

Ambiguity in Privacy Policies and the Impact of Regulation

Joel R. Reidenberg, Jaspreet Bhatia, Travis D. Breaux, and
Thomas B. Norton

ABSTRACT

Website privacy policies often contain ambiguous language that undermines the purpose and value of privacy notices for site users. This paper compares the impact of different regulatory models on the ambiguity of privacy policies in multiple online sectors. First, the paper develops a theory of vague and ambiguous terms. Next, the paper develops a scoring method to compare the relative vagueness of different privacy policies. Then the theory and scoring are applied using natural language processing to rate a set of policies. The ratings are compared against two benchmarks to show whether government-mandated privacy disclosures result in notices that are less ambiguous than those emerging from the market. The methodology and technical tools can provide companies with mechanisms to improve drafting, enable regulators to easily identify poor privacy policies, and empower regulators to more effectively target enforcement actions.

1. INTRODUCTION

Privacy policies often contain ambiguous language describing website practices for data-processing activities such as collection, use, shar-

JOEL R. REIDENBERG is the Stanley D. and Nikki Waxberg Professor of Law at Fordham University and a Visiting Research Collaborator at the Center for Information Technology Policy, Princeton University. JASPREET BHATIA is a Ph.D. candidate at the Institute for Software Research, Carnegie Mellon University. TRAVIS D. BREAU is an Associate Professor of Computer Science at Carnegie Mellon University. THOMAS B. NORTON was a Privacy Fellow at the Center on Law and Information Policy, Fordham University. Work on this project was supported in part by National Science Foundation awards 1330596 and 1330214, the Carnegie Mellon Requirements Engineering Laboratory, and a Fordham Faculty Fellowship. The authors would like to thank Stephen Broomell for his advice on testing validity and Stephanie Tallering for her assistance with coding. The authors benefited from very helpful comments on earlier drafts by participants at the 2015 Privacy Law Scholars Conference and the University of Chicago Coase-Sandor Institute conference Contracting over Privacy.

[*Journal of Legal Studies*, vol. 45 (June 2016)]

© 2016 by The University of Chicago. All rights reserved. 0047-2530/2016/4502-0023\$10.00

ing, and retention. While scholars have shown weaknesses in the readability of privacy policies (McDonald and Cranor 2008; Pollach 2007; Jensen and Potts 2004) and weaknesses in the substantive protections (Marotta-Wurgler 2016; Pollach 2007), they have not focused carefully on policy ambiguity. Ambiguity regarding these practices undermines the purpose and value of a privacy policy for website users. Without clear affirmative statements, privacy policies are, in effect, meaningless. They would convey no true indication to users of the website's actual practices, and they would provide declarations that would be unenforceable. On a practical level, ambiguity also challenges the usability of privacy technologies for user empowerment: clarity in privacy practices is a necessary prerequisite to empowering users to make informed decisions.¹

This paper explores the problem of ambiguity in policy language. In Section 2, we develop a theory for the definition of ambiguous terms and for the measurement of such terms. In Section 3, we develop a scoring method to compare the relative vagueness of different privacy policies. In Section 4, we apply the theory and method using natural language processing (NLP) techniques to score a set of privacy policies for clarity and comparison. We then use these comparative rankings to examine whether regulation improves the clarity of privacy policies. To test the impact of regulation, we compare the ambiguity of policy language under three conditions: no privacy regulation, regulation under the Gramm-Leach-Bliley Act, and regulation under the US-EU Safe Harbor intergovernmental agreement. The results provide normative insight into the role of privacy notice regulation. In Section 5, we address a number of practical public policy considerations resulting from our scoring. The techniques and corresponding technical tools can provide companies with a useful mechanism to improve the drafting of their policies. At the same time, automated tools embodying our theory and scoring method will enable regulators to easily scan industries and companies for poor language in their privacy policies. Such inexpensive scans revealing problems with privacy policy language then empower regulators to more effectively target defective privacy policies for remedial action.

1. For example, the joint Carnegie Mellon University, Fordham University, and Stanford University Usable Privacy Project seeks to combine crowdsourcing, natural language processing (NLP), and machine learning (ML) to develop browser plug-in technologies that will automatically interpret privacy policies for users (see The Usable Privacy Policy Project, Towards Effective Web Privacy Notice and Choice [<http://usableprivacy.org>]). If policies are too ambiguous, automated processing will be frustrated.

2. DEFINING AND MEASURING AMBIGUITY

2.1. Taxonomy of Vague and Ambiguous Terms through Grounded Theory

Ambiguity arises when a statement is incomplete and missing relevant information or when a word or phrase has more than one possible interpretation and the reader is uncertain about which interpretation the author intended. Linguists often address vagueness as a form of ambiguity (Massey et al. 2014). In contract theory, vagueness connotes a distribution around a norm without a clear delineation, while ambiguity refers to situations in which a word may have at least two meanings (Farnsworth 1999, sec. 7.8). In each case, multiple interpretations can arise when a statement is incomplete or when a generic word or phrase is used in place of a more specific word or phrase. When a website privacy policy uses vague or ambiguous terms, the language choices dilute the ability of a policy to describe the website's actual practices. This study will, however, focus on vagueness when terminology lacks specificity or context.

Because privacy policies summarize an organization's data practices, it is not surprising that policies include vagueness. There are at least two motivations for introducing vagueness: the practices include divergent or separate situations when actions do and do not occur, in which case the action may occur, depending on what situation the individual encounters, and there are foreseeable, yet unrealized actions that may occur in the future, and the policy authors wish to be flexible to accommodate those future actions without changing the policy. In the case of the first motivation, we believe that changes to some policy statements can clarify under what situations the action does or does not occur, which would result in a less vague policy. However, the second motivation to accommodate flexibility is at best a form of inaccuracy, whether the result of hedging to cover unknown existing internal practices or unknown changes, and at worst misleading and misrepresentative.²

To demonstrate this effect, we show a few illustrative examples from the Barnes & Noble privacy policy concerning personal information that are commonly found in other policies.³ The Barnes & Noble policy includes two statements that describe the possibility of collection:

2. The desire for flexibility might also be seen as a reservation by the website to hold an option on the ability to engage in unstated data practices. Even if the language explicitly describes an option, the terminology still encompasses alternative meanings, and the user will not know the subjective intent of the website, thus creating vagueness.

3. The Barnes & Noble privacy policy (dated May 7, 2013) is available from the authors on request.

1. “Depending on how you choose to interact with the Barnes & Noble enterprise, we may collect personal information from you.”

2. “We may collect personal information and other information about you from business partners, contractors and other third parties.”

In statement 1, the collection is conditioned on how the user interacts with the company. This is vague, because the statement summarizes multiple situations, some of which will include the collection of personal information and some of which will not. To achieve clarity, it would be reasonable to exclude those situations in which personal information is not collected and to focus on where personal information will be collected. In contrast, statement 2 is vague because the conditional situations are not described; thus, all third-party transactions are summarized in a single statement. By separating these statements and iterating over the different categories, the policy’s authors can exclude prospective collections (envisioned but not actual collections) and those situations in which personal information is not collected.

Another attribute of vagueness concerns the vague conditions and purposes under which information is used. In statement 3, below, the Barnes & Noble policy links the collection to a broad purpose (improving customer experience) under a general assumption about the “necessary” situations that define this broad purpose. An alternative statement would replace the phrase “as necessary” with specific purposes intended to improve customer service.

3. “We collect your personal information in an effort to provide you with a superior customer experience and, as necessary, to administer our business.”

Vagueness pertaining to standard third-party transactions is also evident in the following two statements from the Barnes & Noble privacy policy:

4. “In addition, we disclose certain personal information to the issuer of the MasterCard.”

5. “If you are accessing our goods and services using a Microsoft account, Microsoft may share your personal information with us.”

In statements 4 and 5, the mechanisms for exchanging personal information are coded in software. In the case of credit card transactions, MasterCard, card issuers, and acquiring banks all have transaction-processing rules that are updated from time to time, but the technical specifications of their respective electronic payment processing infrastructures are less likely to change. The Barnes & Noble policy could thus re-

strict the kind of personal information it shares to payment information or information for the purpose of completing a purchase.⁴ In statement 5, Microsoft's Live Connect API for OAuth 2.0 access to the Microsoft account is also very explicit about what information may be shared (first and last names, email address, gender, and age), and Barnes & Noble can further commit to which of these information types it will collect as encoded by its software.

As the case study reveals, the contours of vagueness are very complex. The measurement of vagueness thus becomes a valuable marker to signal whether a privacy policy is a meaningful notice of a website's actual policies and practices and a notice that might give rise to a contractual commitment. The first step in the measurement of vagueness in privacy policies is the development of a rigorous and validated taxonomy of terms that can be used to examine a diverse set of online sectors such as shopping, news, and financial services.

Linguistic scholars have identified various forms of ambiguity in the use of language and have classified textual ambiguity in various ways (Hoffman, Ralph, and Rogers 2013; Massey et al. 2014; Pollach 2007). Some of these classifications reflect terms that may have inherent vagueness (Massey et al. 2014). For example, many privacy policies use the modal verb "might" to describe data-processing activities ("we might collect") that may or may not occur in the future. In addition, policies use conditional phrases, such as "when," "upon," and "during," that indicate an event upon which a particular statement becomes true ("upon consent, we will share"). When multiple modal verbs and conditional terms are used together, readers struggle to determine if the described practices actually occur, or in what combination or under which specific conditions they occur, or how to satisfy those conditions.

For a rigorous analysis of textual ambiguity, the starting point is, thus, the establishment of a typology of ambiguous terms. Since our objective is to provide a qualitative rating of vagueness, we have chosen to focus on this narrower aspect of ambiguity. This means that our scoring will provide relational comparability but underrate overall ambiguity.

4. Statement 4 is also separately confusing because Barnes & Noble would not typically be able to share data directly with a card-issuing bank. Barnes & Noble exchanges data from the point of sale to its acquiring bank, which in turn shares information specified by MasterCard to the customer's card-issuing bank through the MasterCard network. The circumstance that might give rise to direct sharing from Barnes & Noble to a card-issuing bank would be the case of a cobranded card for which Barnes & Noble would have a direct relationship with the card issuer.

Table 1. Categories of Vague Terms

Category	Description
Condition	Action(s) to be performed are dependent on a variable or unclear trigger
Generalization	Action(s) or information types are vaguely abstracted with unclear conditions
Modality (including modal verbs)	Vague likelihood of action(s) or ambiguous possibility of action or event
Numeric quantifier	Vague quantifier of action or information type

We define our typology using grounded theory (Glaser and Strauss 1999) to identify vague terms and classify those terms into categories. Three researchers manually performed this analysis using coding (Saldaña 2012) to examine a set of policies across a variety of sectors (Bhatia et al. 2016). Five policies were used for the initial identification and classification of relevant terms (Bhatia et al. 2016).

The analysis resulted in a taxonomy with four categories as shown in Table 1. From a legal perspective, conditional terms are inherently vague because the performance of a stated action or activity will be dependent on a variable trigger (Farnsworth 1999, sec. 8.2). Similarly, generalizations are terms that vaguely abstract information practices using contexts that are unclear (for example, “typically” or “generally”). From the linguistic perspective, modality (modal verbs, adverbs, and nonspecific adjectives) creates uncertainty with respect to actual action (von Fintel 2006); this includes whether an action is possible, likely, permitted, or obligatory, among others. If the action is only permitted, it may never occur, whereas obligatory actions are expected to occur in the future (the difference between “we may” and “we will”). Similarly, numeric quantifiers that are nonspecific create ambiguity as to the actual measure. To assure the completeness of the typology, three researchers reviewed 15 policies first in their entirety and then statement by statement to identify vague phrases and determine if they fit these categories or if new categories were required (see Table A1).⁵

To see how a sentence may reflect these categories, the phrase shown in Figure 1 contains a condition, generalization, modal verbs, and nu-

5. This also included an evaluation of cases in which the language might appear as a borderline classification.

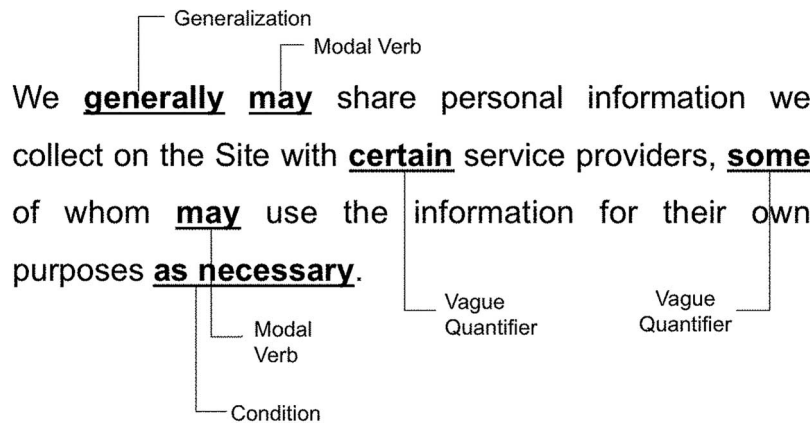


Figure 1. Sentence annotation demonstrating four categories of vague terms

meric quantifiers.⁶ These vague terms are annotated in Figure 1. In combination, these six forms of vagueness combine to allow any organization sharing personal information under this statement to share it with anyone for any purpose, as long as the recipient is a service provider. The combination of these six terms further leaves unclear the conditions under which information is shared and the number or proportion of service providers that engage in this practice.

2.2. Vague Terms

To complete the lexicon of vague terms, we used an established coding frame based on the taxonomy in Table 1, and three researchers analyzed a set of 15 policies (Table A1). To assure saturation of terms, we examined three diverse sectors (shopping, telecommunications, and employment) and five policies in each sector reflecting a diversity of types of websites in each category.⁷ We chose privacy policies of major sites that are visited by large numbers of users.⁸ In this study, we reached saturation after analyzing five policies (Barnes & Noble, Lowe's, Costco,

6. The original privacy policy language was taken from the Lowe's website (Lowe's, Lowe's US Privacy Statement [http://www.lowes.com/l/privacy-and-security-statement.html]) and was modified to include "generally" and "as necessary" for illustration purposes.

7. We considered examining policies from top site rankings, but the various rankings did not assure diversity of sectors.

8. The small number of policies in each category precludes broad generalizations within and across categories but does enable us to show the value of a score for comparison purposes, including comparison against the financial services benchmark.

Table 2. Results from Applying Taxonomy to Privacy Policies

Category	Key Words and Phrases	Distribution
Condition	Depending, necessary, appropriate, inappropriate, as needed, as applicable, otherwise reasonably, sometimes, from time to time	7.20
Generalization	Generally, mostly, widely, general, commonly, usually, normally, typically, largely, often, primarily, among other things	3.63
Modality (including modal verbs)	May, might, can, could, would, likely, possible, possibly	70.60
Numeric quantifier	Anyone, certain, everyone, numerous, some, most, few, much, many, various, including but not limited to	18.60

Note. The distribution represents the number of vague terms in the 15 policies belonging to the category divided by the total number of vague terms in 15 policies. See Bhatia et al. (2016).

AT&T, and Comcast), which reflects that our taxonomy of terms was complete for the privacy policy domain. The resulting set of terms for the taxonomy is shown in Table 2.

2.3. Comparative Levels of Ambiguity

While the taxonomy results in Table 2 present the terms that obscure the clarity of the policy descriptions, the taxonomy does not address the relative levels of ambiguity among the terms. For example, the following two statements appear to have different levels of ambiguity:

1. “We may generally collect . . .”
2. “We may collect as necessary . . .”

Each uses a modal verb (“may”), but statement 1, which contains the generalization “generally,” seems less clear than statement 2, which contains the condition “as necessary.” The practice described “as necessary” suggests that collection will occur only in exceptional cases, while “generally” suggests that collection is likely to occur under broader circumstances. This qualitative difference in clarity may also be linked to the degree of flexibility that the textual language provides to the website. Language designed to give websites greater flexibility is likely to be perceived as more ambiguous. The statement “may collect generally” provides greater flexibility to the website than “as necessary.” Consequently, the generalization term “generally” obscures for the user the website’s activities more than the conditional term “as necessary.”

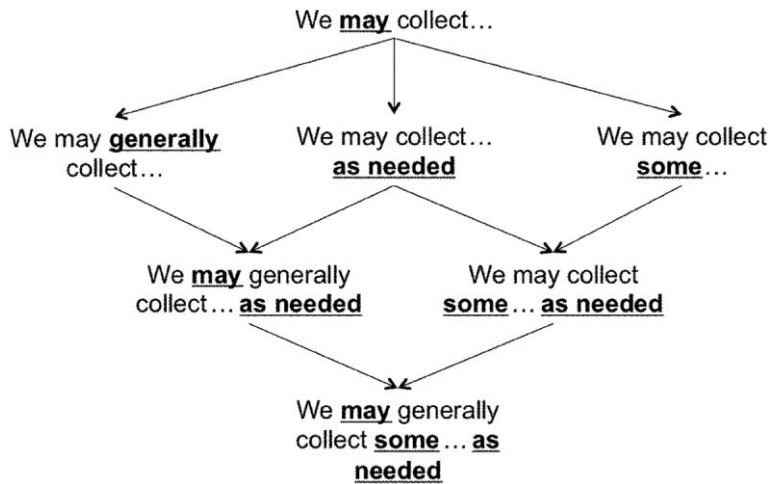


Figure 2. Cumulative effect of vague terms

In addition to variations in clarity among the categories, the combination of terms from different categories in the same sentence may also affect the level of ambiguity perceived in descriptions of privacy practices. Figure 2 shows our initial hypothesis regarding the possible cumulative effect of vague terms represented as a lattice.

The lattice begins with a modal statement “we may collect” and then in the next row adds a term from each of the remaining three categories: a generalization term “generally,” a conditional term “as needed,” and a numeric quantifying term “some.” Our initial assumption was that additional terms would increase the vagueness of the statement, that is, reduce the clarity of the description of the data collection practice. With each successive combination of vague terms, from the first to the second, third, and fourth rows in Figure 2, vagueness would increase until some degree of saturation would occur (that is, adding additional vague terms would have no significant impact on increasing vagueness).

The relative impact of each possible combination is critical to the development of an accurate score for a privacy policy’s ambiguity. To evaluate our initial assumption, we conducted a paired comparison survey.⁹ The survey results show the relationship of combinations of terms

9. A paired comparison survey is a standard statistical technique that collects multiple preferences between two statements from multiple judges and, through the aggregation of the results, establishes a matrix of rating comparisons for all possible combinations of the terms being studied.

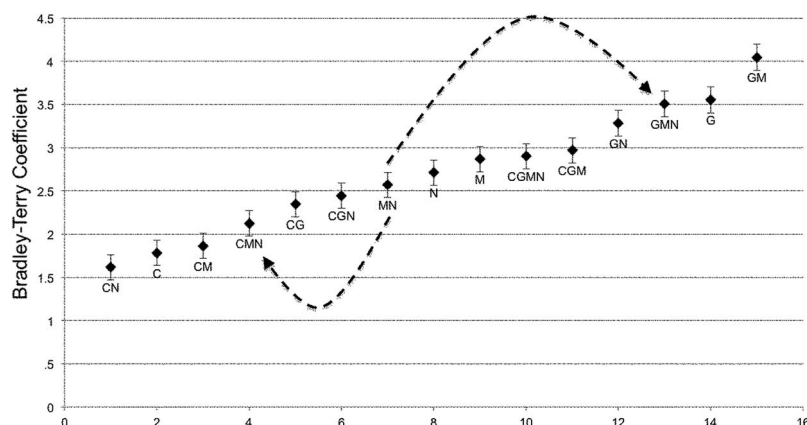


Figure 3. Bradley-Terry coefficients for intercategory combinations

on the level of ambiguity, enabling the assignment of relative weights to different combinations of terms from one or more categories. We used the Bradley-Terry model that scales preferences among different pair comparisons to calculate the weights from the paired comparison data (Turner and Firth 2012).

These results (from Bhatia et al. 2016) are presented in Figure 3 and Table A2. Figure 3 shows the Bradley-Terry coefficients for the combinations of conditions (C), generalizations (G), modal terms (M), and numeric quantifiers (N).

These results show the quantity that each combination of vague terms contributes to the overall concept of vagueness in the survey: data practices described with combinations at the left of the chart (CN, C, CM, and so on) have greater clarity than those practices described with combinations at the right of the chart (GMN, G, GM, and so on). While phrases with both a conditional term and a vague numeric quantifier (CN) are indistinguishably clear from phrases with just a conditional term alone (C), we can observe how the vagueness taxonomy influences overall vagueness. The arrow pointing left in Figure 3 shows that conditional terms increase clarity and reduce vagueness: for example, statements with both a modal term and numerical quantifier (MN) are significantly vaguer than similar statements with the addition of a conditional term (CMN). The arrow pointing right in Figure 3 illustrates how generalizations significantly increase vagueness: for example, the MN statements with the addition of a generalization (GMN) are significantly vaguer. By compari-

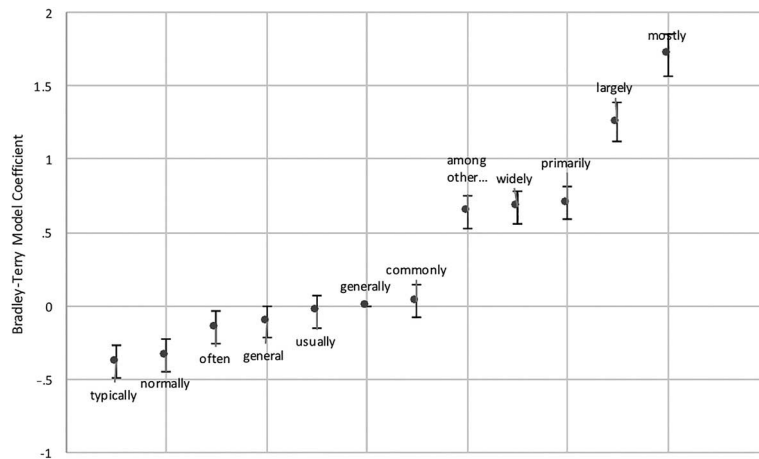


Figure 4. Bradley-Terry coefficients for generalization terms

son, statements with a generalization and modal term (GM) are twice as vague as statements with a condition and a modal term (CM).

The results in Figure 3 present the intercategory vagueness. To measure the intracategory vagueness between terms within each of the categories, we conducted additional surveys (Bhatia et al. 2016). The survey results indicate that terms in each category have different levels of vagueness. The intracategory vagueness results for the generalization category are presented in Figure 4, and the results for all categories appear in Table A3.

The results in Figure 4 show that within the generalization category, vagueness appears to increase as the adverbs transition from the routine (for example, “typically,” “normally,” or “usually”) to the unrestricted (for example, “widely,” “largely,” or “mostly”). The full results in Table A3 show similar differentials. In the condition category, the term “as appropriate” was several times vaguer than the term “as necessary.” In the modality category, the past-tense verbs “might” and “could” are perceived to be vaguer than the present-tense variants “may” and “can,” respectively. These three observations led Bhatia et al. (2016) to conjecture that vagueness increases along three dimensions: authority, wherein discretionary practices are perceived to be vaguer than mandatory practices (for example, “as appropriate” is permissive, whereas “as necessary” is obligatory); certainty, which is the absoluteness with which practices are performed (for example, “typical” is certain with respect to common

cases, whereas “widely” blurs the boundary between common and exceptional cases); and likelihood, which is the possibility that the practice is performed (for example, “likely” is more likely than “possibly,” and thus less vague).

2.4. Ambiguity through Incompleteness

Finally, silence in a privacy policy can often introduce ambiguity (Marotta-Wurgler 2016). For example, if the policy is silent on sharing data with third parties, then the policy fails to convey whether and under what conditions data may be transferred to others. As a result, completeness of the privacy policy will have an impact on the scoring of ambiguity. While there are no legal requirements spelling out all the terms that must be contained in a privacy policy, various templates might be used to determine completeness. For example, the Gramm-Leach-Bliley Act stipulates only that financial service companies provide notice to customers of their privacy policies and that the notice at a minimum contain certain types of disclosures (Gramm-Leach-Bliley Act, 15 U.S.C. 6803 [1999]). The Federal Trade Commission and the Department of Commerce have each articulated several sets of fair information practices (Federal Trade Commission 1998, 2000; Department of Commerce, Notice: Issuance of Safe Harbor Principles and Transmission to European Commission, 65 Fed. Reg. 45,665–86 [July 24, 2000]). For purposes of this analysis, the existence of four elements are inventoried: collection, retention, sharing, and use. These elements reflect the most significant privacy harms demonstrated through litigation that we believe can be resolved by unambiguous privacy policy statements (Reidenberg et al. 2015).

3. SCORING VAGUENESS

With the vagueness taxonomy populated using key words and phrases corresponding to each category, a comparative classification and a completeness indicator can be constructed to score the degree of affirmation or certainty associated with data practices for specific types of personal information. Privacy policy statements about companies that “might collect” are less certain than statements that they “will collect” or “will not collect” a particular information type. Highly uncertain statements can more easily accommodate a company’s future practices, thus providing these companies more flexibility in the interim to alter those practices.

However, highly uncertain statements allow for interpretations that may be untrue, thus giving users a false sense of privacy. By contrast, if an organization has a policy that is more certain, particularly with more restrictive practices, any new unstated practices would require a change in the policy. Such changes would trigger opportunities for users to reevaluate their relationships with those companies under the new practices. This opportunity to evaluate policy changes is necessary if the privacy principle of user consent is to have any meaning. Policies containing more certain statements are more likely to increase the opportunity for choice, since those policies will need to be revised each time a new practice is to be covered.

To score privacy policies, the first step is to determine if vague terms are common in privacy policies through an analysis of the landscape of terms found in privacy policies. The frequent existence of vague terms leads to the definition of a scoring model that can then be applied to privacy policies to rank their vagueness against each other.

3.1. The Landscape of Vagueness in Privacy Policies

When the taxonomy is applied to the set of 15 privacy policies (Table A1), every policy in the data set contains vague terms (Bhatia et al. 2016). As shown in Bhatia et al. (2016) and Table 2, the most frequently observed ambiguous terms are modal verbs (M) followed by numeric quantifiers (N); conditions and generalizations lag far behind (see Table A4). This suggests that the use of modal terms will dominate all other terms in the calculation of overall vagueness in a privacy policy. However, the Bradley-Terry coefficients show the significant impact that conditions, generalizations, and numeric quantifiers have on modality: while modality alone (M) scores at $2.865 \pm .147$, the addition of generality to modality (GM) scores vaguer at $4.045 \pm .156$, and the addition of conditions to modality (CM) scores less vague at $1.864 \pm .146$ (Bhatia et al. 2016; Table A2). In the extreme case that these additional categories of vagueness always appear with modal terms, then over one-third of the total 70.6 percent of modal terms will score well above or well below the coefficient for modality alone (Bhatia et al. 2016). This can lead to significant differences between policies in terms of overall vagueness and especially pronounced differences within a single category of data practice (for example, collection, retention, sharing, and usage) in the event that vagueness is concentrated in one area of the policy.

3.2. The Scoring Model

Simply counting the number of vague terms in a privacy policy will not provide an adequate measure of ambiguity. For example, the AT&T policy contains 70 vague phrases, which places it at the median of 70 vague phrases and just below Time Warner's, which has 85 vague phrases. But this frequency count does not indicate the relative context. Context matters, and a granular scoring model needs to take into account three key variables: the existence of vague terms and their relation to specific categories of data practice (for example, collection, retention, sharing, and usage), the relative impact that a combination of vague terms may have on overall ambiguity, and the completeness of the policy.

To accomplish this goal, we propose a scoring model based on a relative comparison of vagueness in phrases for each policy. This score is based on a statistical measure that scales the overall vagueness of individual statements in each policy based on the Bradley-Terry model for paired comparisons.

The coefficients that were computed by this method serve for these calculations to rank the vagueness of every phrase in each policy containing a vague term or combinations of vague terms associated with an action-information pairing in which one of the four identified data practices (action) is applied to a type of information (information).¹⁰ The data practices were extracted using the technique developed by Bhatia, Breau, and Schaub (2016). The vagueness scores appropriately ignore phrases that do not specifically describe a data-processing activity or that do not contain any vague terms. This means that irrelevant language, such as a corporation's philosophy relating to privacy, or unambiguously described data practices will not factor into the vagueness score.

For each policy, we can then calculate an aggregate vagueness score by taking the sum of the coefficients for each action-information pair containing vague terms. This policy-specific aggregate score is not, however, sufficient to compare two policies. For example, if a policy is long, it may contain more action-information pairs containing vague terms than a shorter policy but proportionately be much clearer. To account for this situation, we normalize the aggregate vagueness score by dividing the aggregate score by the total number of action-information pairs in the policy; we call this normalized score the vagueness score. The vagueness

10. If a statement contained more than one action-information pairing, then all the pairs were included and contributed to the vagueness score.

score reflects positively on the policy and improves if a policy has more action-information pairs that clearly describe data practices and reflects negatively on the policy and worsens if the policy has more pairs that include vague terms. Moreover, it reflects the total unit vagueness independent of policy length but relative to the level of contribution to vagueness by each category of vague terms in Table 2. This can be represented by the following equation:

$$V = \frac{\sum(BTC^{A-I})}{\sum(A-I)},$$

where V is vagueness, BTC is the Bradley-Terry coefficient, and $A-I$ is the action-information pair.

Finally, in case a policy has a high level of ambiguity in paragraphs pertaining to key elements that may be masked by clear language elsewhere in the policy, we calculate the vagueness scores for the collection of policy statements addressing each of the four key data practices: collection, retention, sharing, and usage. These scores are calculated in the same manner as those for the overall policy.

Separately, we report on the completeness of the privacy policies using a scale from 0 to 4. For each element missing from the four data practices (collection, retention, sharing, and use), the policy is assigned 1 point. Thus, a policy containing any description for all four elements will score a 0, and a policy missing all four elements will score a 4.

4. COMPARATIVE SCORES AND THE IMPACT OF REGULATION

4.1. Company Scores for Unregulated Disclosures

Applying the scoring model to the privacy policies of companies that do not have specific notice obligations results in the vagueness scores reported in Table 3. Where the ratios are in proximity to each other, they indicate that those policies have similar levels of ambiguity. Where a ratio is double another, the ratios indicate that the policy with the higher ratio is twice as vague as the policy with the lower ratio.

Table 3 shows that the most ambiguous policies among the unregulated entities belong to Time Warner, with Comcast, Overstock.com, and Charter Communications clustered close behind. These policies use large numbers of ambiguous modal verbs and numeric quantifiers (Table A4). For example, the Comcast policy describes sharing with third parties us-

Table 3. Unregulated Companies' Privacy Policy Vagueness Scores

Company	Total Score	Collect	Retain	Share	Use
Costco	1.02 (.06)	.68 (.04)	.95 (.05)	1.51 (.08)	.63 (.04)
JCPenney	1.19 (.07)	1.32 (.08)	1.44 (.08)	1.16 (.07)	1.07 (.06)
Lowe's	1.28 (.07)	.87 (.05)	2.15 (.11)	2.06 (.11)	1.25 (.07)
Overstock.com	1.71 (.09)	1.56 (.09)	1.44 (.08)	2.03 (.11)	1.62 (.09)
AT&T	1.04 (.06)	.92 (.05)	.45 (.04)	1.25 (.07)	.99 (.06)
Charter Communications	1.64 (.09)	1.54 (.09)	1.02 (.07)	1.72 (.09)	1.84 (.10)
Comcast	1.80 (.10)	1.71 (.10)	1.75 (.10)	1.96 (.11)	1.66 (.09)
Time Warner	2.09 (.11)	2.1 (.11)	2.79 (.15)	1.72 (.09)	2.17 (.12)
Verizon	1.38 (.08)	1.41 (.08)	.80 (.05)	1.48 (.08)	1.34 (.08)
Simply Hired	1.56 (.09)	1.44 (.08)	.64 (.04)	1.12 (.07)	1.97 (.11)
Mean	1.36	1.34	1.60	1.45	1.47

Note. Standard errors are in parentheses. Completeness = 0 for all policies.

ing both a modal verb ("may") and numeric quantifier ("certain"): "In certain situations, third-party service providers may transmit, collect, and store this information on our behalf to provide features of our services" (Comcast 2015). By contrast, Costco's language describing sharing with third parties is more direct: "We do not otherwise sell, share, rent or disclose personal information collected from our pharmacy pages or maintained in pharmacist records unless you have authorized such disclosure, or such disclosure is permitted or required by law" (Costco Wholesale Corporation 2013). In comparison to the vaguest policies, the policies belonging to Costco and AT&T are almost twice as clear.

Table 3 also shows the vagueness scores for actions to collect, retain, share, and use information. The overall mean vagueness across these four actions varies little from 1.34 to 1.60; however, the mean variance is not homogenous across practices (collect variance = .21, retain variance = .52, share variance = .10, and use variance = .30). This variance across practices shows divergent uses of vague terms across companies, with

the least consistency across policy descriptions of retention practices and the most consistency around descriptions of sharing practices. Notably, companies such as Comcast and Time Warner score higher than average for vagueness in all four practice categories. For the website user, however, Overstock.com's high vagueness score for sharing (2.03) presents a more significant, or fundamentally different, privacy risk than Comcast's vagueness regarding collection (1.71) and retention (1.75). Vagueness with respect to sharing is significant because third parties are rarely identified in privacy policies, and most privacy policies disclaim responsibility for the data practices of the unnamed third parties. Vagueness with respect to collection and retention affords companies greater flexibility in broadening what kinds of information they are potentially collecting. This may or may not present heightened privacy risks. However, when these are combined with ambiguous sharing terms, website users will not be able to ascertain exactly what information may be at risk of sharing with third parties. All the policies not subject to regulation were complete.

4.2. Scores for Regulated Disclosures

Because the score ratios are designed to compare the clarity of policies against each other and do not provide a minimum level of acceptability for ambiguity, the model privacy form adopted under the Gramm-Leach-Bliley Act can serve as an informative target benchmark for a regulated notice (Final Model Privacy Form under the Gramm-Leach-Bliley Act, 74 Fed. Reg. 62,890–994 [December 1, 2009]). This model form was adopted by regulatory agencies after careful analysis and testing of language options (Levy and Hastak 2008). In fact, eight federal financial service regulatory agencies approved the language used in this standardized privacy disclosure statement. Financial service providers may use the model form to satisfy their obligations under the Gramm-Leach-Bliley Act, though they are not required to adopt its language.

Table 4 presents the vagueness score calculations for a set of five large national financial institutions that adopted privacy policies based on the model privacy form. Another benchmark can be derived from the US-EU Safe Harbor Agreement (65 Fed. Reg. 45,665–86).¹¹ The US-EU

11. Although the Safe Harbor Agreement was struck down in *Maximilian Schrems v. Digital Rights Ireland* (CJEU C362/14, judgment of October 6, 2015 [<http://tinyurl.com/occ4mvxj>]), the policies previously written under the agreement are nevertheless a valuable benchmark for a regulatory rule that does not provide explicit language for privacy notices.

Table 4. Financial Services' Privacy Policy Vagueness Scores

Company	Total Score	Collect	Retain	Share	Completeness
Bank of America	.96 (.05)	.48 (.03)	2.87 (.15)	1.03 (.06)	0
Capital One	.52 (.03)	.58 (.03)	2.87 (.15)	.38 (.02)	0
Citigroup	.45 (.03)	.58 (.03)		.43 (.03)	1
J. P. Morgan	.36 (.02)	.48 (.03)	0 (0)	.56 (.03)	0
PNC	.35 (.02)	.58 (.03)		.31 (.02)	1
Mean	.52	.54	1.91	.54	

Note. The Citigroup and PNC policies do not mention retention practices, so the mean score for retention is computed excluding these policies. Standard errors are in parentheses. Use = 0 for all policies.

Safe Harbor Agreement identifies data practices that must be contained and described in a privacy policy to satisfy European data export requirements but stops short of providing model language like the model privacy form in the United States. The framework was negotiated between the United States and Europe and then approved by the US Department of Commerce. Companies may benefit from the US-EU Safe Harbor Agreement if they include specified provisions in their privacy notices and register with the US Commerce Department.

Of the 15 companies in our data set (Bhatia et al. 2016), five are on the US-EU Safe Harbor List (Department of Commerce 2015). Table 5 applies the scoring model to these five privacy policies.

The mean vagueness score for the financial services policies is considerably lower than that of the Safe Harbor policies: .52 versus 1.20. This striking more-than-twofold difference means that financial services policies are more than twice as clear as the Safe Harbor policies. Similarly, the vagueness scores show that the descriptions of three of the four data practices found in the financial services policies have greater clarity than those found in the Safe Harbor policies. As a benchmark, the model privacy form for the financial services industry holds privacy policies to a higher standard of clarity and allows less ambiguity than the US-EU Safe Harbor Agreement. All the benchmark policies were complete with the exception of the Citigroup and PNC policies, which were silent on data retention.

Table 5. Safe Harbor Companies' Privacy Policy Vagueness Scores

Company	Total Score	Collect	Retain	Share	Use
Barnes & Noble	2.07 (.12)	2.19 (.12)	1.49 (.09)	2.3 (.13)	1.78 (.10)
CareerBuilder	.84 (.05)	.83 (.05)	.81 (.05)	.89 (.05)	.85 (.05)
Glassdoor	1.36 (.08)	1.41 (.08)	1.23 (.07)	1.54 (.09)	1.26 (.07)
Indeed	.96 (.05)	.8 (.04)	1.08 (.06)	1.04 (.06)	.94 (.05)
Monster	.79 (.05)	.86 (.05)	.72 (.04)	1.12 (.06)	.58 (.04)
Mean	1.20	1.22	1.07	1.38	1.08

Note. Standard errors are in parentheses. Completeness = 0 for all policies.

4.3. Normative Role of Privacy Notice Regulation

Comparing the vagueness scores for the regulated benchmark policies against the unregulated policies shows that the unregulated policies have notably higher scores and use significantly more ambiguous language. The findings indicate that more specific regulation of policy language has a positive impact on the clarity with which privacy policies describe data practices. Table 6 compares the mean vagueness scores of the two benchmarks to the privacy policies of the unregulated companies.

The overall mean score for vagueness across data collection, retention, sharing, and usage is 1.03. While the number of policies per sector is presently small, we see a few mean differences when accounting for outliers or extreme differences. Among those surveyed, telecommunications companies with the exception of AT&T show more vagueness about collection practices than do the employment or shopping sectors.

The mean vagueness score for the unregulated policies in Table 6 is 1.36, which is over 2.5 times higher than the mean for financial services policies (.52). Similarly, the mean vagueness score for the Safe Harbor policies is 1.20, which is slightly lower than that of the unregulated companies but well over twice as high as the financial services benchmark.¹² With respect to the descriptions of data practices (collection, retention, sharing, and use), the unregulated privacy policies have significantly less

12. This slight difference may be a reflection of self-selection; companies choosing to adhere to the US-EU Safe Harbor Agreement may be more transparent than typical companies. The closeness of the Safe Harbor scores to the unregulated scores, however, suggests that the self-selection effect is limited at most.

Table 6. Mean Vagueness Scores by Policy Type

Privacy Policy	Vagueness Score	Collect	Retain	Share	Use
Financial services	.52	.54	1.91	.54	.00
Safe Harbor	1.20	1.22	1.07	1.38	1.08
Unregulated	1.36	1.34	1.60	1.45	1.47
Mean	1.03	1.04	1.53	1.17	.85

clarity than the financial services benchmark in all areas except retention and have mildly less clarity than the Safe Harbor policies in all areas.

The more-than-double difference between the means for financial services policies (.52) and the Safe Harbor policies (1.20) indicates that a regulatory nudge providing specific model language (financial services) results in less-ambiguous policies than a regulatory framework (Safe Harbor) that states only what kind of data practices must be described in a policy.¹³ The difference also shows that Safe Harbor does a poor job of assuring clear descriptions of data practices when compared with the model privacy form. The difference between the means for unregulated policies (1.36) and Safe Harbor policies (1.20) similarly shows that the Safe Harbor Agreement did not largely improve the clarity of privacy policies.

Overall, the comparisons with the benchmarks indicate that the market produces privacy policies that are more ambiguous than those subject to some form of regulation. The least ambiguous policies were those adopted by financial services companies after government regulators provided their imprimatur to specific language. While financial institutions were not required to adopt the model privacy form, doing so assured compliance with their legal obligations under the Gramm-Leach-Bliley Act. At the same time, the policies with the highest level of vagueness were those of the unregulated companies.

5. PUBLIC POLICY CONSIDERATIONS: TECHNICAL TOOLS, LINGUISTIC GUIDELINES, AND REPORTING

Because the vagueness scores show that privacy policies for unregulated companies are notably ambiguous when compared with the federal

13. Interestingly, the unregulated policies are less vague with respect to sharing than the Safe Harbor policies.

benchmark privacy policies for financial institutions, there is a critical need to improve the clarity of online privacy policies. The finding calls for three policy tools. First, technological mechanisms that can easily score the vagueness of large numbers of policies to identify those that are problematic must exist. Second, linguistic guidelines need to be developed so that drafters have a framework to reduce vagueness in policies and so that regulators can point to a set of norms that reduce ambiguity for users. Finally, a framework is needed to assist drafters of privacy policies with reducing vagueness and ambiguity in policies.

5.1. Technical Tools

The ability to identify and score large numbers of privacy policies for vagueness can help drafters recognize and improve the clarity of privacy policies and help regulators identify industries or companies for improvement or enforcement actions. Tools for NLP and machine learning (ML) can provide this functionality. The NLP tools can take annotated privacy policies and, using the taxonomy in Table 2 and the vagueness scoring method based on the coefficients in Figure 3, generate comparable vagueness scores for the policy.¹⁴ The ML tools can be developed, trained, and used to extract policy language relevant to specific data practices and used to annotate policies for analysis.¹⁵ These extracted paragraphs can then be scored separately by NLP tools. These processes would enable drafters to easily identify ambiguity issues in their policies. The cause of vagueness may be due to a desire for flexibility, or it may be due to the policy author's incomplete knowledge about the data practices. In the latter respect, a vagueness score could be used as motivation for conducting internal audits to rationalize the use of less vague language for specific practices. Such audits could improve internal accountability and transparency, which would result in a company's rise to a higher standard of care with regard to data protection and privacy.

Because these processes are automated, they can be easily executed on large numbers of privacy policies. With a database of thousands of privacy policies, scores could be generated automatically and outliers flagged for investigation by relevant regulatory authorities. For example,

14. These natural language processing tools were developed for this paper and will be released as part of the Usable Privacy Project (<https://github.com/cmu-relab/privacy-vagueness>).

15. These ML tools are under development by the Usable Privacy Project and will be released once completed.

the Global Privacy Enforcement Network composed of data privacy regulators around the world conducted a manual sweep of Internet sites to identify transparency issues for privacy (Office of the Privacy Commissioner of Canada 2013). This can be done on a much larger scale using these automated tools.

5.2. Linguistic Guidelines

We propose a set of guidelines designed to yield statements that will, through the use of these NLP tools, measurably show improvement in clarity. The goal of these guidelines is a reduction in the vagueness score.

Four principles can be applied to improve the vagueness score corresponding to clarity. First, avoid terms in Table 2 that are shown to be problematic, specifically generalizations and modal verbs. Second, avoid terms in Table A3 that increase vagueness. Third, use a glossary of key terms so that a company's legal counsel can standardize terminology with a broad range of software developers (website developers, mobile app developers, database administration, and back-end office administration). Finally, signal when the meaning of terms changes in a policy such that automated tools can readily detect the changed meaning.

5.3. Reporting Framework

Beyond these guiding principles, we propose a framework to enable public reporting of the vagueness scores to enhance the public transparency of ambiguous privacy policies. The framework consists of two elements. First, we propose that the vagueness scores represent measurements that are comparable across privacy policies. These scores represent the vague quality of a policy rather than the actual practice of the website or the substantive content of the privacy protections being offered. Second, we propose that the scores for the model privacy form be used as a benchmark standard for acceptable ambiguity against which other policies are measured. In the future, regulators may wish to present new scores and thresholds that companies can seek to achieve through better policy language that consumers can understand.

6. CONCLUSION

If privacy policies are to describe data practices in a meaningful way, the clarity of language is a critical feature. Policies that obfuscate data prac-

tices fail to provide consumers with adequate or appropriate notice of the treatment of their personal information. The measurement and comparison of vagueness across different privacy policies can be used to reveal these failures. On a normative level, the scoring and comparison of ambiguity show that descriptions of data practices suffer a lack of clarity when regulation does not address the kinds of terms to be used in a policy.

And more broadly, the approach and techniques for scoring ambiguity can be generalized for application to other consumer-oriented legal documents. The tools can, for example, be deployed to analyze the clarity of end-user agreements. To deploy the tools in this and other similar contexts, several steps need to be taken. First, some of the terms in Table 3 may be domain specific to privacy policies, and, as a result, these terms need to be updated to reflect the type of document that will be analyzed. Updates may be accomplished by crowdsourcing the review of a handful of documents, since saturation can be achieved without an assessment of large numbers of documents. Second, the set of documents to be scored need annotations that identify the actions and information types to be assessed. These annotations can be crowdsourced or potentially accomplished through the use of ML. Finally, the granular elements that would be valuable to score separately need to be identified. With these elements in place, the scoring tool can then provide vagueness scores for the documents and their subelements. To enhance the interpretation and identify threshold problems, a benchmark document can provide scores for comparison.

APPENDIX: DATA

Table A1. Selected Privacy Policies by Sector

Company	Last Policy Update ^a
Shopping:	
Barnes & Noble	May 7, 2013
Costco	December 31, 2013
JCPenney	May 22, 2015
Lowe's	April 25, 2015
Overstock.com	January 9, 2013
Telecommunications:	
AT&T	September 16, 2013
Charter Communications	May 4, 2009
Comcast	March 1, 2011
Time Warner	September 2012
Verizon	October 2014
Employment:	
CareerBuilder	May 18, 2014
Glassdoor	September 9, 2014
Indeed	2015
Monster	March 31, 2014
Simply Hired	April 21, 2010

Source. Bhatia et al. (2016).

^a These were the most recent versions of each policy as of the date of our analysis.

Table A2. Bradley-Terry Vagueness Coefficients

Category of Vagueness	Coefficient	SE
CN	1.619	.1461
C	1.783	.1458
CM	1.864	.1457
CMN	2.125	.1455
CG	2.345	.1456
CGN	2.443	.1457
MN	2.569	.1458
N	2.710	.1461
M	2.865	.1466
CGMN	2.899	.1467
CGM	2.968	.1470
GN	3.281	.1485
GMN	3.506	.1500
G	3.550	.1503
GM	4.045	.1555

Source. Bhatia et al. (2016).

Note. C = condition; G = generalization; M = modality; N = numeric quantifier.

Table A3. Intracategory Vagueness

Term	Coefficient	SE
Condition:		
As needed	.00	.00
As necessary	.01	.15
As appropriate	.70	.14
Depending	.77	.14
Sometimes	1.20	.15
As applicable	1.37	.15
Otherwise reasonably determined	1.52	.15
From time to time	1.81	.15
Generalization:		
Typically	-.38	.11
Normally	-.34	.11
Often	-.15	.11
General	-.11	.11
Usually	-.04	.11
Generally	.00	.00
Commonly	.03	.11
Among other things	.64	.11
Widely	.67	.11
Primarily	.70	.11
Largely	1.25	.13
Mostly	1.71	.14
Numeric Quantifier:		
Certain	-.53	.22
Most	-1.21	.24
Some	.00	.00
Modality:		
Likely	-.32	.13
May	.00	.00
Can	.42	.13
Would	.60	.13
Might	.76	.13
Could	.96	.14
Possibly	1.78	.15

Source. Bhatia et al. (2016).

Table A4. Frequency of Relevant Vague Terms by Vector

Company	Condition	Generalization	Modality	Numeric Quantifier
Shopping:				
Barnes & Noble	12	4	98	24
Costco	6	7	50	7
JCPenney	6	0	29	12
Lowe's	2	0	62	11
Overstock.com	1	1	19	3
Telecommunications:				
AT&T	3	0	52	15
Charter Communications	8	4	81	41
Comcast	20	9	91	13
Time Warner	1	6	47	31
Verizon	14	1	101	19
Employment:				
CareerBuilder	1	3	28	8
Glassdoor	5	3	42	11
Indeed	0	1	33	6
Monster	3	0	28	2
Simply Hired	1	3	55	12

Source. Bhatia et al. (2016).

Note. The frequency counts reflect each instance in which a word in the taxonomy represented in Table 2 is used in association with a type of personal information and a data-processing action applied to the information. See Breaux and Schaub (2014, pp. 163–72). Completeness = 0 for all policies.

REFERENCES

- Barnes & Noble. 2013. Barnes & Noble Privacy Policy. Revised May 7. <http://www.barnesandnoble.com/u/nook-for-windows-privacy-policy/379003719>.
- Bhatia, Jaspreet, Travis D. Breaux, Joel R. Reidenberg, and Thomas B. Norton. 2016. A Theory of Vagueness and Privacy Risk Perception. Paper presented at the Institute of Electrical and Electronics Engineers 24th International Requirements Engineering Conference, Beijing, September 12–16.
- Bhatia, Jaspreet, Travis D. Breaux, and Florian Schaub. 2016. Privacy Goal Mining through Hybridized Task Re-Composition. *ACM Transactions on Software Engineering Methodology* 25(3), art. 22.
- Breaux, Travis D., and Florian Schaub. 2014. Scaling Requirements Extraction to the Crowd: Experiments on Privacy Policies. Paper presented at the Institute of Electrical and Electronics Engineers 22d International Requirements Engineering Conference, Karlskrona, August 25–29.
- Comcast. 2015. Comcast Customer Privacy Notice. August 1. <http://www.xfinity.com/Corporate/Customers/Policies/CustomerPrivacy.html>.

- Costco Wholesale Corporation. 2013. Your Privacy Rights. December 31. <http://www.costco.com/privacy-policy.html>.
- Department of Commerce. 2015. US-EU Safe Harbor List. <https://safeharbor.export.gov/list.aspx>.
- Farnsworth, E. Allan. 1999. *Contracts*. 3d ed. New York: Aspen.
- Federal Trade Commission. 1998. *Privacy Online: A Report to Congress*. Washington, DC: Federal Trade Commission. <https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf>.
- . 2000. *Privacy Online: Fair Information Practices in the Electronic Marketplace: A Report to Congress*. Washington, DC: Federal Trade Commission. <https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission-report/privacy2000-report/privacy2000.pdf>.
- Glaser, Barney G., and Anselm L. Strauss. 1999. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New Brunswick: Aldine Transaction.
- Hoffman, Paul, Matthew A. Lambon Ralph, and Timothy T. Rogers. 2013. Semantic Diversity: A Measure of Semantic Ambiguity Based on Variability in the Contextual Usage of Words. *Behavioral Research* 45:718–30.
- Jensen, Carlos, and Colin Potts. 2004. Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. Paper presented at the Special Interest Group on Computer-Human Interaction conference Human Factors in Computing Systems, Vienna, April 24–29.
- Levy, Alan, and Manoj Hastak. 2008. *Consumer Comprehension of Financial Privacy Notices: A Report on the Results of the Quantitative Testing*. Interagency notice research project. December 15. <http://www.sec.gov/comments/s7-09-07/s70907-21-levy.pdf>.
- Marotta-Wurgler, Florencia. 2015. Does “Notice and Choice” Disclosure Regulation Work? An Empirical Study of Privacy Policies. Working paper. New York University Law School, New York.
- . 2016. Self-Regulation and Competition in Privacy Policies. *Journal of Legal Studies* 45:S13–S39.
- Massey, Aaron K., Richard L. Rutledge, Annie I. Antón, and Peter P. Swire. 2014. Identifying and Classifying Ambiguity for Regulatory Requirements. Paper presented at the Institute of Electrical and Electronics Engineers 22d International Requirements Engineering Conference, Karlskrona, August 25–29.
- McDonald, Alecia M., and Lorrie Faith Cranor. 2008. The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society* 4:540–65.
- Office of the Privacy Commissioner of Canada. 2013. Privacy Enforcement Authorities Launch First-Ever International Internet Privacy Sweep. May 6. https://www.priv.gc.ca/en/opc-news/news-and-announcements/2013/nr-c_130506/.
- Pollach, Irene. 2007. What’s Wrong with Online Privacy Policies? *Communica-*

tions of the ACM 50:103–8.

Reidenberg, Joel R., N. Cameron Russell, Alexander J. Callen, Sophia Qasir, and Thomas B. Norton. 2015. Privacy Harms and the Effectiveness of the Notice and Choice Framework. *I/S: A Journal of Law and Policy for the Information Society* 11:485–524.

Saldaña, Johnny. 2012. *The Coding Manual for Qualitative Researchers*. London: Sage.

Turner, Heather, and David Firth. 2012. Bradley-Terry Models in R: The Bradley-Terry2 Package. *Journal of Statistical Software* 48(9):1–21.

von Fintel, Kai. 2006. Modality and Language. Pp. 20–27 in *Encyclopedia of Philosophy*, edited by Donald M. Borchert. 2d ed. Detroit: MacMillan Reference.