not Track: Implementations, http://donottrack.us/implementations.

[51] See N. Sadeh, A. Acquisti, T. Breaux, L. Cranor, A. McDonals, J. Reidenberg, N. Smith, F. Liu, C. Russel, F. Schaub, and S.Wilson (2013). "The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi-Automatically Answer Those Privacy Questions Users Care About,". Carnegie Mellon University, School of Computer Science, Institute for Software Research Technical Report, CMU-ISR-13-119 (2013) http://reports-archive.adm.cs.cmu.edu/anon/isr2013/abstracts/13-119.html; Steve Bellovin and Sebastian Ziemeck, *Machine Learning Analysis of Privacy Policies*, [UNIV. OF MICHIGAN TELECOMM. L. REV.] (forthcoming); Sebastian Zimmeck and Steven M. Bellovin, *Privee: An Architecture for Automatically Analyzing Web Privacy Policies*, In Proceedings of 23rd USENIX Security Smposium, August 2014, USENIX Association, https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/zimmeck.

explored the possibility of using automated processing and crowdsourcing to interpret website privacy policies.[52] The study relied on data provided by ToSDR.org, a crowdsourcing project that examined a limited set of privacy policies and that does not use a scientifically-based rating approach for those policies. The study did, however, find inherent limitations due to "ambiguity of language" and variant human interpretations.

Other studies further investigated the feasibility of leveraging natural language processing and machine learning techniques to tackle the problems of automatic categorization of privacy policies[53] and grouping segments of policies based on the privacy issues they address.[54] These studies shed light on automatic methods of understanding privacy policies; however, it is not clear if the existing natural language techniques are able to fully decode the sophistication and ambiguity of privacy policies. A more promising approach will likely involve combining such techniques with machine learning and crowdsourcing, hence the importance of this study.

Another study examined the manual translation of privacy policies into a specialized mathematical logic.[55] The study results include heuristics for mapping variant interpretations into a single, canonical representation expressed in logic and a demonstration of how this logical representation can be used to answer questions about information collection, use and sharing. For example, one heuristic includes mapping certain verbs, such as "transfer," "share," and "access" to events in which a data holder shares personal information with a third party.[56] In particular, the verb "access" is ambiguous because it can map to collection, use, or sharing depending on the stakeholder viewpoint, i.e., who has access. Other ambiguities, such as omissions, generic terms, and terms that have varying technical interpretations can lead to variant interpretations, some of which may be unintended by the policy authors. While the formalization does enable automated reasoning to detect policy conflicts due to ambiguous policy statements, it does require special training to perform the translation into logic. As with any policy document, the logical representation must also be maintained as the natural language policy changes. This prior work leaves open the question of how automated processing and crowdsourcing might function on an enormously broad set of privacy policies with a systematic approach to rating those policies. While preliminary results[57] are promising, it is not clear from this prior work whether a level of automation can be reached that would enable the process to be conducted on a web scale.

---

[52] *Id.*

[53] Waleed Ammar, Shomir Wilson, Norman Sadeh, Noah A. Smith. *Automatic Categorization of Privacy Policies: A Pilot Study*. Carnegie Mellon University, School of Computer Science, Technical Report, CMU-ISR-12-114, CMU-LTI-12-019, December 2012.

[54] Fei Liu, Rohan Ramanath, Norman Sadeh, Noah A. Smith. *A Step Towards Usable Privacy Policy: Automatic Alignment of Privacy Statements*. Proceedings of the International Conference on Computational Linguistics (COLING 2014), Dublin, Ireland, August 2014; Rohan Ramanath, Fei Liu, Norman Sadeh, Noah A. Smith. *Unsupervised Alignment of Privacy Policies using Hidden Markov Models*. Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2014), Baltimore, MD, June 2014.

[55] Travis Breaux, Hanan Hibshi, Ashwini Rao, *Eddy, A Formal Language for Specifying and Analyzing Data Flow Specifications for Conflicting Privacy Requirements.* Requirements Engineering Journal, December 2013.

[56] Id.

[57] Travis Breaux, Florian Schaub, *Scaling Requirements Extraction to the Crowd: Experiments on Privacy Policies.* 22nd IEEE International Requirements Engineering Conference (forthcoming).

4. Unanswered Questions for Automated and Crowdsourced Understanding

In light of the present state of research, this study tests the comprehension and clarity of privacy notices on a larger scale, with the aim of making a prognosis about the viability of large-scale semi-automated, analysis and review of privacy policies. Prior work shows that policy ambiguity may challenge the ability of natural language processing to be effective. Crowdsourcing may not fully remedy these ambiguities, but might help overcome some of these limitations depending on how far the interpretation of non-ambiguous elements might be scaled (e.g., does it always require expert annotators and could one leverage the so-called "wisdom of the crowds"?) Accordingly, this study is designed to explore the clarity of privacy policies in more detail by examining how three groups with different levels of expertise understand privacy notices. The goal is to elicit commonalities and differences in the comprehension and interpretation of websites' privacy policies across groups of participants varying in legal background and training.

III. METHODOLOGY

The research methodology was designed to discover how three different user groups would each interpret specific language in privacy policies. As discussed below, the participant groups were chosen to reflect expert, knowledgeable, and general users. Privacy policies were systematically collected from the web, and a survey was created to probe user understanding of the policies. In addition, background information was collected from the survey respondents.

A. The Participant Groups

Three groups participated in this study: 1) crowd workers representing general users; 2) knowledgeable users; and 3) privacy policy experts. These groups were recruited as follows:

1) *Crowd workers* were recruited on Amazon Mechanical Turk (MTurk) as a representative sample of the general population.[58] Previous studies have shown that MTurk provides a suitable participant pool for conducting research studies and that the demographic distribution of crowd workers on MTurk is comparable to the general US population.[59] These workers were paid $6.00 per reviewed policy. To be eligible to participate, these workers were required to have at least a 95% approval rating for 500 completed tasks on MTurk and be US residents.[60] US residency was verified with a question asking about the worker's country of residence. Multiple screening checks were applied in order to determine whether a crowd worker made an honest effort in completing the task. This vetting consisted of checking for the duration spent on the task,

[58] MTurk is an Amazon website that pays users to complete proposed tasks.

[59] *See* T. S. Behrend, D. J. Sharek, A. W. Meade, Eric N. Wiebe, "The Viability of Crowdsourcing for Survey Research." Behavior Research Methods 43 (3): 800–813. 2011; G. Paolacci, J. Chandler, P. G. Ipeirotis, "Running Experiments on Amazon Mechanical Turk." Judgement and Decision Making 5 (5): 411–19. 2010; and G. Paolacci, J. Chandler, "Inside the Turk: Understanding Mechanical Turk as a Participant Pool." Current Directions in Psychological Science 23 (3): 184–88. 2014.

[60] The MTurk rating level was set to assure that workers would take the task seriously and the U.S. residency requirement was set to assure that workers would not assume rights that exist in foreign countries.

whether question responses were accompanied by meaningful text selections (see below), and whether the participant provided actual words for the answers to a cloze test. All crowd worker submissions satisfied these checks, likely because the required qualification (95% approval rating on 500 tasks) and the relatively high pay ($6.00) were sufficient to motivate honest participation in the study.

2) *Knowledgeable users* consisted of five graduate students with a background in law, public policy, or computer science who were recruited from Fordham University, Carnegie Mellon University, and the University of Pittsburgh. These five knowledgeable users were hired as research assistants.

3) *Privacy policy experts* consisted of four of the study authors who are experienced law and public policy scholars. The purpose of these expert annotations was to determine the degree of agreement between experts, as well as to investigate the deviation of professional interpretation from the interpretation by knowledgeable users and crowd workers.

B. Privacy Policy Data Set

We collected 1,010 unique privacy policies from the top websites ranked by Alexa.com. These policies were collected during a period of six weeks during December 2013 and January 2014. They provide a snapshot of privacy policies from mainstream websites covering fifteen of Alexa.com's seventeen website categories.[61] The fifteen categories are listed below:

Business	Computers	Games	Health	Home
News	Recreation	Shopping	Arts	Kids and Teens
Reference	Regional	Science	Society	Sports

Locating a website's policy is not a trivial task. Though many well-regulated commercial websites provide a "privacy" link on their homepages, not all do. Neither is there standardized URL format for privacy policies. Even once the policy's URL is identified, extracting the policy text presents the usual challenges associated with scraping documents from the web. Since every site is different in its placement of the document (e.g., buried deep within the website, distributed across several pages, or mingled together with Terms of Service) and format (e.g., HTML, PDF, etc.), and since we aimed to preserve as much document structure as possible (e.g., section labels), full automation was not a viable solution.

Therefore, we crowdsourced the privacy policy document collection using MTurk. For each website, we created a "human intelligence task" or HIT in which a worker was asked to copy and paste the following privacy policy-related information into text boxes: (i) privacy policy URL; (ii) last updated date (or effective date) of the current privacy policy; (iii) privacy

[61] Of the seventeen categories, two were excluded: the "Adult" and the "World" category. The "world" category was excluded since it contained mainly popular websites in different languages, and we opted to focus on policies in English in this study.

policy full text; and (iv) the section subtitles in the top-most layer of the privacy policy. To identify the privacy policy URL, workers were encouraged to go to the website and search for the privacy link. Alternatively, they could form a search query using the website name and "privacy policy" (e.g., "Amazon.com privacy policy") and search in the returned results for the most appropriate privacy policy URL. Each HIT was completed by three workers who were paid $0.05. per HIT. The collected privacy policies were further validated through manual review by one of the authors to ensure quality annotations.

After excluding duplicates, the dataset contained 1,010 unique documents. It is of note that different websites may be covered by the same privacy policy provided by the parent company. For example, espn.go.com, abc.go.com, and marvel.com are all covered under the Walt Disney privacy policy.

In an earlier exploratory study, fifteen websites were selected from each of the "news" and "shopping" categories and used for initial crowdsourcing analysis. The websites were selected in a top-down fashion using the rankings provided by Alexa.com. Additionally, two websites (amazon.com, yahoo.com) were set aside as a development data set and used for testing the crowdsourcing interface.

In this study, we focus on U.S commercial websites. From the policy data set, three privacy policies were manually selected from the "news" category and three policies were selected from the "shopping" category. The selected privacy policies are listed below along with the date of their last revision at the moment of collection:

News sites:

ABC News: http://abcnews.go.com/ (December 30, 2013)
Washington Post: http://www.washingtonpost.com/ (November 15, 2011)
Weather Underground: http://www.wunderground.com/ (October 30, 2013)

Shopping sites:

Barnes and Noble: http://www.barnesandnoble.com (May 7, 2013)
Lowe's: http://www.lowes.com (April 25, 2013)
Overstock: http://www.overstock.com (January 9, 2013)

C. Privacy Policy Survey and Annotations

The study focused on three key privacy policy elements: the collection of information, sharing of information, and deletion of information. These were chosen to reflect important user concerns and were selected based on an analysis of FTC privacy enforcement actions, which identified surreptitious collection, unauthorized disclosure, and wrongful retention of personal information as the most significantly contested online information practices.[62] The study asked

[62] *See* Joel R. Reidenberg, N. Cameron Russell, Alexander Callen, and Sophia Qasir, *Privacy Enforcement Actions* (Fordham CLIP: 2014), http://law.fordham.edu/assets/CLIP/CLIP_Privacy_Case_Report_-_FINAL.pdf [hereinafter

about four information types that have been shown to be highly relevant to users in previous studies.[63] These information types were: contact information, financial information, current location information, and health information.

To discover commonalities and differences in interpretation between our different participant groups, we created a survey for participants that asked nine questions about different data practices described in a website's privacy policy. These questions (four collection questions, four sharing questions, one deletion question) are described below.

Each study participant was asked to answer the set of survey questions for each of the respective policies. For each answer, the participant was asked to select the text from the policy sections corresponding to the chosen answer. Each of the experts annotated the same set of six privacy policies specified above. These six policies were the only policies considered among those surveyed for the other participants.[64] The annotation process was completed using an online tool created for the task. Participants would select sentences and text passages in the policy with the mouse and then add those passages into a text field under the question by clicking a button. Participants could add one or multiple policy statements for their answers. All answer responses other than the *not applicable* response option required the selection of at least one accompanying text segment.[65]

The annotation tool and wording of questions and response options were refined over multiple iterations of pilot testing and an exploratory experiment. The experiment consisted of six participants (law and computer science graduate students) who each annotated fifteen policies and provided feedback in semi-structured interviews.

"Privacy Enforcement Actions"]. A fourth aspect – inadequate security for personal information – was not considered in this study, because privacy policies often contain only vague statements on security measurements.

[63] *See* Ackerman, Mark S., Lorrie Faith Cranor, and Joseph Reagle. 1999. "Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences." In Proceedings of the 1st ACM Conference on Electronic Commerce (EC '99), p. 1–8, New York, NY, USA: ACM ; A. N. Joinson, U.-D. Reips, T. Buchanan, and C. B. P. Schofield. Privacy, trust, and self-disclosure online. Human–Computer Interaction, 25(1):1–24, 2010 ; C. E. Wills and M. Zeljkovic. A personalized approach to web privacy: awareness, attitudes and actions. Info. Mgmt. & Comp. Security, 19(1):53–73, 2011; P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L. F. Cranor. What matters to users?: factors that affect users' willingness to share information with online advertisers. In Proc. SOUPS '13, page 7. ACM, 2013.

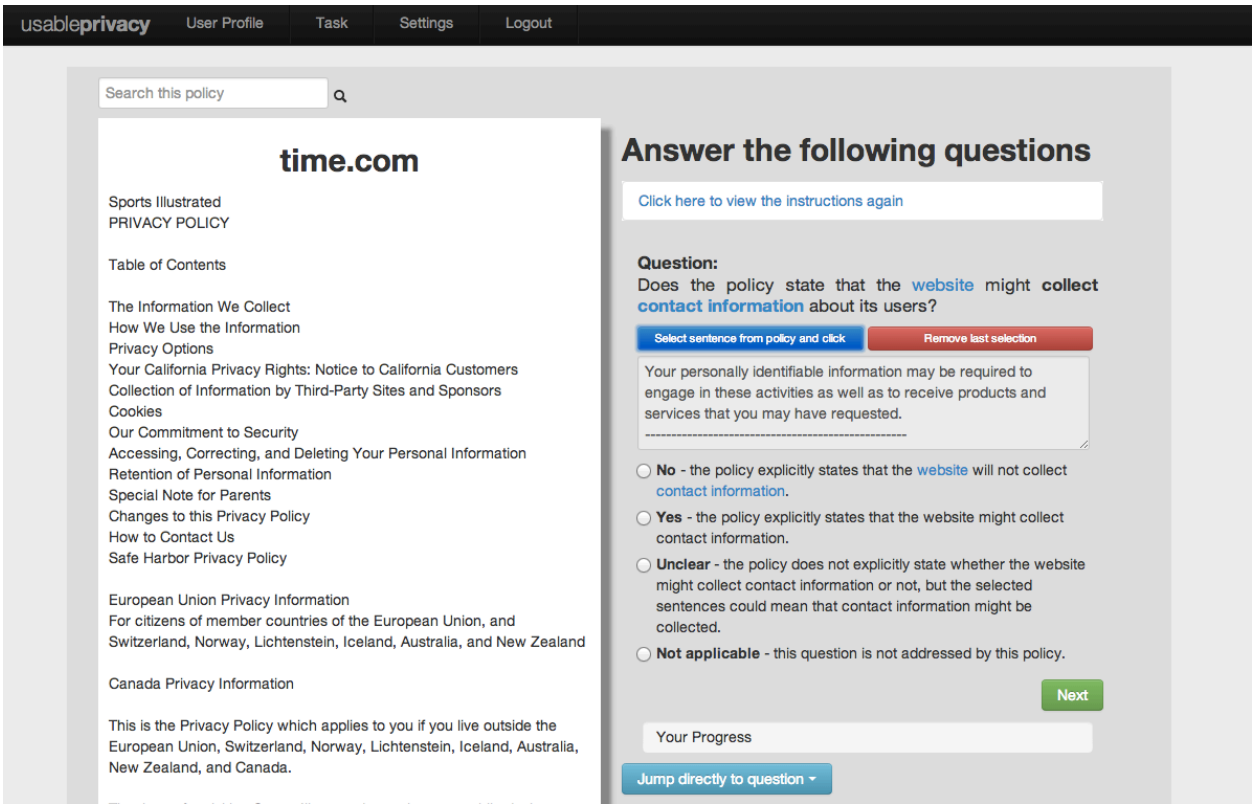
[64] MTurk crowd workers could choose to annotate only one policy or multiple policies, each compensated separately. Crowdsourcing tasks were created in such a way that we obtained at least five annotations from different crowd workers per policy. The majority of MTurk crowd workers chose to only annotate a single policy. Each knowledgeable user annotated 26 privacy policies, in total. We further conducted semi-structured interviews with all five knowledgeable users to gain deeper insights into their annotation strategies and their interpretation of our elicitation questions and policy statements.

[65] The online annotation tool further provided participants with detailed instructions on how to complete the annotation task. Participants were instructed to answer questions only for the company's main website and ignore privacy policy statements pertaining to other aspects of a company's business, such as mobile applications, physical stores, or other websites operated by the same company. Participants were further asked to ignore statements pertaining to a specific subset of users, such as statements addressing California privacy laws, EU Safe Harbor regulation, or COPPA. The instructions further clarified that the most fitting option should be selected based on the information given in the shown privacy policy and that "unclear" should be selected if multiple options would seem to apply, statements are ambiguous or contradicting, or if access to additional linked policies (i.e., separate cookie policy) would likely be required to answer a question conclusively. We also provided definitions for common terms in the questions and response options (see blue highlights in Figure 1) as further clarifications.

The final version of the online tool is illustrated in Figure 1 below. The scrollable privacy policy is displayed on the left site of the screen, and one question is shown at a time in a sidebar on the right. Participants could either progress through the questions sequentially or jump between questions in order to enable participants to quickly translate discovered policy statements into responses to our questions.

Figure 1

Online tool for privacy policy annotations.



Users could select policy statements with the mouse to provide evidence for their response. Hovering over defined terms highlighted in blue would display a tooltip with a definition of the respective term.

The survey questions on collection of personal information (Q1–Q4) inquired whether contact information, financial information, current location information, or health information is being collected by the given website. Participants could choose between four answer options:

> **No** – the policy explicitly states that the website will not collect [specified type of information (i.e., contact, financial, etc.)].
> **Yes** – the policy explicitly states that the website might collect [specified type of information (i.e., contact, financial, etc.)].
> **Unclear** – the policy does not explicitly state whether the website might collect [specified type of information (i.e., contact, financial, etc.)] or not, but the selected

sentences could mean that [specified type of information (i.e., contact, financial, etc.)] might be collected.
Not applicable – this question is not addressed by this policy.

While the *Yes* and *No* options capture explicit statements in the policy, the *Unclear* option enabled participants to note ambiguity in the policy regarding the collection of a specific information type. The *Not applicable* option, on the other hand, allowed for distinguishing between a policy containing ambiguous statements or no statement at all.

The questions on sharing of personal information (Q5–Q8) inquired whether a website would share contact information, financial information, current location information, or health information with third parties. If the policy stated that personal information would be shared with third parties, participants could indicate whether the information would be shared for the purpose of fulfilling a core service (e.g., payment processing or delivery of purchased goods), for purposes other than core services, or for purposes other than core services but only with explicit consent. Hence, the sharing of personal information questions offered six response options:

No sharing – the policy explicitly states that the website will not share [specified type of information (i.e., contact, financial, etc.)] with third parties.
Sharing for core service only – the policy explicitly states that the website might share [specified type of information (i.e., contact, financial, etc.)] with third parties, but only for the purpose of providing a core service, either with explicit or implied consent/permission from the user.
Sharing for other purpose – the policy explicitly states that the website might share [specified type of information (i.e., contact, financial, etc.)] with third parties for other purposes. The policy makes no statement about the user's consent/permission or user consent is implied.
Sharing for other purpose (explicit consent) – the policy explicitly states that the website might share [specified type of information (i.e., contact, financial, etc.)] with third parties for a purpose that is not a core service, but only if the user provided explicit permission/consent to do so.
Unclear – the policy does not explicitly state whether the website might share [specified type of information (i.e., contact, financial, etc.)] with third parties or not, but the selected sentences could mean that *contact information* might be shared with third parties.
Not applicable – this question is not addressed by this policy.

The question on deletion of personal information (Q9) asked about the website's respective deletion policy statements. We explicitly excluded any statements concerning retention for legal purposes, as we sought to assess policy statements that relate to issues of wrongful retention of personal information.[66] If the policy explicitly stated that information could be removed, participants could indicate whether information would be removed fully or whether some or all of the information may be retained for other purposes:

No removal – the policy explicitly states that the user will not be allowed to delete their personal data.

[66] *See* Privacy Enforcement Actions, *supra* note 62.

Full removal – the policy explicitly states that users may delete their personal data and that no data will be retained for any purpose, whether the data was provided directly by the user, generated by the user's activities on the website, or acquired from third parties.
Partial removal – the policy explicitly states that users may delete their personal data but some/all of the data might be retained for other purposes, whether the data was provided directly by the user, generated by the user's activities on the website or acquired from third-parties.
Unclear – the policy does not explicitly state whether users may delete their personal data or not (e.g., it only talks about editing information).
Not addressed – this question is not addressed by this policy.

After answering all nine annotation questions for a given privacy policy, participants were shown an additional screen with three questions asking them whether they had ignored parts of the privacy policy because they did not refer to the company's main website (yes/no), whether the privacy policy contained pointers or links to other policy documents (yes/no), and to rate "[…] how easy or difficult it was to answer the previous nine questions for this privacy policy" on a 5-point Likert scale (from "very difficult" to "very easy").

D. Background Demographics

Additionally, after completing the privacy policy annotations, participants were further asked to complete a background questionnaire, which consisted of multiple parts. First, participants were asked to rate their ability to understand legal texts on a five-point scale (from "very difficult" to "very easy"). This self-assessment was followed by questions about their level of received legal training and whether they worked in a position that required legal expertise. The second part collected basic demographic information, namely, gender, education level, and primary occupation. In the third part, participants were presented with a cloze test – a test that requires participants to replace several missing words in a piece of text – to assess their general reading comprehension.[67] The background questionnaire closed with a number of questions about the annotation experience. Participants were asked to rate the perceived ease or difficulty of answering each of the nine annotation questions on a seven-point scale (from "very difficult" to "very easy") and the helpfulness of provided instructions and terms definitions on a seven-point scale (from "not at all helpful" to "very helpful"). Lastly, an open-ended question allowed participants to further comment on difficulties with terms, questions, or answer options.

In terms of gender and age, the thirty-one crowd workers who annotated the six relevant privacy policies were 58% male (18) and 42% female (13). They ranged in age from 22 to 63 (with a median age of 29). Our group of five knowledgeable users were 40% male (2) and 60% female (3). Their ages were slightly younger (23 to 35 years of age with a median age of 24). The four privacy policy experts were 75% male (3) and 25% female (1). They were slightly older in comparison (34 to 53 years with a median age of 42).

With respect to education, all four experts have a graduate degree, and all five knowledgeable users have at least a bachelors degree (one has a graduate degree). The crowd

[67] The employed cloze test was taken from page 14 of the University of Cambridge's Handbook for Teachers on Cambridge English Proficiency: Certificate of Proficiency in English (CPE), CEFR Level C2 (2013).

workers were less educated, with only 42% having a bachelors degree or higher (4 high school graduates, 11 some college without degree, 3 associate/2-year degrees, 8 bachelors/4-year degree, 5 graduate degrees).

Primary occupations of the crowd workers were diverse, including administrative support (7), service industry (5), unemployed (4), student (3), art/writing/journalism (3), business/management/financial (3), education (2), and other (4). None of the crowd workers selected "legal (e.g., lawyer, law clerk)" as a primary occupation. Twenty-one crowd workers indicated they had no legal training at all. Seven indicated that they had no legal training, but that their background in another field provides them with some legal experience. Finally, three indicated that they were knowledgeable in legal matters, but had no formal training. Privacy policy experts were all researchers and scholars; two of them studied law. Knowledgeable users were all current students of law, computer science, or public policy.

IV. DATA COMPARISONS

This part reports on the data collected as a result of the methodology described above. The part will first analyze the empirical data on intra-group annotator agreement. Then it will analyze the empirical data on inter-group annotator agreement. As will be shown, high levels of agreement within a group ('intra-group" agreement) does not guarantee that this group converged on the same answers as other groups ("inter-group" agreement). The part will conclude with a discussion of qualitative trends observed in the data.

A. Intra-Group Annotator Agreement

The degree of agreement within each user group is shown in Tables 1-3.[68] Table 1 shows intra-group agreement for all questions and over the 6 policies annotated by the experts group for data collection, while Tables 2 and 3 show intra-group agreement for data sharing and deletion respectively. Intra-group agreement is measured by the median level of group member agreement on the same answer across all the policies.[69] First, the most frequently chosen answer (mode) for each question in each policy was identified. Then, to determine the level of agreement with the mode answer, the percentage of annotators selecting that mode answer was calculated for each question in each policy. The median level of agreement across all policies for the same question was calculated to reflect the group consensus on an answer choice. A median value of 100% means that all group members agree on the same answer choice for the survey question for at least 4 out of the 6 policies and share the same understanding of those privacy policies. As the median agreement declines, the group members understand some of the policies differently from each other. Any value less than 100% reflects that the annotator group had no consensus answer on the same question for 3 or more policies. The lower the value, the greater the disagreement. The tables also reflect the median level of agreement when related

[68] The tables do not distinguish between the website policies of news and shopping sites. The agreement rate among annotators was almost identical between the news and shopping categories for knowledgeable users. Crowd workers had similar agreement rates to the knowledgeable users on news sites, but slightly less agreement on the policies for shopping sites.

[69] Because of the small number of annotators, mean and standard deviation calculations would not provide an accurate representation of group member consensus.

answer choices were combined.[70] By combining answer choices, we can determine if there is at least a consensus on the way the policy broadly treats information. When answer choices are combined, the table shows the median level of agreement for each answer combination. Differences in the mode answer choice across the policies are reflected by median calculation based on those policies with the same mode.[71] The difference between the median level of agreement across all answer choices and the median agreement when several answer choices are combined will reflect that group members recognize that the policy addresses a particular point, but do not share the same understanding of the nuances.

Table 1

Data Collection: Intragroup Agreement

Level of Agreement on the Same Answer (Median Across All Policies)	Collect Contact	Collect Financial	Collect Location	Collect Health
Experts				
All choices	100 %	87.5 %	100 %	50 %
Answer Choice 1	n/a	n/a	n/a	n/a
Answer Choice 2	100 %	100 %	100 %	n/a
Answer Choice 3-4	n/a	75 %	n/a	100 %
Knowledgeable Users				
Using All Answers	100 %	100 %	70 %	80 %
Answer Choice 1	n/a	n/a	n/a	n/a
Answer Choice 2	100 %	100 %	90 %	n/a
Answer Choice 3-4	n/a	90 %	100 %	100 %
Crowd Workers				
Using All Answers	90 %	50 %	90 %	70 %
Answer Choice 1	n/a	n/a	n/a	n/a
Answer Choice 2	90 %	100 %	90 %	n/a
Answer Choice 3-4	n/a	60 %	n/a	100 %

Complete agreement by all members within each of the groups was uncommon. As shown in the tables, agreement was not distributed evenly across questions. For data collection, experts had a consensus on contact and location information (100% median level of agreement

[70] In Table 1, we combined the answer options "unclear" and "not applicable" as they were not differentiated consistently by all annotators. In Table 2, we combined answer choices 2-4 ("Sharing for core service only," "Sharing for other purpose," and "Sharing for other purpose (explicit consent)"), as all three of them describe that sharing with third parties is taking place, but differentiate between consent models; we further combined answer choice 5-6 ("Unclear" and "Not applicable") for the same reasons as above. In Table 3, we combined answer choices 2-3 ("full removal" and "partial removal") as they describe that removal is possible but vary in whether data is retained; as well as answer choices 4-5 ("Unclear" and "Not applicable") for the same reasons as above.

[71] This means the median score for a given answer choice combination may be based on fewer than 6 policies, and is independent from the median calculations of the other answer choice combinations.

across all policies), but varied in their understanding of the collection of financial and health information (87.5% and 50% respectively). Knowledgeable users shared the same understanding on the collection of contact and financial information (100% media level of agreement across all policies), but not location (70%) or health information (80%). Interestingly, the knowledgeable users had greater agreement on health information than the experts. This is likely to indicate that knowledgeable users missed important ambiguity in the privacy policy. Lastly, crowd workers had the lowest level of shared understanding compared to the other groups for contact information (90%) and location information (90%) than financial information (50%) and health information (70%). The crowd workers had a higher level of agreement on location information than the knowledgeable users, though less than the experts.

Table 2

Data Sharing: Intragroup Agreement

Level of Agreement on the Same Answer (Median Across All Policies)	**Share Contact**	**Share Financial**	**Share Location**	**Share Health**
Experts				
All Choices	75 %	62 .5 %	62 .5 %	50 %
Answer Choice 1	n/a	n/a	n/a	n/a
Answer Choice 2-4	87 .5 %	87 .5 %	75 %	n/a
Answer Choice 5-6	n/a	n/a	75 %	75 %
Knowledgeable Users				
All Choices	60 %	50 %	60 %	80 %
Answer Choice 1				
Answer Choice 2-4	80 %	80 %	100 %	n/a
Answer Choice 5-6	60 %	60 %	100 %	100 %
Crowd Workers				
All Choices	60 %	60 %	40 %	80 %
Answer Choice 1	n/a	n/a	n/a	n/a
Answer Choice 2-4	100 %	60 %	60 %	n/a
Answer Choice 5-6	n/a	100 %	70 %	100 %

For data sharing, the experts did not have a high level of consensus on the meaning of the privacy policies (ranging from 75% median level of agreement on contact information across all policies to 50% on health data). However, the level of agreement improves when the disclosure choices are aggregated (i.e., when answer choices 2-4 are collapsed into one). This means that the experts recognize, though never unanimously, that sharing occurs, but disagree as to the conditions for sharing as set out in the privacy policies. The knowledgeable users had weak agreement on their interpretations of sharing for contact information (60% median level of agreement across all policies), financial information (50%), and location information (60%). Oddly, the knowledgeable users level of agreement for health information across all policies was greater than that for the experts (80% median level vs. 50% median level). This means that the knowledgeable users did not perceive as much ambiguity as the experts and suggests that the

knowledgeable users may have misunderstood the sharing terms of the privacy policies for health information. The crowd workers similarly had weak agreement on the interpretation of policy statements on sharing and, similarly, had a greater consensus on the sharing of health information as compared to the experts.

Table 3

Data Deletion: Intragroup Agreement

	Level of Agreement on the Same Answer (Median Across All Policies)
Experts	
All Choices	75 %
Answer Choice 1	n/a
Answer Choice 2-3	100 %
Answer Choice 4-5	n/a
Knowledgeable Users	
Using All Answers	60 %
Answer Choice 1	n/a
Answer Choice 2-3	100 %
Answer Choice 4-5	n/a
Crowd Workers	
Using All Answers	50 %
Answer Choice 1	n/a
Answer Choice 2-3	100 %
Answer Choice 4-5	n/a

For deletion, the experts were not in complete agreement on the specific terms for data deletion (75% median level of agreement across all policies). They did, however, agree (100%) on whether some deletion options were available when the full and partial deletion options were aggregated. Knowledgeable users had less agreement than the experts on the terms of deletion policies (60%), but like the experts, when full and partial deletion choices were combined, the knowledgeable users had a complete consensus that at least some deletion was possible (100%). Crowd workers had the least agreement on data deletion (50%), but also reached complete consensus on policies that allowed at least partial deletion (100%).

B. Inter-Group Annotator Agreement

The shared understandings across user groups are shown in Tables 4-6. Table 4 shows the inter-group agreement levels for data collection. Table 5 reflects the level of inter-group agreement on data sharing and Table 6 presents the interpretation of data deletion. The tables were calculated using combined answer choices to determine whether there were shared

understandings of the broad practices.[72] Inter-group agreement is measured across all the policies by comparing the answer choice of the knowledgeable users and Crowd workers with the mode answer choices of the experts. First, to establish the control values, the experts' most frequently chosen answer on each question for each policy was identified. The level of agreement among the experts selecting that mode answer was then calculated and the median level of agreement across all 6 policies became the control value for the comparisons. A value of 100% means that all experts agreed on the same answer choice on the survey question for at least 4 of the 6 policies. As the value declines, the level of disagreement on the correct answer choice increases. For the knowledgeable users and crowd workers, the level of agreement with the experts' mode answer was calculated for each question and for each policy. The median levels of agreement for the knowledgeable users and crowd workers on the experts' mode answer across all 6 policies shows the alignment between these respective groups and the experts. For the knowledgeable users and crowd workers, a value of 100% on a particular question means that all group members agreed on the experts' mode answer choice and shared the same understanding of the privacy policy as the experts for at least 4 of the 6 policies.[73] As the value declines, the group members understand the policies differently from the experts.

Table 4

Data Collection: Intergroup Agreement

	Collect Contact	**Collect Financial**	**Collect Location**	**Collect Health**
Experts Selecting Mode Answer Choice (Median Across All Policies)	100 %	87 .5 %	100 %	100 %
Knowledgeable Users Selecting Experts' Mode Answer Choice (Median Across All Policies)	100 %	100 %	70 %	100 %
Crowd Workers Selecting Experts' Mode Answer Choice (Median Across All Policies)	90 %	40 %	90 %	100 %

For data collection, the three groups shared the same understanding of the collection of health information (100% agreement with experts' answers). With respect to contact information, the knowledgeable users matched the experts (100%), but the crowd workers lagged in their comprehension (90%). The intra-group disagreements, though, on health information

[72] In Table 4, we combined the answer options "unclear" and "not applicable" as they were not differentiated consistently by all annotators. In Table 5, we combined answer choices 2-4 ("Sharing for core service only," "Sharing for other purpose," and "Sharing for other purpose (explicit consent)") as all three of them describe that sharing with third parties is taking place, but differentiate between consent models; we further combined answer choice 5-6 ("Unclear" and "Not applicable") for the same reasons as above. In Table 6, we combined answer choices 2-3 ("full removal" and "partial removal") as they describe that removal is possible but vary in whether data is retained; as well as answer choices 4-5 ("Unclear" and "Not applicable") for the same reasons as above.

[73] Agreement between knowledgeable users and crowd workers does not necessarily mean the agreed-upon answer is objectively correct. To find out whether these two groups were likely to choose correct answers, we compared their answers to the answers given by the experts on the same policies.

reflect that ambiguities in the policies are being interpreted in the same way. The knowledgeable users had a consensus on financial information that matched the experts' most frequently chosen answer choice (100%). Since the experts were not unanimous on the answer choices with respect to financial information, the knowledgeable users may not have been cognizant of the ambiguities seen by the experts. The crowd workers, however, were far off from the experts (40% median level of agreement with the experts' mode answer choice across all policies). This indicates that typical users have a much more difficult time understanding the collection of financial information. Lastly, the knowledgeable users had a significantly different interpretation of the collection of location information than the experts (70% median level of agreement with the experts' mode answer choice across all policies). The Crowd workers were, however, in closer agreement with the expert answer choices (90%).

Table 5

Data Sharing: Intergroup Agreement

	Sharing Contact	**Sharing Financial**	**Sharing Location**	**Sharing Health**
Experts Selecting Mode Answer Choice (Median Across All Policies)	87 .5 %	87 .5 %	75 %	75 %
Knowledgeable Users Selecting Experts' Mode Answer Choice (Median Across All Policies)	80 %	50 %	60 %	100 %
Crowd Workers Selecting Experts' Mode Answer (Median Across All Policies)	100 %	40 %	60 %	100 %

For data sharing, the experts had no complete agreement on answer choices for any of the types of data. However, with respect to contact information, the Crowd workers had agreement with the most frequently chosen expert answer (100% median level of agreement with experts), while the knowledgeable users lagged in their understanding (80%). This may indicate that the knowledgeable users reflected the difficulties that the experts had with policy ambiguity. For financial information, the knowledgeable users had only a modest level of agreement with the experts (50% median level of agreement), while the crowd workers were even more divergent (40%). For location information, the experts' agreement on an answer choice was modest (75% median level of agreement on the mode). Both the knowledgeable users and the crowd workers had a significantly different understanding (60% median level of agreement with experts' mode answer choice). Lastly, for health information, the knowledgeable users and crowd workers converged on the most frequently chosen expert answer (100% median level of agreement). However, since the experts were divided on the answer choice, this finding suggests that the other groups may have missed nuances in the privacy policies with respect to potential sharing of health information.

Table 6

Data Deletion: Intergroup Agreement

Experts Selecting Mode Answer Choice (Median Across All Policies)	100 %
Knowledgeable Users Selecting Experts' Mode Answer Choice (Median Across All Policies)	100 %
Crowd Workers Selecting Experts' Mode Answer (Median Across All Policies)	100 %

For data deletion, all groups had agreement on the understanding of the privacy policies' language (100% median level of agreement). However, since the answer choices with respect to full and partial deletion were collapsed into one, this level of agreement may only reflect that all groups had the same understanding of the existence of, but not the conditions for, data deletion. The intra-group deviation on data deletion means that the policies have important ambiguity on the terms for data deletion.

C. Qualitative Data Analysis

1. Difficulty

For each policy, we asked annotators to rate on a 5-point Likert scale the difficulty of annotating that policy, where 1 was "very difficult" and 5 was "very easy." Averaged over the six policies, knowledgeable users rated the policies as easier to annotate on average (3.23) than the crowd workers (2.53). Interestingly, the experts found the policies also more difficult to annotate (2.88) than the knowledgeable users. Participants' self-reported ratings concerning their understanding of legal texts varied. The knowledgeable users rated their ability to understand legal texts higher on average (3.8) than the untrained crowd workers (2.29) or the experts (3.5). Looking at the average number of correct answers achieved in the general reading comprehension cloze test, the experts and knowledgeable users performed on a similar level. Both groups exhibit a median score of 7 correct answers (out of 8), with slight variations in the distributions (experts: mean=6.75, std.dev.=1.09; knowledgeable users: mean=6.6, std.dev.=0.49). The crowd workers exhibit lower reading comprehension with a median score of 5 correct answers (mean=4.77, std.dev.=2.03).

These results suggest that crowd workers may be hampered by less-developed reading comprehension skills in general and, according to their self-assessment, more specifically for legal text. This aspect matches the higher perceived difficulty of answering the annotation questions. Because the knowledgeable users and experts exhibit similar general reading comprehension skills, and assuming that experts have more experience in interpreting policy and legal text, the knowledgeable users may have either overestimated their abilities or experts may have been more cautious in their self-assessment and difficulty ratings.

Table 7 below further shows the individual average difficulty ratings of the different groups for each of the six policies. One set of privacy policies (Lowe's, ABC News, and Washington Post) were perceived as considerably more difficult to annotate by the untrained crowd workers as compared to the other two groups. The Barnes and Noble policy was perceived as difficult to annotate by all three groups. For the Overstock and Weather Underground policies, the difficulty ratings of the experts and untrained crowd workers are quite similar, whereas the knowledgeable users rated them as easier in both cases. Also note that none of the average values reach the "easy" level (4.0).

Table 7

Average difficulty rating of answering the annotation questions for the given policies.

Policy	Experts (avg. ease)	Knowledgeable users (avg. ease)	Crowd workers (avg. ease)
Overstock	3.00	3.40	3.00
Lowe's	3.00	3.00	1.80
Barnes and Noble	2.25	2.60	2.40
ABC News	3.50	3.60	3.00
Weather Underground	2.50	3.20	2.80
Washington Post	3.00	3.60	2.20

2. Trends in Selected Text[74]

a. Consensus on Text Selection

In some instances, the experts achieved either exact or near-exact agreement on the language selected as well as the answer option. These instances of agreement are promising, as they suggest that some scenarios might be used to train NLP techniques and thus might be able to provide reliable interpretations of privacy policies. In particular, they suggest that crowdsourcing tools might be developed where only (or mostly) relevant text is shown to crowd workers, rather than tools where a crowd worker is required to read an entire policy to answer a particular question.[75] Where the experts achieved exact agreement, they selected the same words or phrases and the same answer option.[76] Where the experts achieved near-exact agreement, it was typically the case that one annotator selected extra, immaterial words that another did not select, even though the experts selected the same answer option.[77]

[74] This section presents trends based on text selected by the experts and the knowledgeable users.
[75] *See* Rohan Ramanath, Florian Schaub, Shomir Wilson, Fei Liu, Norman Sadeh, Noah A. Smith; Identifying Relevant Text Fragments to Help Crowdsource Privacy Policy Annotations, Conference on Human Computation & Crowdsourcing (HCOMP '14), work in progress, 2014.
[76] *See, e.g.*, barnesandnoble.com Question 1, where all annotators selected the following: "we may collect personal information from you, for example your name, e-mail address, billing address, shipping address, phone number."
[77] *See, e.g.*, abcnews.go.com Question 7, where the annotators selected from the following range of text: "When you allow us to share your personal information with another company, such as: Electing to share your personal information with carefully selected companies so that they can send you offers and promotions about their products and services." Here,for example, two annotators selected the darkly shaded language, where only one selected the more lightly shaded language as well.

Likewise, knowledgeable users were able to achieve exact or near-exact agreement on the language selections as well as the answer option.[78] As was apparent among experts, it was also typical for one or more knowledgeable users to select additional, immaterial words that others may have not selected, despite the fact that all annotators chose the same answer option.[79] Notably, there were also instances in which annotators would select the same text in support of different answers. This conflict was evident between both expert annotator groups and knowledgeable annotator groups. Ultimately, these occurrences testify to the difficulty of gathering uniform interpretations of privacy policy language, and present a potential challenge to the development of an NLP model that can provide reliable interpretations of privacy policies.[80] At the same time, these issues suggest that it might also be possible to build interfaces that boost the productivity of crowd workers by selectively displaying text fragments that are mostly relevant to the particular questions they are requested to annotate.

b. Interpretation of Policy Silence

Experts differed from knowledgeable users with respect to interpretation of policy silence. Generally, where a policy was silent on a particular practice, experts interpreted such silence to mean that the policy permitted the practice. This is the legal interpretation. Knowledgeable users, on the other hand, often misinterpreted silence to mean that the policy was unclear regarding the practice. Examples of this interpretive difference in the context of data collection, data sharing, and data deletion are described below.

1. Data Collection

One of the best examples of the interpretive difference in the data collection context can be found in Question 2 ("Does the policy state that the website might collect financial information?") of the wunderground.com policy, which contains no explicit mention of

[78] *See, e.g.*, washingtonpost.com Question 1. As commonly seen, sentence selection was identical among all five knowledgeable users. To support their answers, every annotator selected the exact same text, which stated that "Washingtonpost.com asks for information such as your name, e-mail address, year of birth, gender, Zip code, country, street address."

[79] *See, e.g.*, loews.com Question 1. All knowledgeable users selected the following range of text: "You may choose to provide us with personal information (such as name, contact details and payment information), such as: Contact information, such as your name, address, telephone number, and email address, and your title or occupation."

[80] Both expert and knowledgeable users tended to select the same text to support different answers when asked if a website permitted users to delete personal data. *See, e.g.*, washingtonpost.com Question 9, in which all experts selected the statement that "[i]f you do not wish to receive the foregoing and therefore unregister from the site, please contact Customer Care and ask to have your registration account deleted." While one expert annotator believed that this statement indicated that users could partially remove their data, the remaining three experts selected this statement in support of the answer that it was unclear whether or not users could delete personal data. With regard to deletion questions, even more frequently than experts, knowledgeable users tended to select the same sentences to support different answers. *See, e.g.*, www.abcnews.go.com Question 9. When responding to this question, all annotators selected the same sentences which stated that "[y]ou may correct, update and delete your registration account" and "[y]ou may request access to the personal information we hold about you and that we amend or delete it and we request third parties with whom we have shared the information do the same." Though they all selected the identical text, two annotators believed that these statements indicated "full removal" of personal data, while two other annotators believed that they indicated "partial removal", and the remaining annotator believed that they indicated that it was "unclear" if users could or could not delete their data.

"financial information."[81] In responding to that question, three out of four experts selected sentences that they interpreted to permit the website to collect financial information;[82] only one expert believed that the policy was unclear on the practice. However, knowledgeable users provided almost exactly opposite responses: one knowledgeable user selected sentences that led him or her to think the website could collect financial information,[83] while three knowledgeable users selected "unclear" and one selected "not applicable."

2. Data Sharing

One of the best illustrations of the interpretive difference in the context of data sharing can be found in Question 7 ("Does the policy state that the website might share location information?") of the barnesandnoble.com policy. The policy contains no explicit mention of sharing location information.[84] In responding to that question, three out of four experts selected sentences that they interpreted to permit the website to share location information;[85] only one expert believed that the policy was unclear on the practice. On the other hand, no knowledgeable

---

[81] *See* wunderground.com Privacy Statement.

[82] *See* wunderground.com (Expert) Question 2. Some relevant textual selections chosen by the three experts who said the policy collected financial information are:

- "We use information collected on the Services . . . to help fulfill your requests or in connection with the operation of the Services, for example to . . . display information and advertisements that we believe match your interests and profile . . . ."
- "When your information is collected on the Services, it may be collected directly by or shared with a selected third parties in connection with the operation of the Services or the provision of services to you ("Third Party Processors"). These Third Party Processors may include, for example, companies that operate our online WunderStore, that process credit card information or handle shipping for Weather Underground . . . ."

Id.

[83] *See* wunderground.com (Knowledgeable) Question 2. The one annotator selected the following text: "These Third Party Processors may include, for example, companies that operate our online WunderStore, that process credit card information or handle shipping for Weather Underground, that deliver materials to you via e-mail or postal service, that organize, administer, process, or provide advertising services, and/ or that analyze data on our behalf to help us provide more relevant offers to you and to eliminate the delivery of duplicate offers as well as correcting and/or updating users' data based on information we provide them." *Id.*

[84] *See* barnesandnonle.com Privacy Statement.

[85] *See* barnesandnonle.com (Expert) Question 7. Two annotators selected Option 3 (Sharing for other purpose) and one chose Option 4 (sharing for other purpose with explicit consent). Some relevant textual selections are:

- "These or related applications may also allow you to provide information directly to social networking sites including information about your purchases, physical location . . . ."
- "Like many online retailers, we and/or our third party providers use cookies to recognize you as you use or return to the Barnes & Noble Websites."
- "In addition, to provide location-based services on Devices or through Apps, Barnes & Noble and third party application providers may automatically collect real-time geographic location information or other location-based information about you . . . ."
- ". . . your data may be transferred to or shared with a third party as part of a sale, merger, or acquisition of Barnes & Noble or one of its affiliates."
- "We provide personal information to our partners that provide product and service offerings or technologies that we think may be of interest to you."
- "In connection with purchases of certain Digital Content, we may need to forward information about you (including, for example, your Internet Protocol (IP) address) to the Digital Content provider in order to enable you to download or purchase Digital Content from or through that provider."

Id.

annotator selected sentences that led them to believe that the website may share users' location information. Instead, two annotators thought the policy was "unclear" on this matter,[86] and three selected "not applicable" to reflect the belief that the policy did not appear to address the question.

3. Data Deletion

There were fewer interpretive differences in the context of data deletion. The annotators frequently agreed on whether a website permitted users to delete personal data.[87] However, annotators often indicated uncertainty in their responses. They would sometimes respond that it was "unclear" if a website allowed users to delete their information. Other times, annotators answered "partial removal" because it was not clear whether or not websites would retain some of users' data after deletion.[88] These issues arose because some policies addressed editing and correction, but did not explicitly discuss deletion and some policies said that users could request deletion, but did not guarantee that the website would actually delete information.

This general doubt among users is likely to be a negative reflection on the clarity of the privacy policies themselves. Many policies did not even acknowledge whether users would or would not be able to delete their personal data. For example, with respect to the privacy policy for washingtonpost.com, three knowledgeable users selected that it was unclear if personal data was retained, while two answered that the question was not applicable and not addressed by the policy at all.[89] Similarly, three out of four experts selected unclear.[90] In the washingtonpost.com example, most of the annotators chose the statement "[i]f you do not wish to receive the foregoing and therefore unregister from the site, please contact Customer Care and ask to have your registration account deleted."[91] This statement, however, was in reference to deleting a "registration account" in order to stop receiving items from the website such as legal notices and users' account statuses. This passage was not directed toward the true inquiry of whether or not the website would retain user data for other purposes in the future, and the annotators' responses reflected this ambiguity.

c. Assumptions for the Interpretation of Specific Textual Language

Expert and knowledgeable users also made interpretative assumptions about the meaning of policy language. One interpretive difference arose in the context of permissive or conditional language. For example, the word "may" in a policy was not always interpreted in the same way

[86] *See* barnesandnonle.com (knowledgeable users) Question 7.
[87] *See e.g.*, lowes.com Question 9 (knowledgeable users); wunderground.com Question 9 (knowledgeable users); washingtonpost.com Question 9 (experts).
[88] *See, e.g.*, lowes.com (expert and knowledgeable users) Question 9. Notably, when asked if the website permitted users to delete their personal data, two experts answered "unclear", one "partial removal", and one "not applicable". Knowledgeable users expressed similar uncertainty in their answers to this question, with four knowledgeable users answering "unclear" while one answering "not applicable."
[89] *See* washingtonpost.com (knowledgeable users) Question 9.
[90] *See* washingtonpost.com (experts) Question 9.
[91] *See* washingtonpost.com Privacy Policy.

by each annotator. This was illustrated in cases where experts selected the same text but chose a different answer option.[92] This trend applied to both expert and knowledgeable users.[93]

Another similar divergence of interpretative assumptions arose with respect to "sharing for core service." In at least one instance, the experts made different assumptions with respect to the scope of certain language. One annotator believed that third-party services that "operate [the] online WunderStore" or "process credit card information" to support the website were outside the scope of "core service."[94] Others did not. Knowledgeable users, too, seemed to differ on their interpretations of what constituted a "core service." This confusion was apparent with respect to divergent answer choices for a privacy policy that described third parties as "performing a service" for the website when the nature of the services were not connected to the user's interaction with the website.[95]

---

[92] *See, e.g.*, wunderground.com Question 2. There, in response to a question about whether the website collects financial information, both annotators selected text that contained the following phrase: "companies that operate our online WunderStore, that process credit card information." *Id.* One annotator, in addition to selecting that phrase, selected the following language preceding it: "These Third Party Processors may include, for example," and likely interpreted this permissive language as rendering the language unclear (which corresponds to the answer option they selected). The other annotator, however, did not select this permissive language. Instead, the annotator likely interpreted it to mean that permissiveness means actual practice, and thus chose answer option 2 ("Yes – the policy explicitly states that the website might collect financial information").

[93] *See, e.g.*, Lowes.com Question 5. Three Knowledgeable users selected "unclear" when asked if the privacy allowed the website to collect users' contact information. This is likely due to the permissive language of the privacy policy, which stated that the website "may share personal information we collect on the Site with certain service providers, some of whom may use the information for their own purposes."

[94] *See* wunderground.com Question 6. There, two annotators selected the sentence: "These Third Party Processors may include, for example, companies that operate our online WunderStore, that process credit card information or handle shipping for Weather Underground." One annotator selected the whole sentence along with answer option 3 to notate that "the policy states that the website might share financial information with third parties for other purposes." *Id.* The other annotator selected only the dark portion with answer option 2 to notate that "the policy explicitly states that the website might share financial information with third parties, but only for the purpose of providing a core service." *Id.*

[95] *See* overstock.com Question 5:

> Service Providers
>
> We may share information with companies that provide support services to us (such as a printer, e-mail, mobile marketing, or data enhancement provider) or that help us market our products and services. These companies may need information about you in order to perform their functions. These companies are not authorized to use the information we share with them for any other purpose.

Id.

Two out of five knowledgeable users selected the above passage to support their selection that that the website shared contact information "for core service only." All remaining annotators selected the same sentences. However, two of the remaining annotators selected that the website shared user information "for other purpose." The final remaining annotator selected "unclear." The answers from the Overstock.com survey might suggest that two out of five knowledgeable users may have been influenced by the phrases "service providers" and "support services" when selected the option "sharing for core service only." Though these phrases seem to indicate that sharing would occur for reasons directly related to users' immediate business with Overstock.com, the end of the passage shows that this may not be the case. The final phase "mobile marketing", which is listed as a third example of sharing, demonstrates that contact information may in fact be shared for advertising purposes.

The experts also made different assumptions for the interpretation of "unclear" as described in the answer option choices. For example, in one response to the question about whether the website would share contact information with third parties, two experts selected the following:

> "As part of our ongoing partnership with Microsoft Corporation ("Microsoft"), we may share your personal information with Microsoft and its affiliates and subsidiaries under certain circumstances."[96]

However, one of the two experts who selected the text chose answer option 3 to reflect that "the policy states that the website might share contact information with third parties for other purposes." The other expert selecting the same text chose answer option 5 ("unclear") defined as: "the policy does not explicitly state whether or not the website might share contact information with third parties, but the selected sentences could mean that contact information might be shared with third parties." The second expert likely selected the "unclear" option because of the assumption that "personal information" did not mean or include "contact information."

The knowledgeable users appear to have the same divergence of assumptions over the interpretation "unclear." This divergence was seen particularly in connection with questions about the collection of users' health information. None of the privacy policies reviewed by the group contained an explicit reference to health information. However, knowledgeable users were often divided on whether or not such collection was possible.[97] For example, in one survey, three knowledgeable users believed that it was "not applicable" to ask whether or not the privacy policy collected users' health information. The description for this answer choice stated, "this question is not addressed by this policy."[98] Yet, two knowledgeable users believed that it was "unclear" whether or not health information was collected. These users selected the statement"[w]e collect the following categories of information," which was later defined as "[i]nformation you provide in public forums on our sites and applications" and "[i]nformation sent either one-to-one or within a limited group using our message, chat, post or similar functionality."[99] These users likely picked "unclear" because they read the policy language as ambiguous and as including the possibility that health information would be within the scope of data collection.

Finally, the experts sometimes differed in their interpretations of policy language if inferences could be drawn regarding a website's practices. For example, with respect to the collection of financial information, one expert selected the choice stipulating "the policy does not explicitly state whether the website might collect financial information or not, [but] the selected sentences could mean that financial information might be collected" and assumed that the selected text meant the website might collect financial information.[100] The expert was, in effect,

---

[96] *See* barnesandnoble.com Question 5.
[97] *See, e.g.*, lowes.com Question 4.
[98] *See* abcnews.go.com Question 5.
[99] *Id.*
[100] *See* overstock.com Question 2. One expert selected the following pieces of text to come to their conclusion:
- "You purchase, order, return, exchange or request information about our products and services from the Sites or mobile applications."

responding on the basis of an assumption about the scope of a term in the policy. As it turned out, the policy contained explicit language about the collection of financial information so the assumption was correct.[101]

d. Human Error

Experts appeared to make human errors with occasional instances of mistake. In some cases, experts inadvertently focused on irrelevant material. For example, in response to the question "Does the policy state that the website might collect contact information," one expert annotator selected explicit language from a policy's "Types of Information We Collect" section, while another expert annotator focused instead on the policy's section "We collect information when" and selected language that did not seem to be an appropriate response to the question.[102]

In a few other instances, experts simply made mistakes, as reflected by contradictory choices. As an illustration, one expert chose an answer option that should have been selected only if the accompanying passage mentioned "explicit consent;" the accompanying text passage, however, made no such mention.[103] Similarly, in response to a data deletion question, one expert

---

- "Product and Service Fulfillment"
- "Fulfill and manage purchases, orders, payments, returns/exchanges, or requests for information, or to otherwise serve you"

Id.

[101] "What Information We Collect[:] . . . Your credit/debit card number." *See* overstock.com Question 2.

[102] One expert selected the following language from the "Types of Information We Collect" section of overstock.com's privacy policy:

> Your name
> Your billing and delivery address
> Your e-mail address
> Your phone (or mobile) number
>
> *See* overstock.com Question 1. Another annotator selected the following language from the same policy's "We Collect Information When" section:
>
> You purchase, order, return, exchange or request information about our products and services from the Sites or mobile applications.
> You create an Overstock.com account
> You connect with Overstock.com regarding customer service via our customer service center, or on social media platforms.
> You visit the Sites or participate in interactive features of the Sites or mobile applications.
> You use a social media service, for example, Overstock.com's Facebook page or YouTube channel.
> You sign up for e-mails, mobile messages, or social media notifications from Overstock.com.
> You enter a contest or sweepstakes, respond to one of our surveys, or participate in a focus group.
> You provide us with comments, suggestions, or other input

See id.

[103] *See* wunderground.com Question 5. Here, one expert annotator selected answer option 4, which reads: "Sharing for other purpose (explicit consent) - the policy explicitly states that the website might share contact information

selected an answer choice that should have been chosen if the accompanying selected text mentioned that a user's data might be retained even after a request that it be deleted; yet, the accompanying selected text did not mention retention.[104] It is unclear whether these occurrences resulted from mistaken language selection, mistaken answer option selection, or other factors.

Lastly, experts occasionally missed relevant passages and selected a different portion of text that also accurately responded to the question asked. In these instances, each of the experts could have *also* selected the text that another expert selected.[105] This qualifies as a "mistake" because one expert missed relevant language that another saw. This example reveals that privacy policies can be confusing to a degree that even privacy policy "experts" have difficulty recognizing a policy's full scope. Ultimately, however, these "mistakes" will likely have little effect on adequately informing the NLP tool despite the experts' differences, as the data used to inform the tool will contain both selections. These mistake examples, though, suggest that crowdsourcing solutions may help remedy some of the human errors made while interpreting policies.

---

with third parties for a purpose that is not a core service, but only if the user provided explicit permission/consent to do so." The annotator selected the following text to accompany this answer choice: "we may share demographic information, location data, IP address, aggregate (not individual) usage statistics for our Services, other identifiers and information with advertisers and other third parties. For example, we may share IP address, random or anonymous device identifier, city and state, ZIP code, and specific geo-location with the parties identified in subparagraph F below." *Id.* This selection mentions sharing "for a purpose that is not a core service," but does not require explicit user consent for the sharing it describes.

[104] *See* washingtonpost.com Question 9. Here, one expert annotator selected answer option 3, which reads: "Partial Removal - the policy explicitly states that users may delete their personal data but some/all of the data might be retained for other purposes, whether the data was provided directly by the user, generated by the user's activities on the website, or acquired from third parties." That annotator selected the following policy text, which makes no reference to data retention: "If you do not wish to receive the foregoing and therefore unregister from the site, please contact Customer Care and ask to have your registration account deleted. Once your account has been deleted, you will no longer have access to washingtonpost.com, however, you may reregister at any time." *Id.*

[105] *See, e.g.*, abcnews.go.com Question 4 regarding collection of health information. In response, two experts selected answer option 3, which reads: "Unclear - the policy does not explicitly state whether the website might collect health information or not, but the selected sentences could mean that the health information might be collected." *Id.* To accompany this selection, one annotator chose one string of text ("Information you provide in public forums on our sites and applications [,] Information sent either one-to-one or within a limited group using our message, chat, post or similar functionality, where we are permitted by law to collect this information"), while the other selected a completely different string ("We acquire information from other trusted sources to update or supplement the information you provided or we collected automatically"). *Id.* In this instance, both annotators could have selected both strings of text.

Another example can be seen in lowes.com Question 4. There, in response to a question of whether the website might share contact information, one annotator selected option 3 ("Sharing for other purpose - the policy states that the website might share contact information with third parties for other purposes. The policy makes no statement about the users consent/permission or user consent is implied."), and another selected option 5 ("Unclear - the policy does not explicitly state whether the website might share contact information with third parties or not, but the selected sentences could mean that contact information might be shared with third parties."). Again, here, the annotators selected different language supporting each answer choice. One annotator (the one who chose option 5) selected, among others, the following string of text: "We may share personal information we collect on the Site with certain service providers, some of whom may use the information for their own purposes." *Id.* The other annotator (choosing option 3) selected this text: "We reserve the right to transfer personal information we have about you in the event we sell or transfer all or a portion of our business or assets (including, without limitation, in the event of a reorganization, dissolution or liquidation). Should such a sale or transfer occur, we will use reasonable efforts to direct the transferee to use personal information you have provided to us in a manner that is consistent with our Privacy Statement." And again, each annotator could have selected the text that the other annotator selected.

V. SIGNIFICANCE OF FINDINGS

Discrepancies between privacy experts and law and policy researchers reveal areas for careful attention to the quality of privacy policies. Discrepancies between privacy experts and non-expert users reveal whether website notices, as they are typically worded today, can effectively convey privacy policies to the general public. If websites are not effectively conveying privacy policies to consumers in a way that a "reasonable person" could understand the policies, notice and choice fails as a framework. If consumers cannot successfully decode privacy policies, the underpinnings of the United States approach to privacy are unsustainable, and regulation may be necessary. Indeed, a gap in interpretation indicates that privacy policies are, in fact, misleading the general public and that those policies could be considered legally unfair and deceptive.

This Part will address some of the implications from the findings for common understandings of website privacy policies and for crowdsourcing the interpretation of privacy policies.

A. Implications For Common Understanding and Consumer Deception

The findings show a number of areas where website privacy policies are too ambiguous to be meaningful and reveal a need to clarify specific data practices. The research demonstrates that policies describe websites' data sharing practices poorly. Experts could not reach consensus on interpretation of data sharing practices generally, and agreed *even less* as to the various nuances of data sharing.[106] Website owners must be more candid and clearer in drafting notices of data sharing practices. More precision is also needed with respect to the collection of specific types of users' personal information. The findings showed common understandings among experts on contact and location information, but not financial and health information.[107] This indicates that more clarity is necessary to spell out the specific practices with respect to sensitive information (i.e., financial and health information). General statements concerning "personal information" often introduce ambiguity into a policy and makes it difficult to interpret which information the website is actually collecting.

The findings also showed many instances where a majority of non-expert users interpreted terms differently from experts.[108] This implies that website notices are not conveying accurate information to consumers and that privacy policies may be misleading the public in specific areas. For example, knowledgeable users and crowd workers both lagged substantially behind the experts in their understanding of websites' data practices.[109] Conversely, there were instances when experts could not agree on an interpretation, but non-experts agreed on an interpretation. This may indicate that some language in website notices are commonly misunderstood where non-expert readers fail to recognize the ambiguity in a site's stated practices. For instance, when policies are silent on specific issues, or when conditions of sharing

[106] *See* Table 2.
[107] *See* Table 1.
[108] This arises when experts reach a *unanimous* consensus, yet a majority of non-experts interpret the same terms differently.
[109] *See* Table 5.

with third parties are not described clearly, non-experts will have a tendency to misunderstand the terms.

Lastly, since the findings reveal that complete agreement was uncommon, even among experts, website policies may be difficult to fully interpret through automated means where users are unable to understand the policies accurately. These difficulties are most pronounced at the level of specific practices. In instances where there is consensus on how data is treated *broadly* – for example, whether a policy states that a user may delete his or her personal information generally[110] – confusion exponentially increases with more nuance as to the data practice – such as whether a policy provides for an ability for *full* or *partial* deletion of user information.[111] This applied to each group of annotators. This disparity is very significant because personal privacy preferences are contextually based.[112] To have contextual integrity, the granular aspects of a data practice will need to be understandable to a user – not just whether a website collects, shares or deletes personal information in general. Indeed, a policy statement acknowledging general data collection and/or sharing may do very little to inform readers about the practices relevant for the user.

The lack of agreement and difficulties in interpreting policies suggest that consumers are currently misled by website privacy policies. To the extent that vague and misleading terminology is the result of drafting errors and omissions, this research shows areas where website policies can and need to be improved. To the extent, however, that the vague and misleading terms are intentionally introduced into website privacy policies, then this research suggests that websites are successfully deceiving consumers. The Federal Trade Commission under its unfair and deceptive practice jurisdiction and private litigants under state unfair and deceptive practice laws may be able to address these websites in enforcement proceedings.[113]

B. Implications for Crowdsourcing

Semi-automated extraction of meaning from website privacy policies may, in some circumstances, be a solution to help solve the difficulties users face in the interpretation and comprehension of privacy policies. Crowdsourcing is a critical part of such extraction and the findings of this research raise a number of implications for crowdsourcing opportunities. These implications arise from cases in our study when experts agree on the interpretation of terms in the privacy policies and when experts disagree on interpretation.

1. When Experts Agree

Where there is intra-group agreement among experts, then it is possible to compare the crowdsourced majority answer to the expert-selected answer to see whether a majority of crowd workers are able to arrive at the "correct" answer. Crowd workers in our study did a reasonable job predicting the answer when the experts had intra-group agreement. For example, experts had

[110] *See* Table 3.
[111] *See* Table 3.
[112] *See* Helen Nissenbaum, Privacy in Context: Technology, Policy and the Integrity of Social Life (Stanford Univ. Press: 2009).
[113] For a discussion of these types of enforcement actions, see Privacy Enforcement Actions, *supra* note 62.

a high level of intra-group agreement with respect to the collection of location information[114] and, at the same time, the majority of crowd workers chose the same response as the experts, signifying that both groups interpreted the collection of location information in the same way.[115] This alignment between experts and crowd workers on the interpretation suggests that crowd sourcing could be leveraged to identify collection practices for certain types of data in privacy policies, i.e. those in our study where agreement between experts and crowd workers indicates that the two groups interpretations are very close.

By contrast, where there is intra-group agreement among experts and crowd workers fail to arrive at the experts' "correct" answer, then crowd workers misunderstand policy terms. For example, experts were largely in agreement on whether websites collected financial information,[116] but the majority of crowd workers failed to select the experts' answer.[117] This inconsistency suggests that extracting information practices for these types of data cannot easily be crowd sourced to general users because their interpretation of the policy is likely to be wrong compared to the experts' interpretation.[118]

2. When Experts Disagree

Crowd workers, also, show positive signs at predicting expert disagreement. For example, both experts and crowd workers could not agree on whether websites shared location information.[119] This suggests that crowd worker disagreements can be a proxy for expert disagreements and that those disagreements can be used to identify aspects and potentially specific passages in websites' privacy policies that lack adequate precision.

By contrast, when experts disagree and the majority of crowd workers choose an answer, the crowd workers may either be misunderstanding the policy or interpreting the survey questions differently from the experts. For example, experts had low intra-group agreement on whether websites shared health information[120] and did not agree completely on the best answer choice.[121] Yet, crowd workers all converged on the same response to the sharing question.[122] In a case like this, there appear to be two possible explanations. First, the crowd workers might be failing to see the ambiguity in the policy and consequently would be misunderstanding the terms.

---

[114] *See* Table 1.
[115] *See* Table 4.
[116] *See* Table 1.
[117] *See* Table 4.
[118] At the same time, one NLP technique might be used to recognize text patterns in the policy that prompted experts' answer choices and then build crowdsourcing tools that combine machine learning and natural language processing techniques to improve the performance of crowd workers. The results of an automated extraction would be shown to crowd workers, for whom it may then be easier to verify the correctness of an extracted data practice or be less influenced by other text than to identify the data practice correctly without assistance. The possibility remains to be studied. Another possible opportunity for enhancing the effectiveness of crowd workers and the accuracy of their extractions is to train them with example annotations provided by experts or knowledgeable users.
[119] *See* Table 2. Experts were divided on the sharing of location information with a median level of agreement on all answer choices at 62.5% and crowd workers had a median level of agreement at 40%.
[120] *See* Table 2.
[121] *See* Table 5.
[122] *See* Table 5.

Alternatively, as reflected in the qualitative findings, the two groups might be interpreting the questions and answer choices differently.[123]

These issues require further exploration to inform whether and how crowdsourcing tasks might be organized in support of semi-automated privacy policy annotation to achieve replication of expert interpretations.[124] Similarly, further exploration will be necessary to identify whether and how to augment crowdsourcing with combinations of NLP techniques and machine learning. Our study strongly suggests, however, that for a narrow range of data practices, both crowd sourcing and automated extraction may be within reach. For other data practices, such as information sharing with third parties, where policies are sufficiently ambiguous to cause disagreement between experts, further research will be required to address the challenges of devising crowd sourcing tasks and automated extraction that would enable non-experts to accurately determine a policy's interpretation.

VI. CONCLUSIONS

The results of this study have significant implications for public policy. The disagreements among experts as reflected in the intra-group findings[125] mean that privacy policies are ambiguous on key terms. Specifically, the findings show interpretative challenges with respect to certain types of information, the scope of data sharing, and the scope of data deletion rights. The findings also showed that both knowledgeable users and crowd workers had greater difficulty deciphering privacy policy language than the experts.[126] These findings suggest that privacy policies are written ambiguously and in a way that leads both knowledgeable users and crowd workers to misapprehend websites' data practices as well as cause disagreement among experts with respect to certain data practices.

If the ambiguity in website privacy policies is unintentional, then the findings illustrate where businesses need to improve the clarity of their website privacy policies. The study methodology may also enable industry-wide tracking of how business sectors adjust their policies to more clearly explain their information practices. Specifically, web sweeps on a periodic basis to collect privacy policies as discussed in Part III.B would provide a data set to examine the evolution of privacy policies by industry sectors over time. Then, to the extent that some data practices can be extracted from crowd sourcing, and eventually natural language processing, these techniques could identify if privacy policy terms change over time.

If, however, the ambiguity is intentional, then the study findings suggests that website policies deceive the public. Such deception would be ripe for investigation by the FTC as "unfair and deceptive" business practice.

---

[123] *See* Part IV.C.2.c.

[124] For an experimental approach towards refining a work flow of crowd sourcing tasks for extracting data practices from privacy policies, see T.D. Breaux, F. Schaub Scaling Requirements Extraction to the Crowd: Experiments with Privacy Policies, 22nd IEEE International Requirements Engineering Conference (RE '14) (forthcoming) .

[125] *See* Parts IV.A and C.2.

[126] *See* Part IV.C.1.

For the development of automated and crowdsourcing tools to assist end-users and policy-makers in understanding privacy policies, the findings show initial promise as well as a need for further research into the important, identified challenges. Machine learning and natural language processing techniques might be useful to highlight potentially relevant passages in a privacy policy for crowd workers who are trying to extract information about that practice. Our findings further suggest that crowdsourcing can be used today to infer some aspects of the meaning of textual statements and these interpretations may be used to code for the automatic extraction of answers from those passages. Where crowd workers can predict expert disagreements about the interpretation of policy language, those ambiguous passages might be parsed out. For the remaining policy text, crowd workers' success in predicting answers where experts would agree, may be used to code passages where there is interpretative agreement. This may thus provide a more accurate understanding of the legal meaning of some of the privacy policy terms where an individual non-expert user would otherwise be misled. However, given a high level of disagreement among experts or crowd-predicted disagreement of experts, significant and important terms in a privacy policy will very likely escape effective interpretation by crowd workers because the effectiveness of automated tools and crowdsourced interpretation depends on an accurate baseline meaning for text in a privacy policy. Such a baseline meaning is currently missing for some of the key attributes of the website policies that were analyzed in this study.