A comparison of privacy practices across industries

By Mathew Soares

A Master's Paper Submitted to the Faculty of the

DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS
ELLER COLLEGE OF MANAGEMENT

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2018

Table of Contents

List of Figures

List of Tables

Abstract
A majority of U.S. consumers are concerned about their online privacy. The U.S. Consumer Privacy Index 2016 found that 92% of U.S. consumers have concerns over online privacy (Trust. N.C.S. Alliance, 2016). There is current research that have analyzed the privacy policies of different companies such as Bhatia et. al and Cranor et. al (Bhatia & Breaux, 2017), (Cranor et al., 2013). However, to the best of my knowledge, no studies have analyzed the privacy policies of companies from different industries and compared the analyses of privacy policies by two coders. Therefore, this research focuses on analyzing and comparing the privacy practices of ten companies and five social media services from three different industries: Financials, Retail, and Social Media. The results from this analysis show that there are similarities and differences in the privacy practices both within and across the three industries as well as between the analyses conducted by the two coders.

Introduction
Online privacy is a big concern for U.S. consumers. According to the U.S. Consumer Privacy Index 2016, 92% of U.S. consumers have concerns over online privacy. These concerns have an effect on their behavior. For example, it was found that 89% of U.S. consumers will not purchase from companies that do not ensure their online privacy. Additionally, 28% of consumers did not complete an online transaction due to their concerns about privacy (Trust. N.C.S. Alliance, 2016).

A second study from Deloitte found that 52% of U.S. consumers did not complete a consumer survey due to their concerns over privacy. Figure 1 shows a list of actions that consumers have taken due to concerns about their privacy. An example of such action is 64% of consumers stating that they either deleted or not download a particular app (Pingitore et al., 2017).

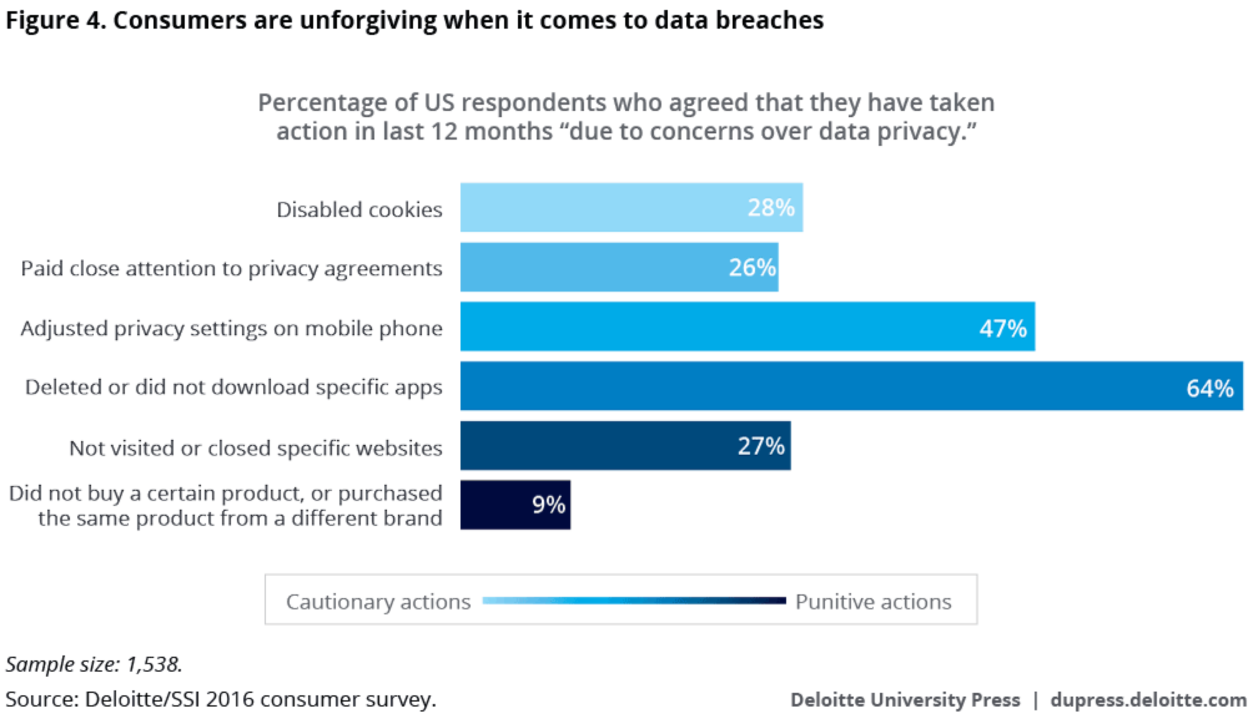


Figure 1 – List of actions taken by consumers due to concerns about privacy

In addition to concerns over online privacy, a significant number of U.S. consumers had trouble understanding a privacy policy that they read. A 2015 Pew Research Center study found that 38% of U.S. consumers had trouble comprehending the information provided to them in a privacy policy (Rainie, 2015). According to the U.S. Consumer Privacy Index 2016, only 31% of U.S. consumers understood how companies shared their information (Trust. N.C.S. Alliance, 2016). Therefore, the objective of this research is to help consumers understand the privacy practices of companies across different industries and therefore make better informed decisions.

Literature Review

This literature review will examine prior research in the following areas:

- Analysis of the privacy practices in privacy policies and disclosures
- Privacy laws
- Intercoder reliability
- Readability analysis of privacy policies

Analysis of the privacy practices in Privacy Policies and Disclosures

There is recent research that have analyzed the privacy policies of different companies. Bhatia and Breaux analyzed the privacy practices of five companies from the retail sector: Amazon, Barnes & Noble, Costco, Lowe's, and Walmart. This analysis was done by identifying the following information from these companies' privacy policies: the purposes given for the use and/or collection of consumers' information, the types of consumers' information that are used and/or collected, and the keywords that denote where a purpose is in a policy (Bhatia & Breaux, 2017).

Prior to conducting the analysis, Bhatia and Breaux preprocessed each of the policies in three steps. The first step was to remove the titles, section headers and other text that was irrelevant to the analysis such as the table of contents section. Next, the text from each the policies were divided into separate paragraphs with a length of about 120 words. Lastly, the paragraphs were put into a file to be analyzed using a task in Amazon Mechanical Turk. The analysis using Amazon Mechanical Turk was done with a tool that allowed annotators to select a piece of text and identify it with one of three labels: purpose, information, and keyword. Figure 2 shows the tool used in this analysis (Bhatia & Breaux, 2017).

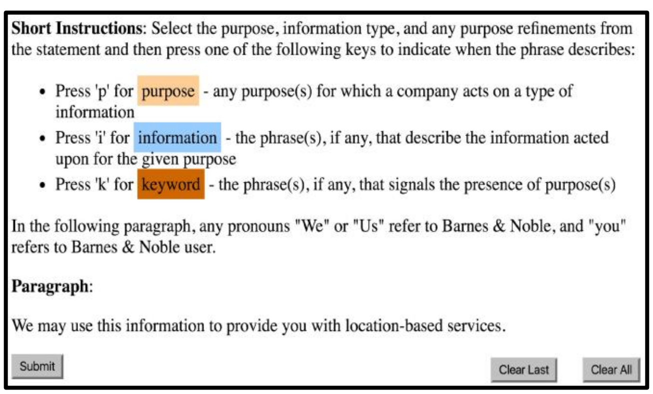


Figure 2 – Annotation tool

The analysis was able to identify 218 purposes across the five policies. Amazon's policy had the largest number of purposes. These purposes were categorized into one of six categories:

- Service Purpose
- Legal Purpose
- Communication Purpose
- Protection Purpose
- Merger Purpose
- Vague Purpose

The Vague Purpose category was used to identify any purposes that were ambiguous or unclear. Service Purpose was the largest purpose category across the policies with about 68% of the 218 purposes falling into this category. The smallest purpose category was Merger Purpose with just 1.8% of the data purposes (Bhatia & Breaux, 2017).

Cranor et al. analyzed the privacy disclosures of 3,422 companies in the financial industry. The privacy disclosures analyzed in this research were based on a standardized model privacy form that was created to help financial companies comply with Gramm-Leach-Bliley's privacy disclosure requirements. Figure 3 shows the four sections where most of the information analyzed came from. The first three sections of the form are used by a company to explain the types of personal information collected and shared, the methods used to collect personal information, and the purposes for collecting personal information (Cranor, Idouchi, Leon, Sleeper, & Ur, 2013). The fourth section is used to explain the third-party companies with whom a company shares personal information with and the reasons for this sharing. In this case, the company must complete each of the three subsections even if they all don't apply to the company. For example, if a company does not share personal information with nonaffiliated companies they can state the following, "[name of financial institution] does not share with nonaffiliates so they can market to you." (Department of the Treasury Office of the Comptroller of the Currency, Federal Reserve System, Federal Deposit Insurance Corporation, Department of the Treasury Office of Thrift Supervision, National Credit Union Administration, Federal Trade Commission, et.al, 2009).

What?	The types of personal information we collect and share depend on the product or service you have with us. This information can include: ■ Social Security number and [income] ■ [account balances] and [payment history] ■ [credit history] and [credit scores] When you are *no longer* our customer, we continue to share your information as described in this notice.

How does [name of financial institution] collect my personal information?	We collect your personal information, for example, when you ■ [open an account] or [deposit money] ■ [pay your bills] or [apply for a loan] ■ [use your credit or debit card] [We also collect your personal information from other companies.] **OR** [We also collect your personal information from others, such as credit bureaus, affiliates, or other companies.]

Reasons we can share your personal information	Does [name of financial institution] share?	Can you limit this sharing?
For our everyday business purposes— such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus		
For our marketing purposes— to offer our products and services to you		
For joint marketing with other financial companies		
For our affiliates' everyday business purposes— information about your transactions and experiences		
For our affiliates' everyday business purposes— information about your creditworthiness		
For our affiliates to market to you		
For nonaffiliates to market to you		

Definitions	
Affiliates	Companies related by common ownership or control. They can be financial and nonfinancial companies. ■ *[affiliate information]*
Nonaffiliates	Companies not related by common ownership or control. They can be financial and nonfinancial companies. ■ *[nonaffiliate information]*
Joint marketing	A formal agreement between nonaffiliated financial companies that together market financial products or services to you. ■ *[joint marketing information]*

Figure 3 – The four sections of the model privacy form

The analysis of these forms was conducted in the following steps. First, an automatic search on Google was conducted for the privacy disclosures of the 6,701 companies listed in the FDIC's directory with a domain name. The automatic search of these companies' privacy disclosures was conducted using a search string with Google's as_sitesearch URL parameter. Cranor et. al decided to only focus on the first page of the search results for each company which resulted in no more than ten links per company. This resulted in the collection of 52,564 files for these companies. The next step was to identify the files that matched the format of the model privacy form.

This was done by attempting to identify 25 keywords and phrases from the model privacy form in each file. Those files that contained less than 21 of these keywords and phrases were removed from the analysis. It was assumed that a file with less than 21 of the keywords and phrases was not based on the model privacy form and therefore was not relevant for the analysis. This step reduced the data set to 3,892 files. Next, to ensure that each company had only one file associated with it, only the file with the most keywords and phrases was kept. The last step removed any disclosures with a company name that did not match the name of the company associated with it. These steps further reduced the data set to 3,422 files. These files were then used for the analysis.

A parser was used to collect data that was mainly from the four above sections of each privacy disclosure. The first section allows companies to include up to six types of information that they collect. While this section does not provide all of the types of information collected by a company, Cranor et al. were able to find that the most common type of information collected was the consumer's Social Security number. All of the companies collected their customers' Social Security number. On the other hand, there were ten opt-out methods proved to consumers by these companies. The two most common opt-out methods were by phone or by postal mail. It is interesting to note that only 466 of the companies offered consumers an opportunity to control the sharing of their information (Cranor, Idouchi, Leon, Sleeper, & Ur, 2013).

Wilson et al. analyzed 115 privacy policies using annotations from experts in privacy and law as well as three data analysis methods. The 115 privacy policies came from a corpus of privacy policies called the OPP-115 corpus. In addition to this analysis, Wilson et al. also created a website that includes a visualization of the different types of data practices of each policy (Wilson et al., 2016).

Reidenberg et al. studied how three different groups of users interpreted the privacy policies of six news and retail companies. The three groups in this study were law and public policy graduate students, privacy experts, and common Internet users. An online tool was created to annotate the policies using nine questions that asked the participants about a company's privacy practices such as the types of personal information collected. The results from this study showed that were both agreements and disagreements in the interpretations of these policies both within the groups as well as between each group (Reidenberg et al., 2014).

Privacy Law

In the United States there are many laws both at the federal and state level that were enacted to protect the privacy of individuals. These laws include the Gramm-Leach-Bliley Act of 1999 (GLB) and the Children's Online Privacy Protection Act of 1998 (COPPA). In addition to these laws, there are also amendments in the U.S. Constitution such as the Fourth Amendment that protect the privacy of individuals (Solove & Schwartz, 2011).

Gramm-Leach-Bliley Act of 1999

The purpose of the GLB Act is to protect the privacy of consumers' financial information in the financial industry. Financial institutions and those organizations who receive consumers' financial information from a financial institution are required to abide by this law. This law requires government agencies such as the Federal Trade Commission to enact regulations that enforce its privacy requirements. An example of such a regulation is the Privacy Rule. The

Privacy Rule applies to both financial institutions and organizations that receive consumer's financial information. One of the requirements from this regulation is that an institution must give consumers the choice to control the sharing of their information (Federal Trade Commission, 2002).

Children's Online Privacy Protection Act of 1998

This law protects the online privacy of children under the age of 13 by imposing certain restrictions on websites or online services that obtain personal information from children. (Federal Trade Commission, "Children's Online Privacy Protection Rule ("COPPA"), n.d."). These requirements include gaining the parent's consent before collecting, using or disclosing a child's personal information. In this case, the types of information that are considered to be personal information include photos and Social Security numbers (Electronic Code of Federal Regulations, 2018).

Intercoder Reliability

Intercoder reliability is a measure used to determine the degree to which independent coders examine a piece of information and come to the same conclusion. This measure can be calculated using many different indices such as Cohen's kappa and Krippendorff's alpha. These indices each come with their own benefits and drawbacks. For example, Krippendorff's alpha is considered to be an index that is very adaptable and therefore can be applied to different scenarios. This index can be used to determine intercoder reliability when there are multiple coders analyzing the same piece of information as well as when there are different types of variables. However, this index does require complex calculations (Lombard, 2010).

Cohen's kappa is a commonly used index that takes into account the possibility of a chance agreement between the coders. This index is best suited for scenarios where two coders analyze a piece of information and assumes independence between the coders' analyses (Cho & Lavrakas, 2018). Cohen's kappa has a range from 0 to 1 and Table 1 shows what each range means for intercoder reliability. A kappa value of 0.81 or higher means that there was a high amount of agreement between the two coders. On the other hand, a kappa value of 0.2 or less means that there was little agreement between the coders. A kappa value of zero means that there was no agreement between the coders (Stephanie, 2014).

Kappa ranges	Amount of Agreement
0	The agreement between the two coders is due to chance.
0.1-0.20	There is only a small amount agreement between the two coders.
0.21-0.40	There is a fair amount of agreement between the two coders.
0.41-0.60	There is a moderate amount of agreement between the two coders.
0.61-0.80	There is a large amount of agreement between the two coders.
0.81-0.99	There is almost perfect agreement between the two coders.
1	There is perfect agreement between the two coders.

Table 1 – Ranges of Cohen's kappa and the Amounts of Agreement (Stephanie, 2014)

Readability of Privacy Policies

There is also research that have analyzed the readability of privacy policies. This includes research by Fabian et al., Ermakova et al., and Meiselwitz (Fabian, Ermakova, & Lentz, 2017), (Ermakova et al., 2015), (Meiselwitz, 2013). There is no commonly used metric to determine the readability of a privacy policy. Instead the research in this area tends to use multiple metrics to determine a privacy policy's readability. These methods include: Flesch Readability Ease Score (FRES), Laesbarhedsindex (LIX) and the Fry Readability Graph (Fry) (Fabian, Ermakova, &

Lentz, 2017). However, the most commonly used metric is FRES. This metric uses the length of the words and sentences in a text to determine its readability. A text with longer words and sentences has a lower FRES score. On the other hand, a text with a higher FRES score is easier to read (Meiselwitz, 2013).

Fabian et al. analyzed the readability of over 49,000 online privacy policies from different top-level domains such as .edu and .gov using readability metrics such as FRES and LIX. This study found that the average privacy policy from their collection was difficult to read and comprehend. In order to comprehend these policies, a consumer would need to be either a high school graduate or have some college experience (Fabian, Ermakova, & Lentz, 2017). Ermakova et al. found similar results when determining the readability levels of privacy policies from the healthcare and e-commerce industries. This study used similar metrics to Fabian et al. The analysis found that the policies from the healthcare industry generally had a high readability level than those from the e-commerce industry. In addition, the policies from the healthcare industry we also shorter in length than the e-commerce policies (Ermakova et al., 2015).

Similar to the previous two studies, Meiselwitz also used multiple metrics to analyze the privacy policies of twenty social media websites. This study used four metrics to determine the reading level necessary for a consumer to comprehend each privacy policy. Meiselwitz found that the majority of the policies require consumers to have a minimum reading level of a college student (Meiselwitz, 2013).

Research Gaps and Questions

While there is research that analyzes different privacy policies, there is no research that analyzes and compares the privacy practices of companies from different industries as well as compares the analyses of privacy policies by two coders using intercoder reliability. Therefore, I decided to address the following two questions with this research:

1. What are the topics in the privacy practices of companies from different industries?
2. Do privacy practices differ either within or across industries?

Research Testbed

The data used for this research were the privacy policies of ten companies and five social media services from three different industries: Financials, Retail and Social Media. Figure 4 shows the companies that were chosen for this research. I chose five companies from the Financials industry that represented some of the largest U.S.-based financial companies based on the size of their market capitalization as of March 2017 (Fortune, 2017). The five companies from the Retail industry are the same companies that were previously analyzed by Bhatia et al. (Bhatia & Breaux, 2017). Finally, the five social media services from the Social Media industry represent the largest U.S.-based social media services in terms of the number of users as of 2017 (Dunn, 2017). Facebook and Facebook Messenger share the same privacy policy (Facebook, 2016). Therefore, fourteen privacy policies were analyzed for this research. These policies were collected from the companies' websites in October 2017. The methodology I used to analyze these policies is discussed in the next section.

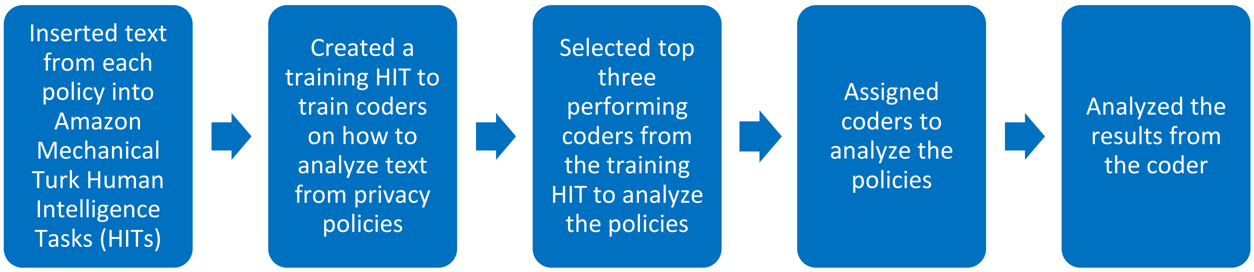
Financials	Retail	Social Media
Bank of America	Amazon	Facebook
Citigroup	Barnes & Noble	Facebook Messenger
Goldman Sachs	Costco	Instagram
J.P. Morgan	Lowes	Twitter
Wells Fargo	Walmart	WhatsApp

Figure 4 – List of companies and social media services by industry

Research Design

The policies were analyzed by two different coders. I was first coder and the second coder was a worker from Amazon Mechanical Turk (AMT). The analysis was done using thirteen types of privacy practices that can be found in Figure 18 of Appendix G. I analyzed the policies using the commenting feature of Microsoft Word. Figure 15 in Appendix C provides a screenshot that shows how this was done. The second coder analyzed the policies in a series of Human Intelligence Tasks (HITs) on AMT. This analysis was completed in the following five steps:

1. Inserted the text from each policy into a separate AMT HIT.
2. Created a training HIT to train the coders on how to analyze the policies.
3. Selected the top three performing coders from the training HIT to analyze the policies.
4. Assigned the top three coders to analyze each of the fourteen policies.
5. Analyzed the results from the coder and determined intercoder reliability (**Error! Reference source not found.**).



Error! Reference source not found.

The following five sections further explain each of these steps.

Step 1: Inserted text from each policy into a separate AMT HIT

Prior to inserting the text from each policy into a separate AMT HIT, I first removed the titles, section headers, and any irrelevant sections from each of the policies. Next, the text of each policy was divided into separate paragraphs. A question with thirteen possible answers was then added to each paragraph. The question asked the coder to read a part of a policy and select one or more of the thirteen answers that applied to the paragraph. These thirteen answers were the labels that the AMT coder used to analyze the text of each paragraph and corresponded with the thirteen types of privacy practices used in this analysis. Figure 14 in Appendix B provides a table with the definitions of these thirteen labels. The paragraphs were then inserted into an AMT HIT (Figure 5).

6. Please select all of the labels that apply to the following paragraph:

Aggregated and De-identified Data is data that we may create or compile from various sources, including but not limited to accounts and transactions. This information, which does not identify individual account holders, may be used for our business purposes, which may include offering products or services, research, marketing or analyzing market trends, and other purposes consistent with applicable laws. Third party widgets - we may allow certain widgets (e.g., social share buttons) on our Sites that enable users to easily share information on another platform, such as a social media platform. The third parties that own these widgets may have access to information about your browsing on pages of our Sites where these widgets are placed. You may wish to review information at the third party site, such as social media platforms where you have an account, to determine how these third parties collect and treat such information. Also, see Linking to other sites and Social media sites below.

☐ Non-marketing Business Operations
☐ Providing Users with Services and/or Products
☐ Advertising and/or Marketing Analytics
☐ Type of user information collected, used, and/or shared
☐ Choice given to users to control the uses of their information
☐ Method used to collect users' information
☐ Customization of Products, Services, and/or the User Experience
☐ Providing Users with General Information
☐ Protection of the User Accounts, Identities and Information
☐ Legal or Regulatory Compliance and Enforcement
☐ Company Merger or Acquisition
☐ Ambiguous or Other Purpose
☐ None of the Above

Figure 5 – Format of the questions in the HIT

Step 2: Created a training HIT to train the coders on how to analyze the policies

In order to ensure that the group of coders understood how to analyze my policies in AMT, I required them to complete a training HIT that was similar to analyzing the text of a policy. This HIT contained ten questions where the coders were asked to read paragraphs from LinkedIn's privacy policy and then analyze each paragraph using one or more of the thirteen labels. Figure 6Figure 6 shows what the training HIT looked like in AMT.

1. Please select all of the labels that apply to the following paragraph:

To create an account you provide data including your name, email address and/or mobile number, and a password. If you register for a premium Service, we ask you for payment (e.g., credit card) and billing information.

☐ Non-marketing Business Operations
☐ Providing Users with Services and/or Products
☐ Advertising and/or Marketing Analytics
☐ Type of user information collected, used, and/or shared
☐ Choice given to users to control the uses of their information
☐ Method used to collect users' information
☐ Customization of Products, Services, and/or the User Experience
☐ Providing Users with General Information
☐ Protection of the User Accounts, Identities and Information
☐ Legal or Regulatory Compliance and Enforcement
☐ Company Merger or Acquisition
☐ Ambiguous or Other Purpose
☐ None of the Above

Figure 6 – Format of the questions in the training HIT

Step 3: Selected the top three performing coders from the training HIT to analyze the policies

The top three performing coders from the training HIT were selected based on the similarity between my analysis and their analyses of the policy's text. This done to ensure the quality of the analyses of the policies.

Step 4: Assigned the top three coders to analyze each of the fourteen policies

The top three coders were then assigned to analyze the text of the fourteen policies in AMT. A coder would receive $20 for each policy they analyzed. The coders were given five hours to analyze each policy. However, since only one of the coders analyzed the text of each policy, the results section will compare my analyses to this coder's analyses.

Step 5: Analyzed the results from the coder and determined intercoder reliability

The results from the coder were analyzed by comparing my analysis of each policy to the coder's analysis in Microsoft Excel. Figure 13 in Appendix A shows an example of the second coder's analysis of a policy in AMT. On the other hand, Figure 17 in Appendix E shows an example of how the comparison of the two analyses of Citigroup's policy was done. In this case, letters A-M correspond to the thirteen labels used in this analysis and were used to calculate the degree of agreement or intercoder reliability between the two analyses of each policy. Figure 14 in Appendix B shows a table of the labels. For example, the letter A refers to the Providing Users with Services and/or Products label. In some instances, I adjusted my analysis by either adding a label or removing a label that did not describe a particular paragraph of a policy. The statistic used to determine intercoder reliability and the intercoder reliability statistics for each policy are explained in the next section.

Results

Cohen's kappa was the statistic used to determine the intercoder reliability between the two coders' analyses of each policy. An online tool called Recal2 was used to determine the Cohen's kappa of each policy. Figure 16 in Appendix D shows an example of the statistics that this tool offers to calculate intercoder reliability. In this case, the kappa statistics are referring to the degree of agreement between the two coders as to where a particular label applies in a policy. Variables 1-13 in this table refer to the thirteen labels used in this research. For example, variable 1 refers to the Providing Users with Services and/or Products label. Therefore, each of the thirteen labels has a kappa statistic that shows the degree of agreement between the two coders as to where the label applies in a particular policy.

In addition to the range of positive kappa values from Table 1, a kappa can also have a value of "undefined" which means that neither coder used the label in their analysis of a policy (ReCal, 2009). A kappa can also have a negative value which shows that there is a certain degree of disagreement between the two coders. In this case, a larger negative kappa value means that there was more disagreement between the coders about where a particular label applies in a policy (McHugh, 2012). The Cohen's kappa statistics for the thirteen labels as well as for each policy are shown in Figures 8-10 below. These statistics are organized by each of the three industries.

Financials Industry

Figure 7 shows the kappa statistics for the policies in the Financials industry. The Company Merger or Acquisition label had a value of "undefined" for all five of the policies. This means that the label was not used in either coder's analyses of these policies. The Company Merger or Acquisition label was the only label that was never used in the analyses of these policies. However, this label was not the only label that had an "undefined" value. Another label that had an "undefined" value for some of the policies was the None of the Above label. In this case, the None of the Above label was not used by either coder in their analysis of Citigroup's or Wells Fargo's policies. It is interesting to note that all five of the policies in this industry had at least one “undefined” kappa value for a label. J.P. Morgan’s policy had the largest number of labels with an “undefined” value with four out of the thirteen labels having an “undefined” value. The two labels that had negative kappa values were the Providing Users with Services and/or Products and the Protection of the User Accounts, Identities and Information labels. In particular, the Providing Users with Services and/or Products label had a kappa value of -0.33 for J.P. Morgan’s privacy policy. This was the largest negative kappa value seen for any of the policies.

Label	Bank of America	Citigroup	Goldman Sachs	J.P. Morgan	Wells Fargo
Providing Users with Services and/or Products	0.46	-0.25	0.14	-0.33	0.16
Non-marketing Business Operations	1.00	0.55	0.56	0.25	0.72
Advertising and/or Marketing Analytics	0.86	0.00	0.00	1.00	1.00
Customization of Products, Services, and/or the User Experience	0.71	1.00	undefined*	undefined*	1.00
Legal or Regulatory Compliance and Enforcement	0.67	undefined*	0.66	0.71	1.00
Protection of the User Accounts, Identities and Information	1.00	0.55	0.43	-0.20	0.70
Company Merger or Acquisition	undefined*	undefined*	undefined*	undefined*	undefined*
Providing Users with General Information	1.00	0.00	0.00	undefined*	0.00
Ambiguous or Other Purpose	0.63	0.00	0.00	undefined*	undefined*
Type of user information collected, used, and/or shared	1.00	0.29	0.67	0.71	0.36
Choice given to users to control the uses of their information	0.72	1.00	0.63	0.25	0.72
Method used to collect users' information	0.57	0.29	0.27	0.00	1.00
None of the Above	0.00	undefined*	0.00	0.00	undefined*
Overall Cohen’s Kappa	**0.81**	**0.50**	**0.51**	**0.31**	**0.78**

Figure 7 – Cohen’s kappa statistics for the Financials industry.

Undefined means that neither coder used the label in their analysis of the policy (ReCal, 2009).

Retail Industry

Figure 8 shows the kappa values for the policies in the Retail industry. It is interesting to note that none of the labels in this industry had an “undefined” value. This means that each of the labels were used to analyze all of the policies. However, there was no agreement between the two coders as to where the None of the Above label applied in any of the policies in this industry. This label was the only label where this amount of disagreement between the coders occurred. On the other hand, there was perfect agreement for the Company Merger or Acquisition label on every policy except Lowe's policy. In this case, the coders did not agree at all about where this label applied in Lowe's policy.

The only label in this industry that had a negative kappa value was the Ambiguous or Other Purpose label. This label had a kappa value of -0.05 for Barnes and Noble's policy.

Label	Amazon	Barnes & Noble	Costco	Lowes	Walmart
Providing Users with Services and/or Products	0.82	0.55	0.60	0.28	0.70
Non-marketing Business Operations	0.84	0.70	0.25	0.82	0.60
Advertising and/or Marketing Analytics	0.72	0.81	0.31	0.73	0.81
Customization of Products, Services, and/or the User Experience	0.76	0.84	0.64	1.00	0.74
Legal or Regulatory Compliance and Enforcement	1.00	0.23	1.00	0.33	0.83
Protection of the User Accounts, Identities and Information	1.00	0.64	0.74	0.62	0.67
Company Merger or Acquisition	1.00	1.00	1.00	0.00	1.00
Providing Users with General Information	0.42	0.66	0.86	1.00	0.69
Ambiguous or Other Purpose	0.00	-0.05	0.24	0.33	0.23
Type of user information collected, used, and/or shared	0.86	0.73	0.62	0.90	0.86
Choice given to users to control the uses of their information	0.74	0.85	0.60	1.00	0.81
Method used to collect users' information	1.00	0.53	0.64	0.82	0.74
None of the Above	0.00	0.00	0.00	0.00	0.00
Overall Cohen's Kappa	**0.81**	**0.70**	**0.61**	**0.74**	**0.72**

Figure 8 – Cohen's kappa statistics for the Retail industry.

Undefined means that neither coder used the label in their analysis of the policy (ReCal, 2009).

Social Media Industry

Figure 9 shows the kappa statistics for the Social Media industry. There was perfect agreement on where the Company Merger or Acquisition label applied in all five of the policies in this industry. This was similar to the results seen in the Retail industry. There was no agreement between the coders about where the None of the Above label applied in Instagram's, Facebook's, and Twitter's policies. The amount of disagreement for this label is similar to what occurred in the Retail industry.

Similar to the Retail industry, the None of the Above label was the only label in this industry that had a kappa value of "undefined". In this case, the label was not used by either coder to analyze

WhatsApp's policy. The only label that had a negative kappa value in this industry was the Ambiguous or Other Purpose label. This label had a small negative value for both Instagram's and Twitter's privacy policies.

Label	Facebook	Instagram	Twitter	WhatsApp
Providing Users with Services and/or Products	0.25	0.41	0.47	0.59
Non-marketing Business Operations	0.33	0.59	0.50	0.73
Advertising and/or Marketing Analytics	1.00	0.66	0.73	0.66
Customization of Products, Services, and/or the User Experience	1.00	0.44	1.00	1.00
Legal or Regulatory Compliance and Enforcement	1.00	0.63	1.00	0.63
Protection of the User Accounts, Identities and Information	1.00	1.00	1.00	0.67
Company Merger or Acquisition	1.00	1.00	1.00	1.00
Providing Users with General Information	1.00	1.00	0.65	0.42
Ambiguous or Other Purpose	0.00	-0.10	-0.06	0.00
Type of user information collected, used, and/or shared	0.74	0.50	0.59	0.60
Choice given to users to control the uses of their information	0.39	0.60	0.75	0.86
Method used to collect users' information	0.44	0.47	0.64	0.47
None of the Above	0.00	0.00	0.00	undefined*
Overall Cohen's Kappa	**0.62**	**0.60**	**0.73**	**0.73**

Figure 9 – Cohen's kappa statistics for the Social Media industry.

Undefined means that neither coder used the label in their analysis of the policy (ReCal, 2009).

The average Cohen's kappa values for the three industries as a whole can be found in Tables 3 3-5 in Appendix F. These kappa statistics show that there was a good amount of agreement between the two coders for each industry. However, the Retail industry had the largest amount of agreement.

Discussion

The following sections are the conclusions I made based on the analyses of the privacy policies.

The number of paragraphs from each policy categorized into each label

Figure 10Figure 10 shows the average number of paragraphs from each policy that were categorized into each of the thirteen labels by the two coders. In the Financials industry, Bank of America's policy had the largest number of paragraphs categorized into ten of the labels. This is most likely due to the fact that this policy is the longest policy in this industry. The word length of each policy can be found in

Policy	Word Count	Number of Paragraphs	FRES Score	Cohen's Kappa
Bank of America	3006	15	31.6	0.81
Citigroup	681	5	40.9	0.5
Goldman Sachs	1923	12	31.1	0.51
J.P. Morgan	1072	8	29	0.31
Wells Fargo	1256	7	42.3	0.78
Average	**1587.600**	**9.400**	**34.980**	**0.582**
Variance	**830519.300**	**16.300**	**37.717**	**0.044**

Policy	Word Count	Number of Paragraphs	FRES Score	Cohen's Kappa
Amazon	2224	15	41.3	0.81
Barnes & Noble	6010	39	33.1	0.7
Costco	2947	21	36.9	0.61
Lowe's	2965	23	31.2	0.74
Walmart	2992	22	38.2	0.72
Average	**3427.600**	**24.000**	**36.140**	**0.716**
Variance	**2188041.300**	**80.000**	**16.273**	**0.005**

Policy	Word Count	Number of Paragraphs	FRES Score	Cohen's Kappa
Facebook	2308	15	43.5	0.62
Instagram	2334	15	37	0.6
Twitter	3283	24	39.7	0.73
WhatsApp	2248	15	42.8	0.73
Average	**2543.250**	**17.250**	**40.750**	**0.670**
Variance	**244510.250**	**20.250**	**8.977**	**0.005**

Tables 3-5 in Appendix F. Barnes and Noble's policy had the largest number of paragraphs categorized into a single label with 18.5 paragraphs categorized into the Method used to collect user's information label. In the Social Media industry, Twitter's policy had the largest number of paragraphs categorized into eight of the labels.

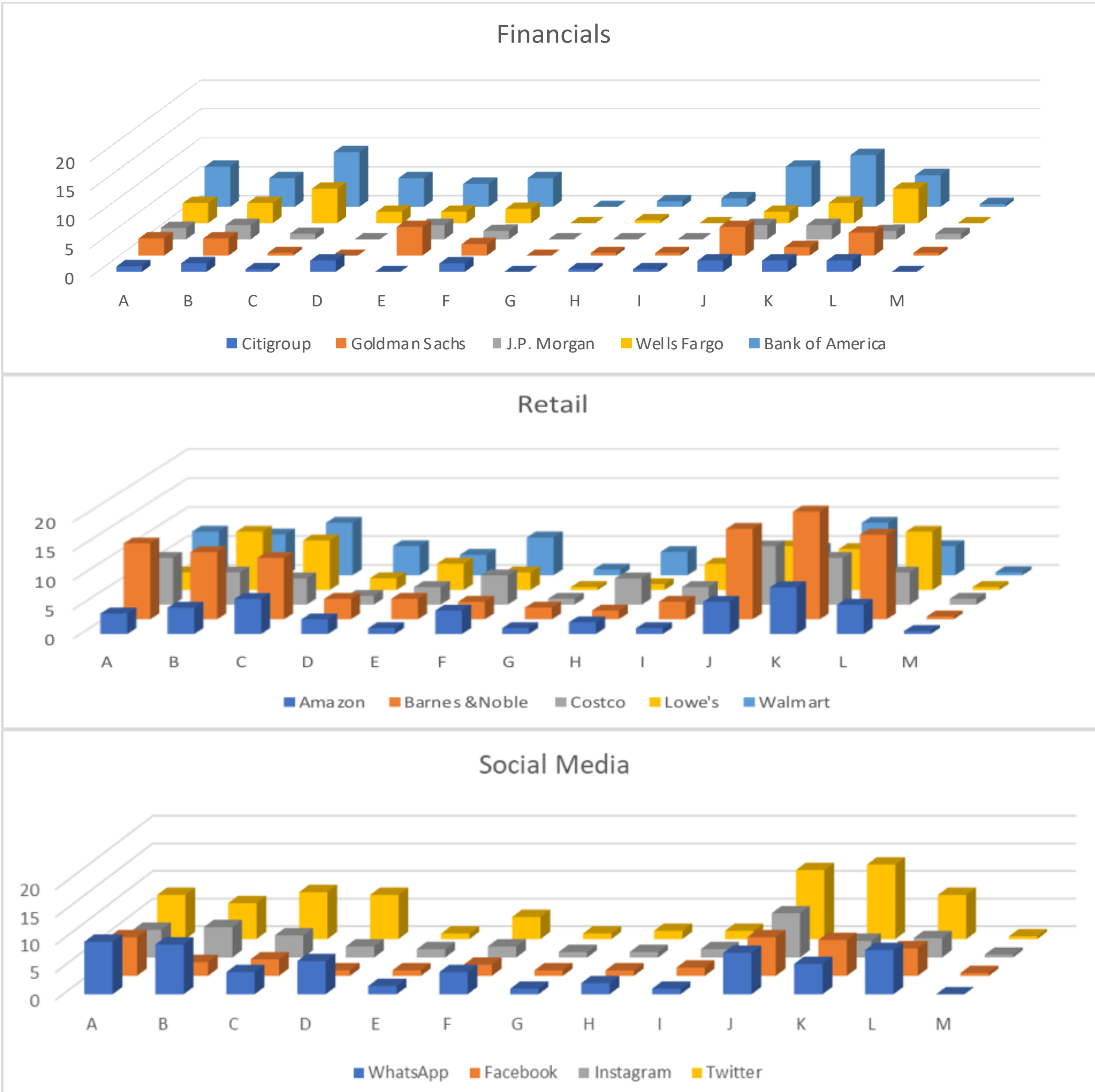


Figure 10 – Number of paragraphs from each policy categorized into each label

The number of paragraphs from each industry categorized into each label

The number of paragraphs from each of the three industries that were categorized into each label is shown in Figure 11. This figure shows that the policies in the Financials industry had the smallest number of paragraphs categorized into most of the labels.

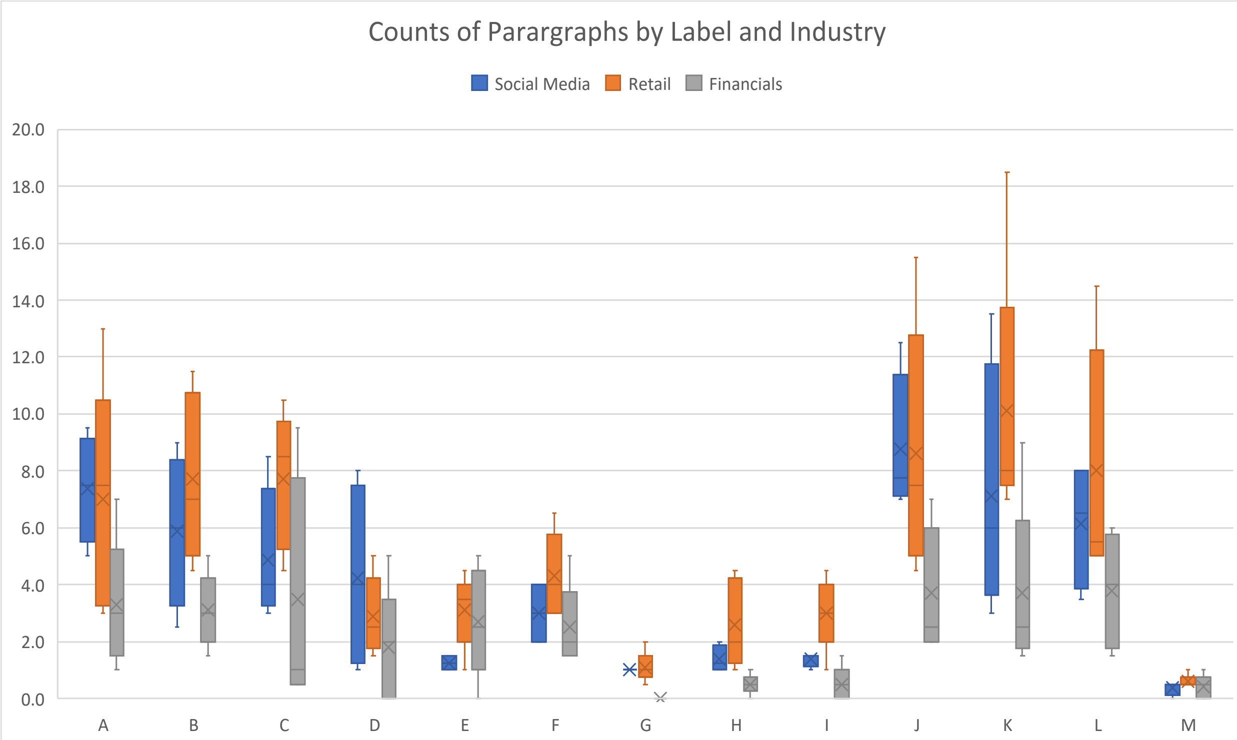


Figure 11 – Number of paragraphs from each industry categorized into each label

Privacy Practices in the Financials Industry

It is interesting to note that none of the companies in this industry stated that users' information may be shared with a third party in the event of a merger or acquisition (Bank of America Corporation, 2014), (Citigroup, Inc., 2014), (Goldman Sachs, 2013), (JPMorgan Chase & Co., 2017), (Wells Fargo, 2017). This industry was the only industry in this analysis where this was the case. There were some differences in the privacy practices observed in the policies from this industry. One example of such a difference can be found in Goldman Sachs' policy. Goldman Sachs is the only company in this industry that mentions compliance with laws and regulations regarding money laundering as a reason for collecting user's personal information (Goldman Sachs, 2013).

Privacy Practices in the Retail Industry

There were also some differences in the privacy practices found in the policies of this industry. One such difference is that Barnes & Noble's policy is the only policy not to mention preventing fraud as a reason for collecting or using user’s information (Barnes & Noble Booksellers, Inc., 2017). This is despite the fact that Barnes & Noble’s policy is more than twice the length of any other policy in this industry. As previously mentioned, each of the thirteen labels were used to analyze each policy. The use of cookies to collect user’s information was a privacy practice seen in all five of the policies in this industry (Amazon.com, Inc., 2017), (Barnes & Noble Booksellers, Inc., 2017), (Costco Wholesale Corporation, 2013), (Lowe’s, 2015), (Walmart Inc., 2017).

Privacy Practices in the Social Media Industry

Similar to the other two industries, there were also differences in the privacy practices found in the policies of this industry. An example of a such difference is the mention of using encryption to protect user’s information found in Facebook’s and WhatsApp’s policies (Facebook, 2016), (WhatsApp Inc., 2016). Similar to the Retail industry, a privacy practice seen in all four of the policies was the use of cookies to collect user’s data (Facebook, 2016), (Instagram, Inc., 2013), (Twitter, Inc., 2017), (WhatsApp Inc., 2016). In terms of the number of words, Twitter’s policy was the longest policy in this industry. However, there was a relatively small difference between the length of Twitter’s policy and the length of other three policies in this industry.

Number of legal purposes across the three industries

The number of legal purposes seen in each of the fourteen policies can be seen in Figure 11. It is interesting to note that the Financials industry had an average number of 2.7 legal purposes seen across the five policies. This average number of purposes is only slightly lower than the average of 3.1 purposes seen in the Retail industry and higher than the average of 1.3 purposes seen in the Social Media industry. Also, Goldman Sachs’s policy had the largest average number of legal purposes of any policy with five purposes.

Disagreement about where the None of the Above label applied in the policies

As shown in Figures 8-10, the two coders disagreed about where the None of the Above label applied for most of the policies. The only exceptions to this were Citigroup's, Well Fargo's, and WhatsApp's policies where neither coder used the label in their analysis. This label had the largest amount of disagreement across all of the policies.

Do Not Track signals

Some of the policies explain how the company's website or online services respond to a Do Not Track signal from a user's web browser. For example, Goldman Sachs' privacy policy states: "Our Web sites are not currently configured to respond to "do not track" signals or similar mechanisms." (Goldman Sachs, 2013).

Statements regarding compliance with laws or regulations

There were statements in some of the policies which explained how the companies complied with certain laws or regulations such as California Civil Code Section 1798.83. This regulation allows Californians to request that companies give them a list of the third parties or corporate affiliates that have received their personal information during the previous year. Barnes & Noble's privacy policy includes a section that explains this regulation to Californian users as well as how to obtain such information (Barnes & Noble Booksellers, Inc., 2017). Another example of a statement about legal compliance can be found in some of the policies where the company explains how its information security measures meet certain legal standards. For example, J.P. Morgan's privacy policy states the following about the security measures used to protect users' information, "We maintain physical, electronic and procedural safeguards that comply with applicable legal standards to secure such information from unauthorized access and use, alteration and destruction." (JPMorgan Chase & Co., 2017).

Ambiguous types of users' information

There were some of types of users' information in the policies that were vague and therefore were classified as ambiguous. These information types were not considered as a type of user information in my analysis. For example, one of the information types in Facebook's privacy policy is "information the developer or publisher of the app or website provides to you or us". This information type is ambiguous because it is not clear what specific user information the developer or publisher is providing to Facebook (Facebook, 2016).

Notice to users visiting or using an unaffiliated website or online service

In addition to informing users about how Do Not Track signals are handled, some of the policies inform users about the privacy practices of unaffiliated websites or online services. These statements inform users that the company itself is not responsible for the privacy practices of any unaffiliated website or online service. For example, Bank of America's privacy policy warns users that the privacy practices of third-party websites may be different than those of Bank of America's (Bank of America Corporation, 2014).

Year of the Most Recent Update

The year when the policy was last updated is included in each of the policies. Table 2 shows the year of the most recent update for each policy. It is interesting to note that nearly half of the policies were most recently updated in 2017. Goldman Sachs, Costco and Instagram were the only companies that have not updated their policies since 2013 (Goldman Sachs, 2013), (Costco Wholesale Corporation, 2013), (Instagram, Inc., 2013).

Policy	Industry	Year of the Most Recent Update
Bank of America	Financials	2014
Citigroup	Financials	2014
Goldman Sachs	Financials	2013
J.P. Morgan	Financials	2017
Wells Fargo	Financials	2017
Amazon	Retail	2017
Barnes & Noble	Retail	2017
Costco	Retail	2013
Lowe's	Retail	2015
Walmart	Retail	2017
Instagram	Social Media	2013
Facebook	Social Media	2016
Twitter	Social Media	2017
WhatsApp	Social Media	2016

Table 2 – Year of the most recent update for each policy (Bank of America Corporation, 2014), (Citigroup, Inc., 2014), (Goldman Sachs, 2013), (JPMorgan Chase & Co., 2017), (Wells Fargo, 2017), (Amazon.com, Inc., 2017), (Barnes & Noble Booksellers, Inc., 2017), (Costco Wholesale Corporation, 2013), (Lowe's, 2015), (Walmart Inc. 2017), (Facebook, 2016), (Instagram, Inc., 2013), (Twitter, Inc., 2017), (WhatsApp Inc., 2016)

The number of words and paragraphs, FRES scores and Cohen's kappa values for each industry

Tables 3-5 in Appendix F show the average and variance of the number of words and paragraphs, FRES scores and Cohen's kappa values for each industry. The policies from the Retail industry had the largest average number of words and paragraphs as well as Cohen's kappa value. This means that this industry had on average the longest policies and the largest amount of agreement between the two coders. On the other hand, the policies of the Financials industry had both the lowest average FRES score and Cohen's kappa value. This means that these policies on average are the most difficult policies to comprehend and had the lowest amount of agreement between the two coders (Ermakova et al., 2015). An interesting point regarding the Financials industry is the amount of variance seen in the FRES scores and Cohen's kappa values for the policies. The amount of variance or degree of variability seen in the FRES scores and Cohen's kappa values from this industry was the largest of the three industries. This shows that there was a significant difference in the amount of agreement and level of readability between the policies of this industry (Stephanie, 2017).

The FRES scores for each policy indicate that none of these policies are easy to read. Facebook's FRES score was the highest of any policy at 43.5. On the other hand, J.P. Morgan's FRES score was the lowest of any policy at 29. This score means that J.P. Morgan's policy is considered to be "very difficult" in terms of its readability. The other thirteen polices are considered to be "difficult" in terms of their readability (Ermakova et al., 2015). This result is similar to what can be seen when comparing the Cohen's kappa values of each policy. J.P. Morgan's Cohen's kappa value of 0.31 was the lowest of any policy.

Conclusion

A few interesting points were found in this analysis. One such point is the differences in the privacy practices seen both with and across the three industries. For example, the policies from the Financials industry had the smallest number of paragraphs categorized into most of the thirteen labels used in this analysis. This result is not surprising since the policies of this industry also had the smallest average number of words and paragraphs. Another interesting point is that the amount of agreement between the two coders and the level of readability for each policy differed both within and across the industries. The level of readability for each of the policies could be seen as a contributing factor to the results seen in the Pew Research study of U.S. consumers mentioned earlier. Overall, this analysis found that were differences both in the privacy practices seen within and between industries and in the analyses of the policies conducted by the two coders.

Future Directions

The future directions for research could focus on: adding more companies and industries to the analysis, automating the analysis and conducting a longitudinal analysis to compare this analysis with the analyses conducted in past research. As an example, companies from the healthcare and technology industries might be included in an automatic analysis. Figure 12 shows the ten companies that could be analyzed.

Healthcare	Technology
Amgen, Inc.	Alphabet
Gilead Sciences, Inc.	Apple
Johnson & Johnson	Intel
Pfizer, Inc.	IBM
UnitedHealth Group Inc.	Microsoft

Figure 12 – Companies to be analyzed in the future

The automatic analysis of these companies' policies might include the use of text analytics software. This analysis could include information such as a rating for each policy that is based on their readability level or the number of choices given to the users.

References

Amazon.com, Inc. (2017). *Amazon Privacy Notice.* Retrieved from https://www.amazon.com/gp/help/customer/display.html/ref=footer_privacy?ie=UTF8&nodeId=468496

Bank of America Corporation. (2014). *Bank of America U.S. Online Privacy Notice.* Retrieved from https://www.bankofamerica.com/privacy/online-privacy-notice.go

Barnes & Noble Booksellers, Inc. (2017). *Privacy Policy.* Retrieved from https://www.barnesandnoble.com/h/help/privacy-policy-complete

Bhatia, J., & Breaux, T. D. (2017). A Data Purpose Case Study of Privacy Policies. *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference, RE 2017*, 394–399. https://doi.org/10.1109/RE.2017.56

Cho, Y. I., & Lavrakas, P. J. (2018). Intercoder Reliability. *Encyclopedia of Survey Research Methods*. https://doi.org/10.4135/9781412963947.n228

Citigroup, Inc. (2014). *Online Privacy Statement.* Retrieved from https://www.citigroup.com/citi/privacy.html

Costco Wholesale Corporation. (2013). *Costco Wholesale Corporation Your Privacy Rights.* Retrieved from https://www.costco.com/privacy-policy.html

Cranor, L. F., Idouchi, K., Leon, P. G., Sleeper, M., & Ur, B. (2013). Are They Actually Any Different? Comparing Thousands of Financial Institutions' Privacy Practices. *The Twelfth Workshop on the Economics of Information Se- Curity (WEIS 2013) , June 11–12, 2013, Washington, DC*.

Dunn, J. (2017). *Facebook totally dominates the list of most popular social media apps.* Retrieved from http://www.businessinsider.com/facebook-dominates-most-popular-social-media-apps-chart-2017-7

Electronic Code of Federal Regulations. (2018). *Title 16: Commercial Practices, Part 312- Children's Online Privacy Protection Rule*. Retrieved from https://www.ecfr.gov/cgi-bin/text-idx?SID=4939e77c77a1a1a08c1cbf905fc4b409&node=16%3A1.0.1.3.36&rgn=div5

Ermakova, T., Fabian, B., & Babina, E. (2015). Readability of Privacy Policies of Healthcare Websites. *Wi*, (April), 1085–1099.

Fabian, B., Ermakova, T., & Lentz, T. (2017). Large-scale readability analysis of privacy policies. *Proceedings of the International Conference on Web Intelligence - WI '17*, (September), 18–25. https://doi.org/10.1145/3106426.3106427

Facebook. (2016). *Data Policy*. Retrieved from https://www.facebook.com/privacy/explanation

Federal Trade Commission. n.d. *Children's Online Privacy Protection Rule ("COPPA").* Retrieved from https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule

Federal Trade Commission. (2002). *How to Comply with the Privacy of Consumer Financial Information Rule of the Gramm-Leach-Bliley Act*. Retrieved from https://www.ftc.gov/tips-advice/business-center/guidance/how-comply-privacy-consumer-financial-information-rule-gramm

Fortune. (2017). *Fortune 500 Companies 2017: Who Made the List (filtered based on market value as of March 31, 2017 for the Financials sector).* Retrieved from http://fortune.com/fortune500/list/filtered?sector=Financials&sortBy=mktval

Goldman Sachs. (2013). *Goldman Sachs Global Privacy Policy.* Retrieved from http://www.goldmansachs.com/privacy-and-cookies/global-privacy-policy.html

Instagram, Inc. (2013). *Privacy Policy.* Retrieved from https://help.instagram.com/155833707900388

JPMorgan Chase & Co. (2017). *Privacy Policy.* Retrieved from https://www.jpmorgan.com/country/US/en/privacy

Lombard, M., Snyder-Duch, J., Bracken, C. (2010). *Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects.* Retrieved from http://matthewlombard.com/reliability/

Lowe's. (2015). *Lowe's US Privacy Statement.* Retrieved from https://www.lowes.com/l/privacy-and-security-statement.html

McHugh, Mary. (2012). *Interrater reliability: the kappa statistic.* Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

Meiselwitz, G. (2013). Readability Assessment of Policies and Procedures of Social Networking Sites. In A. A. Ozok & P. Zaphiris (Eds.), *Online Communities and Social Computing* (pp. 67–75). Berlin, Heidelberg: Springer Berlin Heidelberg.

OCC, Federal Reserve System, FDIC, OTS, NCUA, FTC, … SEC. (2009). Final Rule: Final Model Privacy Form under the Gramm-Leach-Bliley Act, (34815), 74:62890-62994. Retrieved from https://www.sec.gov/rules/final/2009/34-61003.pdf

Pingitore, Gina, Rao, Vikram, Cavallaro, Kristen, Dwivedi, Kruttika. (2017). *To share or not to share - What consumers really think about sharing their personal information.* Retrieved from https://www2.deloitte.com/insights/us/en/industry/retail-distribution/sharing-personal-information-consumer-privacy-concerns.html

Rainie, Lee. (2015). *Americans conflicted about sharing personal information with companies.* Retrieved from http://www.pewresearch.org/fact-tank/2015/12/30/americans-conflicted-about-sharing-personal-information-with-companies/

ReCal. (2009). *ReCal error log entry #1: invariant values (updated 5/12/09).* Retrieved from http://dfreelon.org/2008/10/24/recal-error-log-entry-1-invariant-values/

Reidenberg, J. R., Breaux, T., Carnor, L. F., French, B., Cranor, L. F., Grannis, A., … Schaub, F. (2014). Disagreeable Privacy Policies: Mismatches Between Meaning and Users' Understanding. *Berkeley Technology Law Journal*, *30*(1), 39–88. https://doi.org/10.15779/Z384K33

Stephanie. (2014). *Cohen's Kappa Statistic.* Retrieved from http://www.statisticshowto.com/cohens-kappa-statistic/.

Stephanie. (2017). *Variance: Simple Definition, Step by Step Examples.* Retrieved from http://www.statisticshowto.com/probability-and-statistics/variance/.

Trust. N.C.S. Alliance. (2016). *TRUSTe/NCSA Consumer Privacy Infographic – US Edition.* Retrieved from https://www.trustarc.com/resources/privacy-research/ncsa-consumer-privacy-index-us/

Twitter, Inc. (2017). *Twitter Privacy Policy.* Retrieved from https://twitter.com/en/privacy

Walmart Inc. (2017). *Walmart Privacy Policy.* Retrieved from https://corporate.walmart.com/privacy-security/walmart-privacy-policy

Wells Fargo. (2017). *Wells Fargo Digital Privacy and Cookies Policy.* Retrieved from https://www.wellsfargo.com/privacy-security/privacy/online

WhatsApp Inc. (2016). *WhatsApp Privacy Policy*. Retrieved from https://www.whatsapp.com/legal/#privacy-policy

Wilson, S., Schaub, F., Dara, A. A., Liu, F., Cherivirala, S., Giovanni Leon, P., … Sadeh, N. (2016). The Creation and Analysis of a Website Privacy Policy Corpus. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1330–1340. https://doi.org/10.18653/v1/P16-1126

Appendix A: Second coder's Answers in Amazon Mechanical Turk

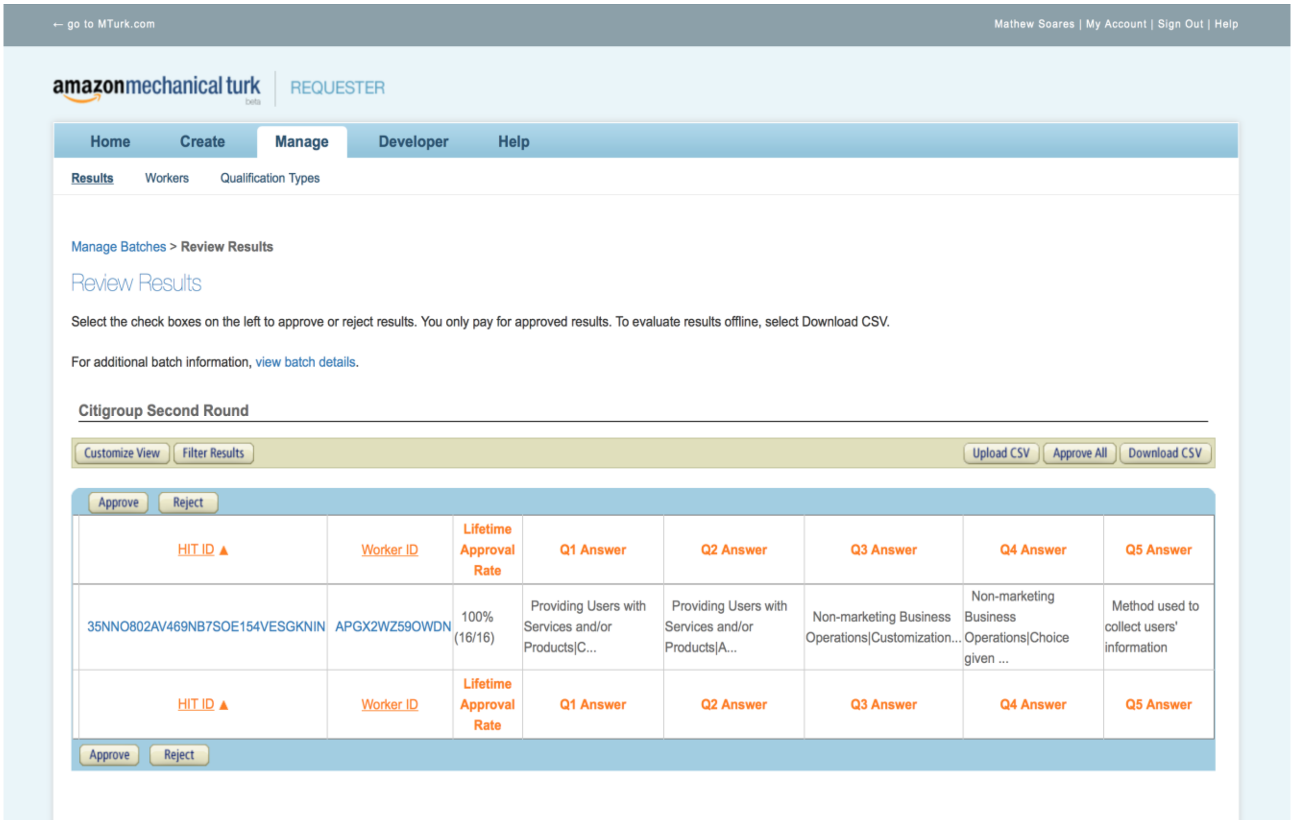


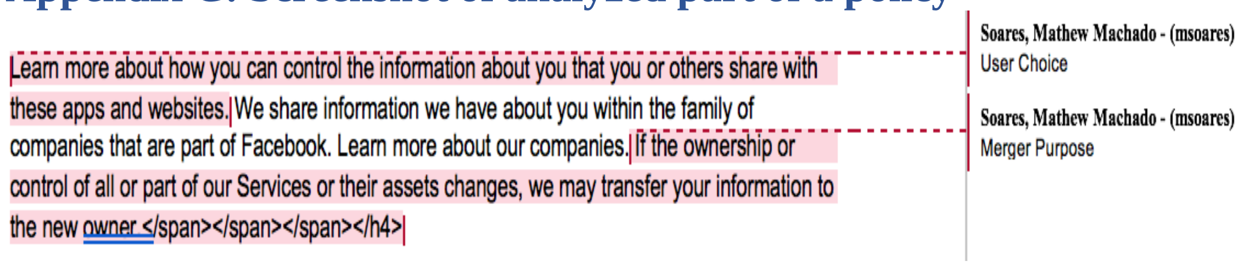
Figure 13 – Second coders answers to Citigroup's privacy policy in Amazon Mechanical Turk

Appendix B: Definitions of the thirteen labels

Label	Definition
Non-marketing Business Operations	A purpose where the company is supporting its business operations, improving its products and services, or assisting another business with its operations that doesn't already fall into a more specific purpose such as authentication or communication. (e.g., Your information helps us understand how users use our site as well as improve our products and services.)
Customization of Products, Services, and/or the User Experience	A purpose where the company customizes their products, services, or the user experience for the user. (e.g., The information is used to help make your user experience on this website more personalized.)
Providing Users with General Information	A purpose where the company provides non-marketing information to the user. (e.g., Your information is used to inform you about important information such as updates to this website)
Protection of the User Accounts, Identities and Information	A purpose where the company verifies a user’s or their device’s identity or uses other security measures to ensure the protection of users' accounts, identities and information from unauthorized uses or disclosure. (e.g., Your information is used for risk control, fraud detection and prevention, and to verify your identity.)
Providing Users with Services and/or Products	A purpose where the company provides services and/or products to the user. (e.g., Your information is used to provide our services and products to you.)
Legal or Regulatory Compliance and Enforcement	A purpose where the company is complying with any legal or regulatory requirements and/or its terms, conditions and policies. (e.g., Your information may be disclosed in order to comply with applicable laws and regulations.)
Company Merger or Acquisition	A purpose where the company is involved in a merger, acquisition, or transfer of its assets. (e.g., In the event of a merger with another company, your information will be disclosed to the successor entity.)
Advertising and/or Marketing Analytics	A purpose where the company advertises its products or services (this includes advertisements that are tailored to the user) and/or conducts marketing analytics. (e.g., Your information is used to provide advertising and help us determine the effectiveness of our advertising.)
Ambiguous or Other Purpose	A purpose that is too vague to classify into one of the above eight data purpose labels. (Please use this label for purposes that do not fit into any of the above data purpose labels.) (e.g., Your information will not be used for any other purposes.)
Types of user information collected, used, and/or shared	The types of users’ information collected, used, and/or shared (e.g., name, email address, telephone number).
Choices given to users to control the uses of their information	The choices given to users to control how their information is collected, used and/or shared (e.g., You can also opt out specifically from interest-based advertising served through our platform for third parties.).
Methods used to collect users' information	The methods used to collect users’ information from users and/or through technologies (e.g., We collect data through cookies and similar technologies.).
None of the Above	Workers were told to use this label if none of the other labels applied to the text.

Figure 14 – Definitions for the labels

Appendix C: Screenshot of analyzed part of a policy



Learn more about how you can control the information about you that you or others share with these apps and websites. We share information we have about you within the family of companies that are part of Facebook. Learn more about our companies. If the ownership or control of all or part of our Services or their assets changes, we may transfer your information to the new owner </h4>

Soares, Mathew Machado - (msoares)
User Choice

Soares, Mathew Machado - (msoares)
Merger Purpose

Figure 15 – Example of how the policies were analyzed using Microsoft's commenting feature

Appendix D: Screenshot of intercoder reliability statistics for Citigroup’s policy

ReCal 0.1 Alpha for 2 Coders
results for file "Citigroup Before Recal.csv"

File size:	266 bytes
N columns:	26
N variables:	13
N coders per variable:	2

	Percent Agreement	Scott's Pi	Cohen's Kappa	Krippendorff's Alpha (nominal)	N Agreements	N Disagreements	N Cases	N Decisions
Variable 1 (cols 1 & 2)	60%	-0.25	-0.25	-0.125	3	2	5	10
Variable 2 (cols 3 & 4)	80%	0.524	0.545	0.571	4	1	5	10
Variable 3 (cols 5 & 6)	80%	-0.111	0	0	4	1	5	10
Variable 4 (cols 7 & 8)	100%	1	1	1	5	0	5	10
Variable 5 (cols 9 & 10)	100%	undefined*	undefined*	undefined*	5	0	5	10
Variable 6 (cols 11 & 12)	80%	0.524	0.545	0.571	4	1	5	10
Variable 7 (cols 13 & 14)	100%	undefined*	undefined*	undefined*	5	0	5	10
Variable 8 (cols 15 & 16)	80%	-0.111	0	0	4	1	5	10
Variable 9 (cols 17 & 18)	80%	-0.111	0	0	4	1	5	10
Variable 10 (cols 19 & 20)	60%	0.167	0.286	0.25	3	2	5	10
Variable 11 (cols 21 & 22)	100%	1	1	1	5	0	5	10
Variable 12 (cols 23 & 24)	60%	0.167	0.286	0.25	3	2	5	10
Variable 13 (cols 25 & 26)	100%	undefined*	undefined*	undefined*	5	0	5	10

*Scott's pi, Cohen's kappa, and Krippendorff's Alpha are undefined for this variable due to invariant values.

Figure 16 –Intercoder reliability statistics for Citigroup’s policy

Appendix E: Comparison of analyses by the two coders using Microsoft Excel

Question	1	2	3	4	5
Person 1	A F K	D J	D	B K L	I
Mathew	K F	L J D C A	B D F H	B J K L	L J

A	Providing Users with Services and/or Products
B	Non-marketing Business Operations
C	Advertising and/or Marketing Analytics
D	Customization of Products, Services, and/or the User Experience
E	Legal or Regulatory Compliance and Enforcement
F	Protection of the User Accounts, Identities and Information
G	Company Merger or Acquisition
H	Providing Users with General Information
I	Ambiguous or Other Purpose
J	Type of user information collected, used, and/or shared
K	Choice given to users to control the uses of their information
L	Method used to collect users' information
M	None of the Above

Figure 17 –Comparing the second coder's analysis of Citigroup's policy with my analysis

Appendix F: Word Length, Cohen's Kappa and FRES Score Statistics for each policy

Policy	Word Count	Number of Paragraphs	FRES Score	Cohen's Kappa
Bank of America	3006	15	31.6	0.81
Citigroup	681	5	40.9	0.5
Goldman Sachs	1923	12	31.1	0.51
J.P. Morgan	1072	8	29	0.31
Wells Fargo	1256	7	42.3	0.78
Average	**1587.600**	**9.400**	**34.980**	**0.582**
Variance	**830519.300**	**16.300**	**37.717**	**0.044**

Policy	Word Count	Number of Paragraphs	FRES Score	Cohen's Kappa
Amazon	2224	15	41.3	0.81
Barnes & Noble	6010	39	33.1	0.7
Costco	2947	21	36.9	0.61
Lowe's	2965	23	31.2	0.74
Walmart	2992	22	38.2	0.72
Average	**3427.600**	**24.000**	**36.140**	**0.716**
Variance	**2188041.300**	**80.000**	**16.273**	**0.005**

Policy	Word Count	Number of Paragraphs	FRES Score	Cohen's Kappa
Facebook	2308	15	43.5	0.62
Instagram	2334	15	37	0.6
Twitter	3283	24	39.7	0.73
WhatsApp	2248	15	42.8	0.73
Average	**2543.250**	**17.250**	**40.750**	**0.670**
Variance	**244510.250**	**20.250**	**8.977**	**0.005**

Tables 3-5 –Word length, Cohen's Kappa and FRES Score for each policy grouped by industry: Financials, Retail and Social Media

Appendix G: Privacy Practice Types and Categories

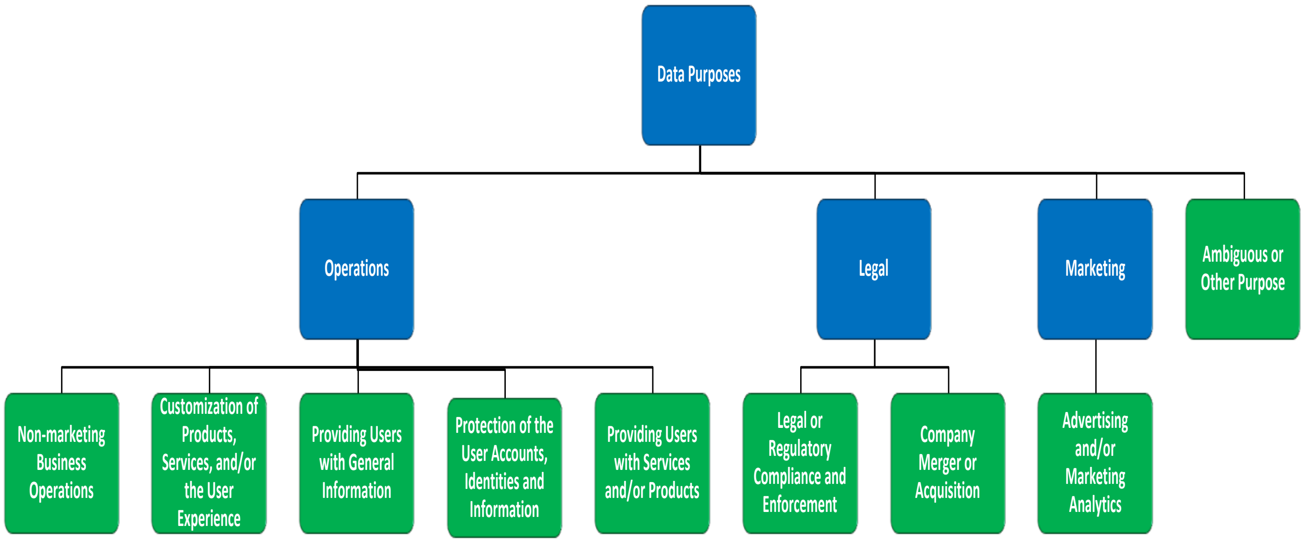


Figure 18 – Privacy Practice types and categories