Accepted To: 2016 IEEE International Conference on Requirements Engineering

A Theory of Vagueness and Privacy Risk Perception

Jaspreet Bhatia[1], Travis D. Breaux[1], Joel R. Reidenberg[2] and Thomas B. Norton[2]
Institute for Software Research, Carnegie Mellon University[1]
Pittsburgh, Pennsylvania, United States
Center on Law and Information Policy, Fordham University School of Law[2]
New York, New York, United States
{jbhatia, breaux}@cs.cmu.edu, {jreidenberg, tnorton1}@law.fordham.edu

***Abstract*—Ambiguity arises in requirements when a statement is unintentionally or otherwise incomplete, missing information, or when a word or phrase has more than one possible meaning. For web-based and mobile information systems, ambiguity, and vagueness in particular, undermines the ability of organizations to align their privacy policies with their data practices, which can confuse or mislead users thus leading to an increase in privacy risk. In this paper, we introduce a theory of vagueness for privacy policy statements based on a taxonomy of vague terms derived from an empirical content analysis of 15 privacy policies. The taxonomy was evaluated in a paired comparison experiment and results were analyzed using the Bradley-Terry model to yield a rank order of vague terms in both isolation and composition. The theory predicts how vague modifiers to information actions and information types can be composed to increase or decrease overall vagueness. We further provide empirical evidence based on factorial vignette surveys to show how increases in vagueness will decrease users' acceptance of privacy risk and thus decrease users' willingness to share personal information.**

***Index Terms*—vagueness, hedging, natural language processing, privacy, risk perception.**

I. Introduction

Companies and government agencies use personal information to improve service quality by tailoring services to individual needs. To support privacy, regulators rely on the privacy notice requirement, in which organizations summarize their data practices to increase user awareness about privacy. These notices, also called privacy policies, further serve to align company privacy goals with government regulations. The underlying vagueness in privacy policies, however, undermines the utility of such notices as effective regulatory mechanisms. Consequently, privacy policies also fail to offer a clear description of the organization's privacy practices to users.

Privacy policies pose a challenging requirements problem for organizations, because policies must: (a) be *comprehensive*, which includes describing data practices across physical places where business is conducted (e.g., stores, offices, etc.), as well as web and mobile platforms; and (b) be *accurate*, which means all policy statements must be true for all data practices and systems. Ensuring privacy policies are comprehensive and accurate means that policy authors can resort to vagueness when summarizing their data practices. Variations in data practices may exist because two or more current practices that are semantically different must be generalized into a broader category of statement. In Figure 1, the data "shipping address" and "ZIP code" are generalized into "address information," and the purposes "order fulfillment" and "marketing purposes" are combined into a vague condition "as needed," to encompass both practices. To account for future practices, a vague modal verb "may" is added to the general policy statement, while "address" is subsumed by "location information."

Vagueness can introduce privacy risks, because the flexibility entailed by vague policy statements may conceal privacy-threatening practices. Moreover, vague statements can limit an individual's ability to make informed decisions about their willingness to share their personal information, which may increase their perceived privacy risk. To ensure accuracy, we believe business analysts and system developers, in addition to legal advisors, must participate in deciding which practices to summarize in a privacy policy, and when to use vagueness.

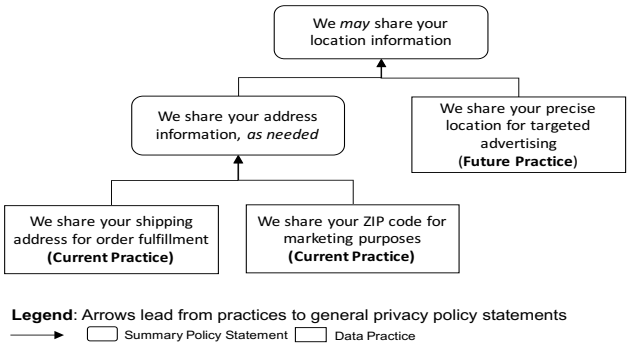

Fig. 1. Example data practices that are generalized into privacy policy statements

Creswell defines a theory as an interrelated set of constructs formed into propositions and hypothesis that specify the relationship among variables, typically in terms of magnitude and direction [14]. To that end, the contributions in this paper include a four-part theory: (1) the construct vagueness is described by multiple, exclusive semantic categories; (2) the categories, independently and through composition, predict how vagueness increases and decreases; (3) semantic functions, called likelihood, authority and certitude, suggest why semantic categories predict vagueness; and (4) as privacy statement vagueness increases, a person's willingness to share personal information decreases. The theory provides an early, novel foundation upon which to improve the summarization of data

practices and readability of privacy policies, which are known to be hard to read [38], and it aims to enhance emerging techniques for automating the extraction of privacy goals [8].

This paper is organized as follows: in Section II, we review vagueness, risk and related work; in Section III, we present our approach to discover a theory of vagueness using content analysis and paired comparison, and to study perceived privacy risk using factorial vignettes; in Section IV, we present our results; in Section V we present threats to validity, and in Section VI, we discuss our results and future work.

II. Background and Related Work

We now review vagueness, risk and related work.

A. Vagueness in Natural Language

The use of vague terms, such as *may, as necessary,* and *generally*, to describe goals in privacy policies introduces uncertainty into the goal's *action* or the associated *information type*. Consider the following statements:

1. We will share your personal information, such as your name, email address and phone number, with our marketing affiliates for advertising purposes.
2. We might share some of your personal information with our third party affiliates as necessary.

In the first statement, the modal phrase *will* is certain, whereas the modal phrase *might* in the second statement leaves open the possibility of sharing, and is thus vague. In addition, the first statement elaborates upon what personal information is included, *name, email address and phone number*, which adds additional clarity missing from the second statement, which mentions sharing *some of your personal information*. Similarly, the description of the purpose *advertising purposes* is more clear than the phrase *as necessary*, which leaves open a range of possible purposes, such as legal, marketing, etc.

Table I presents Massey et al.'s ambiguity taxonomy that was applied to natural language legal texts [37]. In this paper, we focus on vagueness from the use of vague terms.

TABLE I. Ambiguity Categories in Natural Language

Type	Definition
Lexical	a word or phrase with multiple, valid meanings, also called polysemy
Syntactic	a sequence of words with multiple valid grammatical interpretations regardless of context
Semantic	a sentence with more than one interpretation in its provided context
Vagueness	a statement that admits borderline cases or relative interpretation
Incompleteness	a grammatically correct sentence that produces too little detail to convey a specific or needed meaning
Referential	a grammatically correct sentence with a reference that confuses the reader based on the conduct

B. Risk Perception and Privacy Risk

Risk is a multidisciplinary topic that spans marketing, psychology, and economics. In marketing, risk is defined as a choice among multiple options, which are valued based on the likelihood and desirability of the consequences of the choice [7]. Starr, an engineer by training, first proposed that risk preferences could be *revealed* from economic data, in which both effect likelihood and magnitude was previously measured (e.g., the acceptable risk of death in motor vehicle accidents) [48]. In psychology, Fischhoff et al. note that, so-called revealed preferences assume that past behavior is a predictor of present-day preferences, which cannot be applied to situations where technological risk or personal attitudes are changing [22]. To address these limitations, the psychometric paradigm of perceived risk emerged in which surveys are designed to measure personal attitudes about risks and benefits [49]. Two insights that emerged from this paradigm and inform our approach are: (a) people better accept technological risks when presented with enumerable benefits, and: (b) perceived risk can account for benefits that are not measurable in dollars, such as lifestyle improvements [49]. In other words, people who see technological benefits are more inclined to see lower risks than those who do not see benefits. Notably, privacy is difficult to quantify, as evidenced by ordering effects and bimodal value distributions in privacy pricing experiments [4]. Rather, privacy is more closely associated with lifestyle improvements, e.g., private communications with friends and family, or the ability to avoid stigmatization. Finally, the economist Knight argues that subjective estimates based on partial knowledge represent uncertainty and not risk [34].

C. Ambiguity and Requirements

Lakoff noted that NL concepts have vague boundaries and fuzzy edges. Consequently, he introduced the term *hedging* to describe the fuzziness in the truth value of NL sentences, meaning, that they are true to a certain extent, and false to a certain extent, true in certain respects and false in certain other respects [35]. In NLP, ML systems have been developed as part of the CoNLL-2010 shared task to identify hedge cues and their scopes in Wikipedia and Biomedical texts [15].

Requirements are often written in natural language (NL) and thus suffer from inherent NL ambiguity [9]. For example, Yang et al. report that, out of the 26,829 requirements statements that they analyzed, 12.7% had ambiguity due to a coordinating conjunction (and/or), which is a type of syntactic ambiguity [55]. Ambiguity is often considered a potentially dangerous attribute of requirements [12]. Gause and Weinberg note that ambiguity in requirements can lead to *subconscious disambiguation*, wherein readers disambiguate using their first interpretation, unaware of other possible interpretations [25]. This leads different stakeholders with different interpretations of the same requirements. Ambiguity detection is difficult, even if the reader is aware of all the facets of ambiguity [29].

Many attempts have been made previously to address the problem of ambiguity in requirements. Fuchs and Schwitter propose Attempto Controlled English, a restricted NL, to align NL specifications with first order logic to reduce the ambiguity in requirements [21]. However, restricted or formal languages are not as expressive as NL, and incorrectly interpreted NL specifications lead to incorrect formal specifications [50]. Alternatively, Berry et al. introduced the Ambiguity Handbook, which describes ambiguity in requirements and legal contracts, including strategies for avoiding and detecting ambiguity [9].

Pattern based techniques have also been used to identify ambiguity in requirements [30, 18]. Kiyavitskaya et al. propose a tool that combines lexical and syntactic measures applied to a semantic network to identify ambiguous sentences and determine potential ambiguities [33]. Alternatively, object oriented analysis models of the specified system can be used to identify ambiguities [44]. Tjong describes ambiguities found in NL requirements, such as lexical ambiguity, ambiguity due to uncertainty, etc., and guidelines to avoid these ambiguities [50]. The tool called SREE identifies instances of a set of vague words using simple keyword matching and marks it as potentially ambiguous [51]. In our approach, we do not employ keyword matching, because we do not consider all instances of a vague term to be potentially vague (see Section III.A). Instead, we rely on manual annotations to identify vague terms. Requirements quality evaluation tools, such as IBM Doors and QuARS [20] and ARM [53], also identify ambiguous terms. Yang et al. identify speculative requirements and uncertainty cues, using a technique that combines machine learning (ML) and a rule-based approach. They utilize lexical and syntactic features of requirements to identify uncertainty [57]. More recently, researchers have used ML based on heuristics drawn from human judgments to identify nocuous coordination and anaphoric ambiguities in requirements [55, 56]. This approach still requires human interpretation to resolve ambiguity. To our knowledge, this prior work to identify vague requirements terms [9, 30, 50, 51, 20, 53, 57] does not differentiate the relative vagueness of these terms. We address this limitation with a new vagueness taxonomy and predictions of how vague terms increase and decrease vagueness.

III. Vagueness and Risk Perception Study Designs

We now introduce our research questions and three study designs based on content analysis, paired comparisons and factorial vignettes. Our research questions are as follows:

RQ1. What are the different categories of terms in privacy policies that lead to vagueness or lack of clarity?

RQ2. How does the relative vagueness vary within and across different categories of vague terms and their combinations?

RQ3. How do vagueness and risk likelihood affect the overall privacy risk perceived by users and their willingness to disclose their personal information?

Next, we describe our three study designs to answer the above research questions.

A. Content Analysis of Vague terms

Research question RQ1 is exploratory and asks how vagueness appears in privacy policies "in the wild." To answer RQ1, we manually annotated 15 privacy policies (see Table II) using content analysis [47] to identify words or phrases that introduce vagueness into policy statements. We limited our analysis to statements about collection, use, disclosure and retention of personal information, which have also been discussed by Antón and Earp [5]. These policies are part of a convenience sample, although, we include a mix shopping companies who maintain both online and "brick-and-mortar" stores, and we chose the top employment websites and Internet service providers in the U.S. Table II presents the 15 policies by category and date last updated.

TABLE II. Privacy Policy Dataset For Vagueness Study

Company's Privacy Policy	Industry Category	Last Updated
Barnes and Noble	Shopping	05/07/2013
Costco	Shopping	12/31/2013
JC Penny	Shopping	05/22/2015
Lowes	Shopping	04/25/2015
Over Stock	Shopping	01/09/2013
AT&T	Telecom	09/16/2013
Charter Communication	Telecom	05/04/2009
Comcast	Telecom	03/01/2011
Time Warner	Telecom	09/2012
Verizon	Telecom	10/2014
Career Builder	Employment	05/18/2014
Glassdoor	Employment	09/09/2014
Indeed	Employment	2015
Monster	Employment	03/31/2014
Simply Hired	Employment	4/21/2010

The policies are first prepared by removing section headers and boilerplate language that does not describe relevant data practices, before saving the prepared data to an input file for an Amazon Mechanical Turk (AMT) task. The task employs an annotation tool developed by Breaux and Schaub [11], which allows annotators to select relevant phrases matching a category, in this case, the vague terms belonging to a certain category. The first and fourth authors, and a graduate law student, performed the annotation task.

The annotation process employs two-cycle coding [47]. In the first cycle, the first author analyzed five policies to identify an initial set of vague terms, and then applied second-cycle coding to group these terms into emergent categories based on the kind of vagueness introduced by related terms. In addition, guidelines were developed to predict into which category a vague term should be placed. The terms, categories and guidelines were shared with the other two annotators, who independently annotated the same five policies. Next, the three annotators met to discuss results, to add new terms to the categories and to refine the guidelines. After agreeing on the categories and guidelines, the three annotators annotated the remaining ten policies, before meeting again to reconcile disagreements. Saturation was reached after no new vague terms or new categories were discovered, which occurred after analyzing the first five policies (Barnes and Noble, Lowes, Costco, AT&T, and Comcast).

The resulting vagueness categories and their definitions are:

- *Conditionality* – the action to be performed is dependent upon a variable or unclear trigger
- *Generalization* – the action or information types are vaguely abstracted with unclear conditions
- *Modality* – the likelihood or possibility of the action is vague or ambiguous
- *Numeric Quantifier* – the action or information type has a vague quantifier

This approach is also knows as grounded theory in literature [47]. The guidelines help disambiguate the policy statement in a given context, for example, the phrase "as necessary" when

followed by a specific purpose: "We will use your personal information as necessary for law enforcement purposes" states that the information is used for legal purposes, thus disambiguating the condition "as necessary" in this context.

We use the semi-automated privacy goal-mining framework developed by Bhatia et al. to identify statements with privacy goals [8]. This technique was extended to use the Stanford Dependency Parser [39] to automatically identify which annotated vague terms are attached to either an *action* or *information type* in the privacy goal. The resulting vagueness dataset consists only of privacy goals with a vague term attached to either the action or information type.

We applied Fleiss' Kappa, an inter-rater agreement statistic [23], to the annotations-vagueness category mappings. Because Fleiss' Kappa assumes that categories are exclusive, we compute the Kappa statistic for the complete composition of all vagueness categories assigned to each policy statement. A statement that contains one or more *Modality* category terms is assigned to the singleton category M, whereas a statement with terms from a combination of the *Conditionality*, *Generality* and *Modality* categories is assigned to the composite category CGM. The Fleiss Kappa for all mappings from annotations to vagueness categories and the three annotators was 0.94, which is a very high probability of agreement above chance alone.

B. Ranking Vagueness Categories in Paired Comparisons

The RQ2 asks how vagueness varies within and across categories and their combinations. Paired comparison is a statistical technique used to compare *N* different items by comparing just two items at once [17]. The overall results are computed by combining data from all paired comparisons. This technique is especially useful when items are comprised of multiple factors, when the comparison context is difficult to control, or when the comparison order influences the outcome. This technique is beneficial when differences between items are small, and when comparison between two items should be as free as possible from any extraneous influence caused by the presence of other entities [17]. To compare N entities, a total $N * (N - 1)/2$ paired comparisons are performed.

We designed multiple surveys to compare combinations of one or more vague terms, within and across the four vagueness categories. The first survey is an exploratory survey designed to compare statements containing combinations of vague terms from across the four vagueness categories (see Section III.A). We chose one exemplary vague term from each category. The vague terms were then inserted into a baseline privacy policy statement: "*We share your personal information.*" For example, variants 1 and 2 below show two statements that result from inserting the underlined vague terms selected from the corresponding vagueness categories (in parenthesis):

Variant 1 (Modality, Condition): *We <u>may</u> share your personal information <u>as necessary</u>.*

Variant 2 (Numeric Quantifier): *We share <u>some</u> of your personal information*.

For four vagueness categories, we have 2^4-1 or 15 category combinations and thus one statement variant per combination. The 15 statement variants yield 105 paired comparisons.

The survey consists of a scenario, and five of 105 paired comparisons (see Figure 2). The scenario frames the survey rationale for the participants.

Instructions: A company wants to improve the clarity of their website privacy policies. Therefore, they are considering alternative language to help users better understand what their data practices are. For each numbered question, please read each pair of statements, and identify which of the two statements best represents a more clear description of the company's treatment of personal information.

For example, a clear description of the company's treatment of personal information could be "*We share your personal information such as your name and contact details, as needed for legal purposes*."

In the following statement, any pronouns "We" or "Us" refer to the company, and "you" refers to the user.

1. Which one of the following statements is a more clear description of the company's treatment of personal information than the other?

- We may share your personal information.
- We share some of your personal information, as needed.

Fig. 2. Paired Comparison Survey Question

The number of participants needed to judge each paired comparison was based on Pearson and Hartley's data for calculating power for paired comparisons [41, 42]. To attain 95% power, at least four participants are needed to judge each paired comparison. We solicited 60 participants to judge each paired comparison. The additional 56 participants only reduce standard error to further delineate between vagueness levels; four participants are sufficient to discover rank order.

We designed four additional surveys based on the design shown in Figure 2 to measure intra-category vagueness. For the intra-category vagueness surveys, each survey has a total $N * (N - 1)/2$ paired comparisons for N vague terms in the corresponding vagueness category.

The research question RQ2 is answered using the Bradley Terry model, which estimates the probability that one item is chosen over another item using past judgments about the items [17, 28]. Model fitting is either by maximum likelihood, by penalized quasi-likelihood (for models which involve a random effect), or by bias-reduced maximum likelihood in which the first-order asymptotic bias of parameter estimates is eliminated [52]. The Bradley Terry model has been implemented using statistical R package [45, 52].

C. Vagueness, Risk and Factorial Vignettes

Research question RQ3 asks whether changes in statement vagueness correspond to changes in perceived risk. Factorial vignettes provide a method to measure the extent to which discrete factors contribute to human judgment [6]. The factorial vignette method employs a detailed scenario with multiple factors and their corresponding levels, designed to obtain deeper insights, into a person's judgment and decision principles, than is possible using direct questions (i.e., with a prompt "Please rate your level of perceived risk" and a scale). Our factorial vignette survey design measures the interactions between two independent variables, *vagueness* and *likelihood of privacy violation*, and their effect on a dependent variable, the Internet user's *willingness to share their personal information*. This includes whether vagueness or likelihood of violation alone, or neither of these two factors affect willingness to share.

For this study, we chose to control several factors that affect willingness to share. For example, Nissenbaum argues that privacy and information sharing are contextual, meaning that

the factors, data type, data recipient, and data purpose, affect willingness to share [40]. We chose to control these factors by examining a single context that many Internet users engage in: shopping for products online [27]. In addition, Fischhoff et al. argue that individuals should be presented with enumerable benefits before judging the risk of a specific event [22]. We conducted a brief one-hour, four-person focus group to elicit benefits of online shopping (as opposed to visiting a physical store), without considering potential harms of online shopping. The elicited benefits include: convenience, discounts and price comparisons, anonymous and discreet shopping, certainty that the product is available, wider product variety, and informative customer reviews.

When measuring risk, Fischhoff et al. recommend using ratios to represent probabilities, because lay people can better map ratios to physical people then they can map probabilities to people affected [22]. We pilot tested a risk factor with ratio-based levels and found no significant effects, suggesting that participants cannot distinguish among ratios. Alternatively, construal-level theory shows that people correlate larger spatial, temporal, social and hypothetical distances with increased unlikelihood than they do with shorter psychological distances along these four dimensions [54]. Thus, we designed our risk likelihood scale to combine spatial and social distance as a correlate measure of likelihood (see Table III): a privacy harm affecting *only one person in your family* is deemed a psychologically closer and more likely factor level than "one person in your city" or *one person in your country*, which are more distal and perceived less likely.

TABLE III. VIGNETTE FACTORS AND THEIR LEVELS

Factors	Levels
Risk Likelihood ($RL)	only one person in your family
	only one person in your workplace
	only one person in your city
	only one person in your state
	only one person in your country
Vague Statement ($VS)	(C) We share your personal information as necessary.
	(G) We generally share your personal information.
	(M) We may share your personal information.
	(N) We share some of your personal information.

Factorial vignettes are presented using a template in which factors correspond to independent and dependent variables and each factor takes on a level of interest. The two independent factors are Risk Likelihood and Vague Statement with the levels described in Table III. Figure 3 shows the vignette template: for each participant, each factor is replaced by one level. Because the independent variables are within-subjects factors, each participant sees and responds to all combinations of levels (4x5=20). Within-subject designs reduce subject-to-subject variability thereby increasing power.

For each vignette, participants rate their willingness to share their personal information on an eight-point, bipolar semantic scale, labeled: *Extremely Willing, Very Willing, Willing, Somewhat Willing, Somewhat Unwilling, Unwilling, Very Unwilling* and *Extremely Unwilling*. This scale omits the midpoint, such as "Indifferent" or "Unsure," which produce scale attenuation wherein responses are prone to cluster, and these midpoints are more indicative of a vague or ambiguous context than of the respondent's attitude [32].

Please rate your willingness to share your personal information with a shopping website you regularly use, given the following benefits and risks of using that website.

Benefits: convenience, discounts and price comparisons, anonymous and discreet shopping, certainty that the product is available, wider product variety, and informative customer reviews

Risks: In the last 6 months, **$RiskLikelihood** experienced a privacy violation while using this website.

When choosing your rating, given the above benefits and risks, also consider the following website's privacy policy statements. Website privacy policies are intended to protect your personal information.

	Extremely Willing	Very Willing	Willing	Somewhat Willing	Somewhat Unwilling	...
$VagueStatement	○	○	○	○	○	

Fig. 3. Template used for vignette generation (fields with $ sign are replaced with values selected from Table III)

Before the vignettes, participants are presented a pre-survey to elicit their demographic characteristics (gender, age, race, education, income) and frequency of online behavior in six activities: using social networking sites; shopping for products or services; paying bills, checking account balances, or transferring money; searching for health information; using dating websites; and searching for jobs. The semantic scale response options for frequency of online behavior are: *a few times a day, once a day, few times a week, few times a month, few times a year,* and *never*.

Multi-level modeling is a statistical regression model with parameters that account for multiple levels in datasets, and limits the biased covariance estimates by assigning a random intercept for each subject [24]. Multi-level modeling has been used to study interactions among security requirements [26]. In our study, the main dependent variable of interest is willingness to share, labeled `$WtS` in our model. The two fixed independent variables, which are within-subject factors, are risk likelihood labeled `$RL` (with five levels) and vague statement labeled `$VS` (with four levels). The independent exploratory variable `$Shopping` is based on the pre-test online behavior question about online shopping frequency and has two levels: *S1* for participants who shop online a few times a week or more, and *S0* for participants who shop less than a few times a week. For the within-subject design, subject-to-subject variability is accounted for by using a random effect variable `$PID`, which is unique to each participant.

The data is analyzed in R [45] using the package lme4 [10]. Each participant sees all 20 combinations of our two within subject factors. Thus, our analysis accounts for dependencies in the repeated measures, calculates the coefficients (weights) for each explanatory independent variable, and tests for interactions. We test the multi-level models' significance using the standard likelihood ratio test: we fit the regression model of interest; we fit a null model that excludes the independent variables used in the first model; we compute the likelihood ratio; and then, we report the chi-square, p-value, and degrees of freedom [24]. We performed *a priori* power analysis using G*Power [19] to test for the required sample size for repeated measures ANOVA. The power analysis estimate is at least two participants per combination of the within-subject factors to achieve 95% power, and a medium effect size [13].

IV. Results

We now describe our results from the three studies.

A. Vagueness Taxonomy from Content Analysis

Table IV shows the content analysis results applied to the 15 policies in Table II: the categorization was done by the first author and checked by the other two annotators. The frequency represents the number of times the term appeared across all selected statements in the 15 policies. Table V presents a breakdown of number of terms per category that appear across all 15 policies and the privacy goal types present in the policy (C: Collection, R: Retention, T: Transfer, U: Use).

TABLE IV. Taxonomy of Vague Terms

Category	Vague terms	% Freq.
Conditionality (C)	depending, necessary, appropriate, inappropriate, as needed	7.9%
Generalization (G)	generally, mostly, widely, general, commonly, usually, normally, typically, largely, often	4.0%
Modality (M)	may, might, can, could, would, likely, possible, possibly	77.9%
Numeric Quantifier (N)	certain, some, most	10.1%

TABLE V. Frequency of Vague Terms Across Policies

	Policy	Vagueness				Goal Types			
		C	G	M	N	C	R	T	U
Shopping	Barnes & Noble	12	4	98	17	55	7	47	48
	Costco	6	7	50	1	47	12	70	43
	JC Penny	6	0	29	5	31	2	31	30
	Lowes	2	0	62	6	61	16	16	54
	OverStock	1	1	19	3	9	2	10	14
Telecom	AT&T	3	0	52	0	41	4	47	77
	Charter Comm.	8	4	81	12	46	16	70	48
	Comcast	20	9	91	9	30	18	68	56
	Time Warner	1	6	47	18	24	12	29	27
	Verizon	14	1	101	12	57	13	83	87
Employment	Career Builder	1	3	28	4	24	14	13	52
	GlassDoor	5	3	42	6	30	13	19	34
	Indeed	0	1	33	4	19	13	25	57
	Monster	3	0	28	1	31	20	23	38
	Simply Hired	1	3	55	8	37	9	12	44

B. Vagueness Rankings using Paired Comparisons

In Section III.B, we describe a method for rank ordering exemplar terms selected from each vagueness category to answer research question RQ2, how does vagueness vary within and across categories, and how do vague terms interact in combination to affect overall vagueness. The selected terms are *as needed* (C), *generally* (G), *may* (M), and *some* (N). The survey was conducted on AMT, and each paired comparison was judged by 60 participants, who were paid $0.12 to judge five paired comparisons at once. We analyze the paired comparisons using the Bradley-Terry (BT) model; the BT model coefficients and standard error appear in Table VI.

Figure 4 presents the BT coefficients and standard error in an annotated scatter plot to show the linear relationship of vagueness categories and their combination. The coefficients show the quantity that each vague term contributes to the overall concept of vagueness. The data practices described with combinations to the left of Figure 4 (CN, C, CM, …) have greater clarity than practices described with combinations to the right of Figure 4 (GMN, G, GM, …).

TABLE VI. Bradley Terry Coefficients

Vagueness Category	Coefficient	Standard Error
CN	1.619	0.146
C	1.783	0.146
CM	1.864	0.146
CMN	2.125	0.146
CG	2.345	0.146
CGN	2.443	0.146
MN	2.569	0.146
N	2.710	0.146
M	2.865	0.147
CGMN	2.899	0.147
CGM	2.968	0.147
GN	3.281	0.149
GMN	3.506	0.150
G	3.550	0.150
GM	4.045	0.156

C: Conditionality, G: Generality, M: Modality, N: Numeric Quantifier

For example, while phrases with both a conditional term and numeric quantifier (CN) are statistically indistinguishable compared to phrases with only a conditional term (C), we observe how the vagueness taxonomy influences overall vagueness. In Figure 4, the red arrow from MN to CMN shows a condition term increases clarity and reduces vagueness: statements with both a modal term and numerical quantifier (MN) are significantly less clear than similar statements with an added conditional term (CMN). The blue arrow from MN to GMN shows how generalization increase vagueness: the MN statements with the added generalization (GMN) are significantly more vague. By comparison, statements with a generalization and modal term (GM=4.045) are twice as vague as statements with a condition and a modal term (CM=1.864).

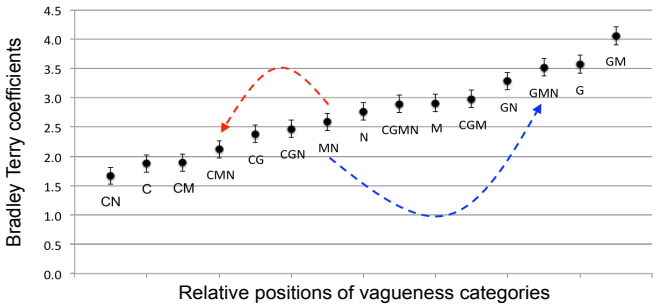

Fig. 4. Bradley Terry Coefficients

Table VII presents the BT coefficients for intra-category vagueness: the shaded rows present the model intercepts, which consist of the vague terms in the inter-category survey. In the Conditionality category, "as appropriate" was several times more vague than "as necessary". Under Generality, the vagueness appears to increase as the adverbs transition from the routine (e.g., typical, normal or usual) to the unrestricted (e.g., widely, largely, mostly). Under Modality, the past tense verbs "might" and "could" are perceived to be more vague than the present tense variants, "may" and "can", respectively.

TABLE VII. BRADLEY TERRY COEFFICIENTS FOR INTRA-CATEGORY VAGUENESS

	Vague term	Coefficient	Standard Error
Conditionality	as needed	0.00	0.00
	as necessary	0.01	0.15
	as appropriate	0.70	0.14
	depending	0.77	0.14
	sometimes	1.20	0.15
	as applicable	1.37	0.15
	otherwise reasonably determined	1.52	0.15
	from time to time	1.81	0.15
Generality	typically	-0.38	0.11
	normally	-0.34	0.11
	often	-0.15	0.11
	general	-0.11	0.11
	usually	-0.04	0.11
	generally	0.00	0.00
	commonly	0.03	0.11
	among other things	0.64	0.11
	widely	0.67	0.11
	primarily	0.70	0.11
	largely	1.25	0.13
	mostly	1.71	0.14
Num. Q.	certain	-0.53	0.22
	most	-1.21	0.24
	some	0.00	0.00
Modality	likely	-0.32	0.13
	may	0.00	0.00
	can	0.42	0.13
	would	0.60	0.13
	might	0.76	0.13
	could	0.96	0.14
	possibly	1.78	0.15

C. Vagueness and Privacy Risk Perception Results

The research question RQ3 asks how vagueness and risk likelihood affect user willingness to share personal information. We recruited 102 participants using AMT, where we paid $3 to completing the survey. We now discuss our results from the privacy risk perception survey (see Section III.C).

1) Descriptive Statistics

A total 102 participants responded to our risk perception survey: 45.1% are female and 54.9% are male; 84.3% reported "white" as their ethnicity; 87.3% reported having at least some college level education; and 84.3% reported having annual household income less than $75,000. Figure 5 shows frequency of online behavior by participants. While 70% of respondents report viewing social networking sites daily, while 33% in a separate survey reported sharing personal information on these sites a few times a week or more.

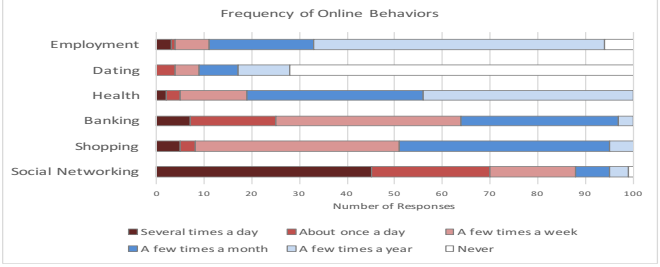

Fig. 5. Frequencies of Online Behaviors

2) Willingness to Share

Equation 1 below is our main additive regression model with a random intercept grouped by participant's unique ID, the independent within-subjects measure `$RL`, which is the likelihood of a privacy violation, and `$VS`, which is the vague privacy statement with a single vague term from one of the four categories (see Table III in Section III.C). The additive model is a formula that defines the dependent variable `$WtS`, willingness to share, in terms of the intercept α and a series of components, which are the independent variables. Each component is multiplied by a coefficient (β) that represents the weight of that variable in the formula. The formula in Eq. 1 is simplified as it excludes the dummy (0/1) variable coding for the reader's convenience.

$$\$WtS = \alpha + \beta_R \$RL + \beta_V \$VS + \epsilon \quad (1)$$

To compare dependent variable `$WtS` across vignettes, we establish the baseline level for the factor `$RL` to be "*only one person in your family*" who experiences the privacy violation and, for the factor `$VS`, we set the vagueness category to Condition, "*We share your personal information as needed*". The intercept (α) is the value of the dependent variable, `$WtS`, when the independent variables, `$RL` and `$VS` take their baseline values.

We found a significant contribution of the two independent factors, for predicting the `$WtS` ($\chi^2(7)=875.15$, $p<0.000$), over the null model, which did not have any of the independent variables. In our model, we did not observe any effect of the interaction term `$RL*$VS`, ($\chi^2(12)=4.7$, $p=0.97$), which means vagueness and risk likelihood did not interact to affect the willingness to share. In Table VIII, we present the model *Term*, the corresponding model-estimated *Coefficient* (along with the p-value, which tells us the statistical significance of the term over the corresponding baseline level), and the coefficient's *Standard Error*. In our survey, the semantic scale option *Extremely Unwilling* has a value of 1, and *Extremely Willing* has a value of 8. A positive coefficient in the model signifies an increase in willingness to share and a negative coefficient signifies a decrease in willingness to share.

TABLE VIII. MULTILEVEL MODELING RESULTS

Term	Coeff.	Stand. Error
Intercept (Family+Condition)	3.133***	0.164
Risk - only 1 person in your workplace	0.162*	0.080
Risk - only 1 person in your city	0.968***	0.080
Risk - only 1 person in your state	1.517***	0.080
Risk - only 1 person in your country	2.118***	0.080
Vagueness - generalization	-0.729***	0.072
Vagueness - modal	-0.155*	0.072
Vagueness - numeric	-0.218**	0.072

*p≤.05 **p≤.01 ***p≤.001

The results in Table VIII show that `$WtS` is significantly different and increasing for decreasing levels of `$RL`, as compared to the baseline level "*only 1 person in your family*". For the `$RL` level "*only 1 person in your workplace*", the `$WtS` increases by 0.16 over the baseline level, which is "*only 1 person in your family*", which denotes an increasing willingness to share. For the baseline `$VS` level "Condition," however, the `$WtS` is at the maximum. The `$VS` level

"*Generalization*" shows a 0.73 decrease in the value of the dependent variable $WtS, as compared to the baseline level, which means generalization reduces the willingness to share.

3) Effect of the Online Behavior Shopping

We computed a new, two-level independent exploratory variable $Shopping based on the participant responses to the online behavior questions. The two levels correspond to the frequency that respondents shop online: *S1*, which is a few times a week or more, and *S0*, which is less than a few times a week. The new additive model in Eq. 2, below, has a component for the $Shopping variable. The new model in Eq. 2 improves the prediction of the $WtS over the model in Eq. 1 ($\chi^2(1)$=4.3, p<0.05), which means respondents who shop more often express increased certainty about their willingness to share their personal information.

$$\$WtS = \alpha + \beta_R \$Rl + \beta_V \$Vs + \beta_S \$Shopping + \epsilon \quad (2)$$

We found that participants who shop online a few times a week or more, are also more willing to share their personal information ($WtS is 0.62 higher than other participants), which means they may be more likely to comprehend the presented benefits of shopping while evaluating the risk. We discuss these results in more depth in Section VI.

V. Threats to Validity

Construct validity addresses whether what we measure is actually the construct of interest [58]. To mitigate threats to construct validity, multiple annotators participated in identifying vague terms during the content analysis, and the annotators met twice to discuss guidelines and reconcile differences. The Fleiss' Kappa statistic K=0.94 shows a very high degree of agreement above chance. For the risk survey, we conducted multiple rounds of pilot testing and we focus-grouped the benefits and risk likelihood levels. The likelihood levels were further motivated by an foundational theory of psychological distance, which has been validated in multiple studies [54]. The theory predicts that spatial and social distance strongly correlate with perceived event likelihood. While we are measuring perceived risk, similar to Fischhoff [22], we assume that a person's willingness to disclose corresponds to their acceptance of the risk; this assumption was used in other study designs by Acquisti and Kobsa to measure privacy-related risk [1, 31]. The semantic scale anchor labels used for $WtS in the factorial vignettes could be interpreted differently by participants [16]. To address this threat, we designed both the fixed effect independent factors $RL and $VS as within-subject factors, so that all participants respond to all levels of these variables. During multi-level modeling, we account for subject–to-subject variability using the random effect variable $PID. Another way to address this threat is to conduct surveys to calibrate the scale options for the dependent variable $WtS.

Internal validity concerns whether our correlation of the effects with the conclusion is valid [58]. With respect to the rank order of vagueness categories, the paired comparison limits judgments to two items at a time, rather than comparing multiple entities at the same time, to avoid the confounding effect caused by the presence of other entities [2, 17]. In our risk perception study, we randomized the order of online behavior questions and of vignettes. To address fatigue effects, we limit pairs to a maximum of five comparisons per question set and allow participants to complete as many or as few as they prefer. The participants spent an average 10.4 minutes to complete the factorial vignette survey.

External validity refers to the extent to which we can generalize the results to other situations [58]. In our study, we analyzed 15 policies from three domains: shopping, telecom and employment. Other policies not included in the 15 policies may contain other vague terms that were not present in our taxonomy. Hence, we believe that our taxonomy is complete for the policies that we analyzed, but new vague terms would need additional evaluation. Furthermore, replication would be needed to see if these vague terms affect risk perception in other domains, such as security. Our target population is the average U.S. Internet user. We recruited participants from AMT who have a 95% approval rating or higher, and between 1000-5000 HITs completed. Demographically, our participant population deviates from other measures of U.S. Internet users. We had less reported Asian, Black and Hispanic participants, and more White participants than were found in 2014 Census data and the 2015 PEW Internet and American Life Survey of Internet users [43]. This sample may skew privacy risk perceptions measured in our study that are influenced by race.

VI. Discussion and Summary

We now discuss our results and their impact on improving awareness of privacy practices and for privacy goal extraction. The terms in the vagueness taxonomy are associated with two semantic roles: the action performed on the information and the information type. While we did not observe an interaction between risk likelihood and vagueness on willingness to share personal information, there may be an interaction with respect to specific roles, e.g., vague disclosure recipients may be perceived as higher risk ambiguities, than the type of information disclosed. From the inter- and intra-category vagueness results, we theorize that differences in clarity may be due to one of three semantic functions: *likelihood*, which is the possibility that something is true; *authority*, which is whether an action is discretionary or mandatory; and *certitude*, which is the absoluteness with which something is true. For example, "likely" is more clear than "possibly," both of which concern the degree or likelihood that a data practice occurs. Authority refers to whether the practice is permitted, required or prohibited, and it may be true that *required* practices are perceived as more clear than *permitted* practices: "as needed" is perceived as more clear than "as appropriate." Similarly, the vague term "may" denotes both *permissibility* and *possibility*, and is perceived to be more clear than "can," which denotes *capability* and not necessarily *authority*. Concerning certitude, "as needed" and "normally" describe minimal versus routine behavior, respectively. These two vague terms may have a higher degree of absoluteness than "generally," which assumes the existence of unstated exceptions, and which is perceived to be more vague and less clear than "as needed" and "normally."

We conclude from the results that willingness to share increases as a participant's social and physical distance from

the person experiencing the privacy violation (`$RL`) increases. This means that the users' perception of privacy risk increases, when they think about a person from their family or workplace experiencing the violation, as compared to the experience of a person somewhere in their state or country. We also found that the willingness to share is highest for the least vague category Condition, as compared to other vague categories, and willingness to share was the lowest for Generalization, which is the most vague category in Figure 4, Table VIII. Furthermore, there was no statistically significant difference between willingness to share for Modality and Numeric Quantifier (p=0.38), which have similar vagueness measures. The inverse decrease in willingness to share due in the presence of increased vagueness is in contrast to Acquisti and Grossklags, who found that a user is less likely to protect their personal information in presence of benefits with missing information about data use [3]. The explanation offered is that the missing information leads the user to not think about the risk [3]. In our study, the vague terms are signals that information is missing, which may explain why users reduce their willingness to share.

Goals are formulated at different levels of abstraction and refined using sub-goals, which provides a natural mechanism for structuring complex specifications at different levels of concern [36]. A theory of vagueness that accounts for variants of summarization, i.e., likelihood, authority, and certitude, can be used to augment goal refinement patterns by introducing formalized notions of vague terms. For example, the coarse-grained privacy goal "*May share personal information*" can be refined into finer-grained sub-goals using OR-refinement to surface the specific situations that a user's personal information will and will not be shared. Regarding certitude, "mostly" implies larger coverage of cases where a goal will be achieved, whereas "typically" could emphasize common cases at the exclusion of boundary cases, and thus yield a lower frequency of achievement. The vague terms "likely" and "possibly" can indicate planned features for a future system version.

We believe our work offers benefits to practitioners. For example, policy writers can use vague terms to effectively summarize their diverse set of data practices; however, for sensitive information types, policy writers should avoid using vague terms that reduce a user's willingness to share personal information due to perceived risk. Regulators may look for vague terms surrounding specific information types as signals of increased privacy vulnerability, in which, they can invite companies to reduce vagueness around such types. For new data practices, vague terms may signal less clarity on the part of companies about how they plan to use specific information, which may be a cause for increased oversight until users have a better understanding of how their data will be used.

We also see benefits for future research. Our manually annotated dataset can be used to train ML models or to derive NLP patterns to automate the identification of vague terms in privacy policies. The content analysis, paired comparison and multi-level modeling approach we described in this paper can be applied to other types of requirements documents, to address new research challenges in the field of requirements ambiguity. These techniques may be combined in semi-automated requirements analysis tools to help requirements authors identify, categorize and measure the vagueness in requirements documents. We have developed such a prototype, which measures the vagueness across 15 privacy policies, computes a vagueness score for each type of data practice in a policy and for the policy as a whole. It then compares the vagueness scores of the policies with the vagueness scores of a set of benchmark model privacy policies in the finance industry [46].

ACKNOWLEDGMENT

We thank Howard Seltman, Stephen Broomell and the CMU RE Lab for their helpful feedback. We thank Stephanie Tallering for help with the annotations. This research was funded by NSF Award #1330596, and NSA Award #141333.

REFERENCES

[1] A. Acquisti, J. Grossklags, "An online survey experiment on ambiguity and privacy", *Communications & Strategies*, 88(4): 19-39, 2012.

[2] A. Agresti, *Categorical Data Analysis*, Wiley, 2013.

[3] A. Acquisti and J. Grossklags."Uncertainty, Ambiguity, and Privacy", *W'shp the Economics of Information Security*, 2005.

[4] A. Acquisti, L.K. John, G. Lowestein. "What is the price of privacy," *Journal of Legal Studies*, 42(2): Article 1, 2013.

[5] A.I. Antón, J.B. Earp, "A requirements taxonomy for reducing web site privacy vulnerabilities," *Req'ts Engr. J.*, 9(3):169-185, 2004.

[6] K. Auspurg and T. Hinz, *Factorial Survey Experiments*, v. 175. SAGE Publications, 2014.

[7] R.A. Bauer, "Consumer behavior as risk-taking, dynamic marketing for changing world." *American Marketing Association*, Chicago, 389.

[8] J. Bhatia, T.D. Breaux, F. Schaub. "Privacy goal mining through hybridized task re-composition", *ACM Trans. Soft. Engr. Method.*, 25(3): Article 22, 2016.

[9] D.M. Berry, E. Kamsties, M.M. Krieger. "From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity", *Univ. of Waterloo, Tech. Rep*., Nov. 2003..

[10] D. Bates, M. Maechler, B. Bolker, S. Walker. "Fitting linear mixed-effects models using lme4," *J. Stat. Soft.*, 67(1): 1-48, 2015.

[11] T.D. Breaux, F. Schaub, "Scaling requirements extraction to the crowd: experiments on privacy policies," *22nd IEEE Int'l Req'ts Engr. Conf.*, pp. 163-172, 2014.

[12] S. Boyd, D. Zowghi, and A. Farroukh, "Measuring the expressiveness of a constrained natural language: an empirical study," *13th IEEE Int'l Req'ts Engr. Conf.*, pp. 339-352, 2005.

[13] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences.* L. Erlbaum Associates, 1988.

[14] R. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 3rd ed*. SAGE Publications, 2008.

[15] R. Farkas, V. Vincze, G. Móra, J. Csirik, G. Szarvas, "The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text," *14th Conf. Comp. NL Learning-Shared Task*, pp. 1-12, 2010.

[16] L. A. Clark and D. Watson, "Constructing validity: Basic issues in objective scale development", *Psychological Assessment*, 7(3): 309-319, 1995.

[17] H. A. David, *The Method of Paired Comparisons, 2nd ed.* Oxford University Press, 1988.

[18] C. Denger, *High Quality Requirements Specifications for Embedded Systems through Authoring Rules and Language Patterns.* M.Sc. Thesis, Fachbereich Informatik, Universität Kaiserslautern, Germany 2002.

[19] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behav. Res. Methods,* 39(2): 175-191, 2007.

[20] F. Fabbrini, M. Fusani, S. Gnesi, and G. Lami, "The linguistic approach to the natural language requirements, quality: benefits of the use of an automatic tool," *26th IEEE Comp. Soc.-NASA GSFC Soft. Engr. W'shp,* pp. 97-105, 2001.

[21] N. E. Fuchs, R. Schwitter, "Specifying logic programs in controlled natural language," *Workshop on Comp. Logic for NLP*, pp. 3-5, 1995.

[22] B. Fischhoff, P. Slovic, S. Lichtenstein, S. Read, B. Combs, "How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits." *Policy Sci.* 9: 127-152, 1978.

[23] J. L. Fleiss, "Measuring nominal scale agreement among many raters." *Psych. Bulletin*, 76(5): 378-382, 1971.

[24] A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models*. Cambridge Univ. Press, 2006.

[25] D.C. Gause, G.M. Weinberg, *Exploring Requirements: Quality Before Design*, Dorset House, 1989.

[26] H. Hibshi, T.D. Breaux, S.B. Broomell, "Assessment of risk perception in security requirements composition." *23rd IEEE Int'l Req'ts Engr. Conf.*, pp. 146-155, 2015.

[27] J. Horrigan, "Online shopping," PEW Internet and American Life Project, Feb. 13, 2008.

[28] D. R. Hunter, "MM algorithms for generalized Bradley–Terry models". *The Annals of Statistics,* 32(1): 384–406, 2004.

[29] E. Kamsties, "Understanding ambiguity in requirements engineering", *Engr. & Managing Soft. Req'ts,* pp.245-266, 2006.

[30] E. Kamsties, D. Berry, B. Paech, "Detecting ambiguities in requirements documents using inspections," *1st Workshop on Inspection in Soft. Engr.* (WISE'01), pp. 68-80, 2001.

[31] B. Knijnenburg, A. Kobsa, "Increasing sharing tendency without reducing satisfaction: finding the best privacy-settings user interface for social networks", *35th Int'l Conf. Info. Sys*., pp. 1-21, 2014.

[32] J.T. Kulas, A.A. Stachowski. "Respondent rationale for neither agreeing nor disagreeing: person and item contributors to middle category endorsement intent on Likert personality indicators," *J. Research in Personality*, v. 47, pp. 254-262, 2013.

[33] N. Kiyavitskaya, N. Zeni, L. Mich, D. M. Berry, "Requirements for tools for ambiguity identification and measurement in natural language requirements specifications," *Req'ts Engr. J.,* 13(3): 207–240, 2008.

[34] F.H. Knight. *Risk, Uncertainty, and Profit.* Houghton Mifflin Company, 1921.

[35] G. Lakoff, "Linguistics and natural logic", *The Semantics of Natural Language*, pp. 545– 665, 1972.

[36] A. van Lamsweerde, *Requirements Engineering - From System Goals to UML Models to Software Specifications*, Wiley 2009.

[37] A. Massey, R.L. Rutledge, A.I. Antón, P.P. Swire. "Identifying and classifying ambiguity for regulatory requirements," *22nd IEEE Int'l Req'ts Engr. Conf*, pp. 83-92, 2014.

[38] A. M. McDonald and L. F. Cranor, "The cost of reading privacy policies", I*/S – A Journal of Law and Policy for the Information Society,* 4(3): 540-565, 2008;

[39] M. C. de Marne, B. MacCartney, C. D. Manning. "Generating typed dependency parses from phrase structure parses." *Intl. Conf. Lang. Res. & Eval.,* pp. 449-454, 2006.

[40] H. Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Stanford Law Books, 2009.

[41] E.S. Pearson, H. O. Hartley (eds). *Biometrika Tables for Statisticians,"* v. I, 2. Aufl. Cambridge University Press, 1962.

[42] E.S. Pearson, H. O. Hartley (eds). *Biometrika Tables for Statisticians,* v. I, 3. Auflage. Cambridge University Press, 1966.

[43] A. Perrin, M. Duggan. "Americans' Internet Access: 2000-2015," PEW Internet and American Life Project, June 26, 2015.

[44] D. Popescu, S. Rugaber, N. Medvidovic, D. M. Berry. "Reducing ambiguities in requirements specifications via automatically created object-oriented models." *Lecture Notes Comp. Sci.,* 5320: 103-124, 2008.

[45] R Core Team, "R: A Language and Environment for Statistical Computing." *R Foundation for Statistical Computing*, 2013.

[46] J. R. Reidenberg, J. Bhatia, T. D. Breaux, T. B. Norton, "Automated comparisons of ambiguity in privacy policies and the impact of regulation," *Journal of Legal Studies*, 2016

[47] J. Saldaña. *The Coding Manual for Qualitative Researchers*, SAGE Publications, 2012.

[48] C. Starr, "Social benefit versus technological risk," *Science*, 165, pp. 1232-1238, 1969.

[49] P. Slovic (ed.), *The Perception of Risk*, Earthscan Pub., 2000.

[50] S.F. Tjong, *Avoiding Ambiguities in Requirements Specifications*, PhD Thesis, Univ. of Nottingham, 2008.

[51] S.F. Tjong, D.M. Berry, "The design of SREE - a prototype potential ambiguity finder for requirements specifications and lessons learned", *REFSQ*, pp. 80-95, 2013.

[52] H. Turner, D. Firth, "Bradley-Terry models in R: the BradleyTerry2 package," *J. Stat. Soft.,* 48(9): 1-21. 2012.

[53] W. M. Wilson, L. H. Rosenberg, L. E. Hyatt, "Automated analysis of requirement specifications," *19th ACM/IEEE Int'l Conf. Soft. Engr.*, pp. 161-171, 1997.

[54] C.J. Wakslak, Y. Trope, "The effect of construal-level on subjective probability estimates." *Psych. Science*, 20: 52-58, 2009.

[55] H. Yang, A. Willis, A. de Roeck, B. Nuseibeh. "Automatic detection of nocuous coordination ambiguities in natural language requirements." *25th IEEE/ACM Int'l Conf. Auto. Soft. Engr.,* pp. 53-62, 2010.

[56] H. Yang, A. de Roeck, V. Gervasi, A. Willis, and B. Nuseibeh. "Analysing anaphoric ambiguity in natural language requirements." *Req'ts Engr. J*., 16: 163-189, 2011.

[57] H. Yang, A. De Roeck, V. Gervasi, A. Willis and B. Nuseibeh, "Speculative requirements: Automatic detection of uncertainty in natural language requirements," *20th IEEE Int'l Req'ts Engr. Conf.*, pp. 11-20, 2012.

[58] R. K. Yin, *Case Study Research: Design and Methods*, *5th ed*. Sage Pubs., 2013.