

**Carnegie Mellon University
Information Networking Institute**

TECHNICAL REPORT

**SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF**

Master of Science in Information Technology – Information Security (MSIT-IS)

Title: Crowdsourcing Privacy Policy Analysis: Evaluating the Comfort, Readability and Importance of Privacy Policies

Presented by: Chaiwut Chaianuchittrakul

Accepted by the Information Networking Institute

Project Advisor(s): **Dr. Jason Hong**
 Print Name (s)

Signature(s)

Date

Project Reader: **Dr. Kilho Shin**
 Print Name

Signature

Date

Academic Advisor: **Dr. Nicolas Christin**
 Print Name

Signature

Date

INI Director: **Dena Haritos Tsamitis**
 Print Name

Signature

Date

Project Presentation Date: **21 November 2013**

I give the Information Networking Institute permission to publish my technical report as an on-line document and as a library resource.

Author(s)' signature(s): _____

TR # _____
Office use only

Crowdsourcing Privacy Policy Analysis: Evaluating the Comfort, Readability and Importance of Privacy Policies

Submitted in partial fulfillment of the requirements for
the degree of
Master of Science
in
Master of Science in Information Technology - Information Security

Chaiwut Chaianuchittrakul

B.A., Asia Pacific Management, Ritsumeikan Asia Pacific University

Carnegie Mellon University
Pittsburgh, PA

December, 2013

Copyright © 2013 by Chaiwut Chaianuchittrakul
All rights reserved except the rights granted by the
Creative Commons Attribution-Noncommercial Licence

Acknowledgements

I would like to express my deep gratitude to Dr. Jason Hong for his patient guidance, enthusiastic encouragement and constructive suggestions during the planning and development of this research work. I would like to thank Dr. Kilho Shin, my reader, for his useful critiques. My grateful thanks are also extended to members of CHIMPS lab, who provided extensive professional feedback and comments.

I would also like to extend my thanks to all staff of Information Networking Institute, all staff of Global Communication Center and Dr. Nicolas Christin, my academic supervisor, for their help and support.

Finally, I wish to thank my parents for their support and encouragement throughout my study.

Abstract

In this decade, Internet users already make up one-third of the world's population. Numerous web services collect, use and share users' personal information, such as location data and search history. Personal information is used to track and identify unique users in order to customize services for each user. Currently, privacy policies are the only tools that inform users about the flow of data and help users make privacy practice decisions. However, past research shows that privacy policies tend to be difficult and time-consuming to read. Building on the previous research regarding approaches that facilitate user comprehension of privacy policies, including Platform for Privacy Preferences (P3P) and Privacy Nutrition Label, this project seeks to develop a privacy policy interface that is more concise and user-friendly. We explore the potential of using crowdsourcing techniques to help improve usability and simplify complexity of privacy policies. To evaluate small segments of privacy policies, our research uses Amazon Mechanical Turk, a crowdsourcing online marketplace. Through asking users about the comfort, difficulty and importance of the individual segments of privacy policies, we can identify surprising, difficult and important segments in each privacy policy. In total, the experiment resulted in eight trials with five different privacy policies. Unlike previous findings, which show that the privacy policies are difficult and confusing, our results suggest that a majority of users think almost all segments are understandable and important but do not raise any privacy concerns. This result shows that users think the context of a privacy policy is important. The result also suggests surprising facts that users can understand privacy policies and the privacy policies we evaluated did not raise any concerns.

Table of Contents

Acknowledgements	ii
Abstract	iii
List of Tables	vi
List of Figures	vii
1 Introduction	1
1.1 Research Objective	2
1.2 Overview of Results	3
2 Background & Related Work	7
2.1 Privacy Policy	7
2.1.1 Background	7
2.1.2 Related Works	9
2.2 Crowdsourcing	12
2.2.1 Background	12
2.2.2 Related Works	12
2.3 Summary	13
3 Methodology	15
3.1 Data Collection	15
3.1.1 Tools	15
3.1.2 Privacy Policies in Our Studies	16
3.1.3 Core Process	17
3.2 Experiments	19

3.2.1	The First Experiment - A Feasibility Test with Klout's Privacy Policy	19
3.2.2	The Second Experiment - An Improved Interface with Duolingo's Privacy Policy	22
3.2.3	The Third Experiment - A More Sensitive Service with M&T Bank's Privacy Policy	25
3.2.4	The Fourth Experiment - A Higher Number of Assignments with M&T Bank's Privacy Policy	28
3.2.5	The Fifth Experiment - A More Difficult Privacy Policy with USCIS's Privacy Policy	29
3.2.6	The Sixth Experiment - Smaller Segments of USCIS's Privacy Policy	30
3.2.7	The Seventh Experiment - More Suprising Use of Information with Wikimedia's Privacy Policy	33
3.2.8	The Eighth Experiment - Binary Scale of Questions with Wikimedia's Privacy Policy	35
3.2.9	Result Summary & Discussion	36
4	Conclusion	39
	Bibliography	41
	Appendix A Question Interfaces	45
	Appendix B Experiments Data	49

List of Tables

Table 3.1	Negative and Positive Meanings of Each Evaluation	22
Table 3.2	Summary Table of Eight Experiments	38
Table B.1	The First Experiment Data	50
Table B.2	The Second Experiment Data	51
Table B.3	The Third Experiment Data	51
Table B.4	The Fourth Experiment Data	52
Table B.5	The Fifth Experiment Data	52
Table B.6	The Sixth Experiment Data	52
Table B.7	The Seventh Experiment Data	53
Table B.8	The Eighth Experiment Data	53

List of Figures

Figure 2.1	Flesh Reading Ease Score, Flesch Grade Level	9
Figure 2.2	Privacy seals	10
Figure 2.3	Privacy Nutrition Label	11
Figure 3.1	Core Process	18
Figure 3.2	The First Experiment Interface	20
Figure 3.3	The First Experiment Results	21
Figure 3.4	The Second Experiment Improved Interface	23
Figure 3.5	The Second Experiment Results	24
Figure 3.6	The Third Experiment Results	27
Figure 3.7	The Fourth Experiment Results	29
Figure 3.8	The Fifth Experiment Results	31
Figure 3.9	The Sixth Experiment Results	32
Figure 3.10	The Seventh Experiment Results	34
Figure 3.11	Binary Questions	35
Figure 3.12	The Eighth Experiment Results	37
Figure A.1	The First Experiment Interface	46
Figure A.2	The Second Interface	47
Figure A.3	The Third Interface	48

1

Introduction

The majority of people still do not fully understand privacy practices, even though many of them are online. However, they are concerned about their privacy and want to understand these privacy practices. According to Miniwatts Marketing Group's Internet Usage Statistics, two billion users are already online. In the United States, 78.6% of the population are Internet users [27]. The Consumer Action 2013 survey shows that a large majority of consumers are unaware of many privacy issues; for example, 29% of participants did not know that their location data could be tracked by their mobile phone, and 49% of participants thought online tracking is illegal. This means that the majority of people still do not fully understand their online privacy. However, 90% of participants are seeking tools to limit tracking, and 95% of participants believe they should have the right to control what information can be collected online. These answers can be interpreted as suggesting that people want to control and make choices about their privacy policy [5].

However, the research shows that while the privacy policy is a tool to help people, it tends to be difficult and time-consuming to read, and, therefore, many people don't read privacy policies. A Microsoft survey shows that the privacy policy is the only

tool that people will use to look for privacy-related questions [26]. However, past research shows that privacy policies are written in confusing linguistic patterns [30]. According to the difficulty estimation algorithm, some of them even require 25 years of education to fully understand all contexts [18]. McDonald et al. estimate that it would require approximately 244 hours for a person to read every unique service privacy policy for every website they visit in a given year [25]. There are several solutions, such as P3P, a machine-readable language that allows browsers to read privacy policies for users, or TRUST seals, which are the seals that help users pre-verify the privacy-friendly services. However, these tools still cannot provide effective help in the privacy notice and choice because only 10% of service providers have P3P, and more than half of people do not recognize TRUST seals [8, 28]. Kelley et al. proposed a privacy nutrition label that allows consumers to scan the brief information [19]. With the privacy nutrition label, participants can find the information in less time with more accuracy and satisfaction.

1.1 Research Objective

Our research aims to develop a new user-friendly approach to inform consumers about privacy practices without having them read every entire privacy policy on every unique service they use. In this master’s thesis, we present the results of a feasibility study using crowdsourcing to simplify privacy policies. Our research aims to summarize the privacy policy into a shorter version, or heat map, of surprising, difficult or important segments in each privacy policy to minimize reading time. We adopt the crowdsourcing concept to read privacy policies for users. Crowdsourcing is a collaboration of a large group of people to achieve a task or objective [10]. Instead of reading privacy policies on their own, humans also can be influenced by the social proof from crowdsourcing [4]. If crowd workers read privacy policies and did not find anything surprising, then users can assume services are privacy-friendly. By applying

Amazon Mechanical Turk (a crowdsourcing service), we aim to produce two main results:

- Summarize the privacy policies into shorter versions
- Visualize the privacy policy in comfort, readability and importance differences

1.2 Overview of Results

We conducted eight experiments in total through Amazon Mechanical Turk, the crowdsourcing marketplace, with different hypotheses and objectives. In each case, the privacy policies were split up into smaller segments. We then asked crowdsourced participants to read and answer four questions regarding each segment, including 1) the main idea, 2) comfort, 3) readability and 4) importance of each segment. Our experiments are as follows:

1. The first study used Klout’s privacy policy as an initial feasibility study to see how well the idea of crowdsourcing privacy policies analysis might work. Surprisingly, we found that crowd workers found each individual segment to be easy to read and not surprising. We conjectured that the results may be due to lazy workers, which led to our next study.
2. The second study used Duolingo’s privacy policy. We also improved the user interface to minimize the effect of lazy workers. The improved user interface displays other segments with navigation buttons in case workers need to read other segments for better understanding. This study also asked for main idea and three 5-point questions of comfort, readability and importance of each segment workers read. However, the results still showed that crowd workers found each privacy policy segment to be easy to read, not concerning and important. We suspected these results may be due to lazy workers or that

Duolingos privacy policy itself was too easy, important and not surprising. Therefore, we continued to test a more important privacy policy in our next study.

3. The third study used M&T Bank's privacy policy, changing social media privacy policy to one involving financial information. We expected to get significant differences in our results due to the more sensitive nature of banking services. Every segment has an average of importance higher than those in Duolingos privacy policy, but it still did not show any significant differences. We still suspected the problem of lazy workers. We conjectured that the results may be due to having too many lazy workers.
4. The fourth study used M&T Bank's privacy policy but we posted 30 assignments per segment (up from 10 assignments per segment) to reduce the potential effect of lazy workers. We expected to see significant variance. However, we did not see any significant differences. The result still showed that crowdsourced workers thought that individual segments of M&T Bank's privacy policy was easy to read, important and not surprising. We conclude that increasing the number of assignments is still not a good solution. It led us to use other privacy policies and other hypotheses in our next study.
5. The fifth study used USCIS's privacy policy, embedding many legal terms and more difficult to understand than the previous three privacy policies. We also monitored time spent on each assignment. We expected to see significant variance of difficulty between segments embedding legal terms and the rest of them. However, the average readability fell into a neutral range (0, 1) again. Even the different difficulty of segments did not produce any significant variance in the results. We thought that results may still be due to lazy workers who

did not want to read long segments, which led us to our next study.

6. The sixth study used USCIS's privacy policy and shrunk normal size segments (< 100 words) into shorter segments (< 40 words). This approach might reduce the number of lazy workers. But the result still showed that the average rating of shorter segments produced no significant differences in comfort, difficulty and importance compared with normal-sized segments at the same content. We suspected other problems or USCIS's privacy policy itself instead of lazy workers, which led to our next study.
7. The seventh study used Wikimedia's privacy policy, which collects suspiring data, such as location data, which might make users more concerned about this. We expected to see that some segments were rated as surprising personal information use (unexpected use). However, we still did not receive any significant results. We thought that 5-point question from $(-2, 2)$ range is too wide for people to make a decision, so we continued to our next study.
8. The eighth study used Wikimedia's privacy policy and used binary questions instead of normal 5-scale questions. We thought that binary questions might help users decide to answer negative value easier than 5-scale questions. Since scales of both questions are different, we normalized all our results on Wikimedia's privacy policy evaluation into $(0, 1)$ range to compare their differences. Once again, we did not observe any differences from binary scale and 5-point. We concluded that scales of questions have no significant impact on the results.

We expected to use those main ideas from each segment to summarize the short version of each privacy policy. Moreover, we intended to use the average of each scale to visualize privacy policies into a heat map.

However, we found that users' summaries of the main idea cannot be easily used to summarize the privacy policy in the same way as same as previous crowdsourcing work because all of them are not written in the same way. It is difficult to do it without manually summarizing. We also found that the majority of the three scales' averages fall in the same range. That is, participants mostly labeled segments as easy to read, important and not surprising. We attempted seven experiments, trying smaller segments, more workers and different scales of answers with five different privacy policies from Klout, Duolingo, M&T Bank, U.S. Citizenship and Immigration Services and Wikimedia [21, 13, 29, 34, 35]. Unfortunately, our results did not expose any significant findings within these eight experiments.

2

Background & Related Work

This section discusses the background of privacy policies and their complexity followed by previous research on privacy policy. Then, this section explains the model and meaning of crowdsourcing and its previous works. In the end of this section, we discuss previous works strengths and weaknesses, which influence our research.

2.1 Privacy Policy

2.1.1 Background

Our research is aiming to reduce complexity of privacy policy, so understanding the role and the background of privacy policy is essential. Today, online services involving individuals' information are rapidly spreading. These services collect, share, receive and use individuals information [7, 32]. Service providers use information to optimize personalized services, such as Amazon's recommendation system or Google's search. Indeed, personalized services are useful and convenient. At the same time, these services also raise many privacy concerns and questions about information that service providers collect, use and share. Because of these concerns, there should be tools allowing visitors to understand the collection and distribution and how they

can control of their own personal information on websites. One efficient tool is a privacy policy. According to a Microsoft-commissioned survey of 1,000 U.S. adults, 39% of people seek answers from their friends and family, and 29% seek answers from companies privacy policies [26]. This simply means that the privacy policy is a significant tool of notice and choice for individuals information-related actions [7]. The privacy policy allows service providers to communicate and clarify the data flow to their visitors.

Many service providers post privacy policies on their websites. They want to obtain users' trust by explaining privacy practices and educating users' decisions [9]. Users benefit from online privacy policies because some users verify privacy-friendly services before making any online purchases [33]. However, privacy policies are mainly left unnoticed, and only a small fraction of users want to read them [?]. In addition, even though some users want to read those privacy policies, reading them has a surprisingly expensive opportunity cost. McDonald et al. estimate that Internet users would need 244 hours to read the privacy policies of every unique web service that they visit in a given year. In the United States, this represents a national opportunity cost of \$781 billion per year [25]. Besides the time cost, Jensen et al. conducted research using the Flesch Reading Ease Score and the Flesch-Kincaid Grade Level to evaluate of the complexity of the text [14, 20]. Their results show that almost 94% of privacy policies require at least a high school education to read, and 12.5% of privacy policies require at least a postgraduate education [18]. The most significant point is that only a small fraction of visitors bother to read it [16].

Nonetheless, privacy policies still play an important role in communicating and helping visitors protect their privacy information. Therefore, "How can we make privacy policies better?" is still an important question to be answered. The next sections provide a discussion of approaches to improve the practical usage of privacy policies.

Flesch Reading Ease Score (FRES): $206.835 - 84.6 * (\text{syllables/words})$ $- 1.015 * (\text{words/ sentences})$ Flesch Grade Level (FGL): $(0.39 * \text{words/sentences})$ $+ (11.8 * \text{syllables/words}) - 15.59$
--

Figure 2.1: Formulas of Flesch Reading Ease Score (FRES) and Flesch Grade Level (FGL)

2.1.2 Related Works

In 2002, the World Wide Web Consortium (W3C) introduced the platform for privacy preferences (P3P). P3P is an XML format of privacy policies that allows browsers to read privacy policies for users. If the users' preferences and the web site's P3P specification do not match, the browser will block those websites. Therefore, users do not need to read every privacy policy on every unique website they visit [6]. However, the browsers decide whether to block each particular website based on the bad tokens. Along with low adoption, past work has found that a large number of websites have errors in their P3P policies [23, 8]. There are many approaches to avoid the compact policy blocking. For example, Amazon uses its own token as AMZN, and Facebook and Google use natural language instead of P3P format or P3P compact policy tokens [6, 11]. Cranor et al. show that P3P has a low rate of adoption (only 10% among most popular websites in 2008) [8]. Therefore, the P3P is an inefficient tool for aiding users' privacy decisions so far.

Next, another approach is providing privacy trustable seals for websites, such as TRUSTe Trustmark (www.truste.com), WebTrust Program (www.webtrust.org), and BBB Online (no longer operating). The service providers must submit the information on the privacy policy and typically pass the evaluation from these seal programs. Basically, these programs will review the privacy policies of websites for



Figure 2.2: TRUSTe Trustmark, WebTrust Program, BBB Online Privacy seals (from left to right)

users. Hence, users can verify these seals instead of spending time to reading privacy policies [2, 28]. Although the seals are capable of reducing reading time, the research shows that many of users do not recognize these seals. Moores mentions that only 60 respondents (42%) recognized TRUSTe, and 41 respondents (29%) recognized BBBOnline. Somehow, 21 respondents recognized the fake seal, which was even more than WebTrust seal recognition [28]. Moreover, there is more research that shows the coverage of privacy issues on privacy policies. Pollach conducted a privacy policies assessment by trying to answer questions on data handling. The results showed that the companies' privacy policies with at least one trust seal could answer only 34.3% of all questions. This is worse coverage than the average privacy policy (39.4% of all questions) [30]. Both points raise questions about how useful the seals are and how much users can rely on them.

Kelley et al. take another approach, designing and evaluating a privacy "nutrition label." The privacy nutrition label is based on the idea of food nutrition facts. For food, nutrition labels are a standard where people can look to find facts and help them to compare with other products (or, in this case, services). Similarly, a privacy nutrition label presents brief information quickly, using just a few symbols and colors. Then, users who want to know more details can read the whole privacy policy. The authors also did a user study by asking questions as well as measuring accuracy and

Acme

information we collect	ways we use your information				information sharing	
	provide service and maintain site	marketing	telemarketing	profiling	other companies	public forums
contact information		opt out	opt out			
cookies						
demographic information		opt out	opt out			
preferences		opt out	opt out			
purchasing information		opt out	opt out			
your activity on this site		opt out	opt out			

Information not collected or used by this site: social security number & government ID, financial, health, location.

Access to your information
This site gives you access to your contact data and some of its other data identified with you

How to resolve privacy-related disputes with this site
Please email our customer service department

acme.com
5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@acme.com

	we will collect and use your information in this way		we will not collect and use your information in this way
opt out	by default, we will collect and use your information in this way unless you tell us not to by opting out	opt in	by default, we will not collect and use your information in this way unless you allow us to by opting in

Figure 2.3: An example of privacy nutrition label displaying information collection, use and sharing.

timing compared for both natural language and a nutrition label of a privacy policy. The nutrition label helps users to find out accurate information for 13 of 14 questions. It also reduces time to find the information for 13 of 14 questions. The participants were also satisfied using nutrition labels more than they were when reading the natural language privacy policy. However, there were a couple participants who misinterpreted the meaning of the symbols [19]. There are also similar approaches to creating icons and labels, such as Raskins Privacy Icons [31]. These labels are colorful and more attractive than normal text, but they look confusing. Users still need to be educated and get familiar with privacy terms in order to use this privacy nutrition label efficiently.

2.2 Crowdsourcing

2.2.1 Background

Crowdsourcing is an approach to tackling complex problems or achieving task objectives with a large network (crowd) of participants. Howe and Robinson, who originally used the term 'crowdsourcing,' describe it as the following: "Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call" [17].

Crowdsourcing is an example of a collective distributed intelligence process for doing tasks that often require a real human. In crowdsourcing, the main task is often broken up into smaller tasks called microtasking. This kind of microtasking allows the workers work faster, cheaper and with less errors [10]. Along with its fast, cheap and accurate properties, crowdsourcing can be done through online web technologies. It allows the individuals to work in the same single environment with decentralized restrictions, such as demographics, cultural background, etc. [17].

2.2.2 Related Works

Instead of hiring professionals to do particular work, crowdsourcing provides decent work within a shorter time. This section introduces two projects run by crowdsourcing. First, Bernsterin et al. introduce Soylent, a word processing tool powered by crowdsourcing. Soylent has three main components: Shorten, a service that cuts text to 85% of the original length; Crowdproof, a service that proofreads spelling and grammar; and The Human Macro, a service that edits, formats and adds figures. This project uses Amazon Mechanical Turk to recruit workers. The authors also manage to improve the quality of works using the Find-Fix-Verify pattern. The result of the work is mostly correct, but it still needs improvement for its deployment

[3].

Lin et al. use Amazon Mechanical Turk to survey user expectations on Android permission use. There are many applications that request unexpected permissions. For example, the Angry Birds game requested network location access, the Backgrounds HD Wallpapers app requested contact list access and the Brightest Flashlight app requested device ID information. The authors looked at the sensitive permissions and asked workers to state the expectation condition and comfort condition. After they got the users' evaluation, they created a mock interface, notifying the percentage of users that surprised each sensitive permission requests at the app download page. The new interface reduced time spent on permissions summary reading, and more people also mentioned concerns compared with the original permission request page [24].

2.3 Summary

Despite its complexity, a privacy policy is still a useful tool for users to understand websites' privacy practices, and our research wants to reduce its complexity. Our research takes an idea from the nutrition label by aiming to provide a shorter or more concise version of a privacy notice. The privacy nutrition label allows users to spend a shorter amount of time obtaining information they need [19]. We take a similar approach to summarizing a long privacy policy into a short version.

In order to create a concise version of privacy policy, our research aims to harvest crowds' intelligence to achieve an objective. The Soylent project could shortened a paragraph by asking crowdsourcing workers to summarize texts into shorter paragraphs [3]. At the same time, Lin et al.'s work on Android permissions shows that crowdsourcing workers have a potential to evaluate or make any decisions on privacy issues. Our research aims to request crowdsourcing workers to pull main ideas out of each privacy policy segment. Moreover, we would like to ask crowdsourcing workers

to evaluate comfort, readability and importance of each segments, too.

3

Methodology

In this section, we introduce Amazon Mechanical Turk, the online crowdsourcing marketplace. We post the tasks, then we explain our process of the data collection. Next, we conduct eight different experiments and discuss problems and keys to take away from each experiment.

3.1 Data Collection

3.1.1 Tools

We use Amazon Mechanical Turk in order to deliver tasks to crowdsourcing workers. Amazon Mechanical Turk is an online marketplace that allows individuals or businesses (a.k.a. Requesters) to submit the tasks (which is known as HITs: Human Intelligence Tasks). Then, Workers (Turkers or Providers) browse and select their tasks and complete them. Amazon Mechanical Turk describes seven steps as follows:

[1]

1. Begin with a project: define the goals and key components of the project.
2. Break it into tasks and design your HIT: design the tasks as microtasking for

workers.

3. Publish HITs to the marketplace: post assignments into the marketplace.
4. Worker accepts assignments: workers decide on assignments to work on.
5. Worker submits assignments for review: he or she submits assignments for requesters.
6. Approve or reject assignments: requesters review the quality of tasks and approve or reject them.
7. Complete the project: requesters download the data after all assignments are done. In order to design and submit the HITs, our first experiment used the sentiment project template provided on the website to submit. However, we decided to customize the HITs from the second experiment to the eighth experiment. Amazon Mechanical Turk provided the SDK in several languages, such as Java, Ruby and .NET platforms. We decided to use Ruby SDK to deliver the task. We will discuss the further details of our HITs in the experiment sections.

3.1.2 Privacy Policies in Our Studies

We used five different privacy policies in all experiments as follows:

- Klout’s privacy policy (<http://klout.com/corp/privacy>) has 2283 words and it can be divided into 41 segments [21].
- Duolingo’s privacy policy (<http://www.duolingo.com/privacy>) has 886 words and it can be divided into 14 segments [13].
- M&T Bank’s internet privacy policy (<https://www.mtb.com/customerservice/>

`Pages/PrivacyPolicy.aspx`) has 885 words and it can be divided into 13 segments [29]

- USCIS’s privacy policy (http://www.uscis.gov/Privacy_Policies) has 2240 words and it can be divided into 41 segments [34].
- Wikimedia’s privacy policy (http://meta.wikimedia.org/wiki/Privacy_Policy/BannerTestA) has 4744 words and it can be divided into 102 segments [35].

We decided to switch from the Klout’s privacy policy to Duolingo’s privacy policy because Duolingo’s privacy policy is much shorter than Klout’s one. After we found the problems, we tried to switch the privacy policy to M&T Bank’s internet privacy policy because we had a hypothesis that users might be more concerned in cases of financial privacy involvement. We then used the USCIS’s privacy policy because we found it was difficult in term of readability due to legal terms that are embedded in it. Finally, we changed to the Wikimedia privacy policy because they collected location data, which might surprise users.

3.1.3 Core Process

This core process is the structure we used to get the result from Amazon Mechanical Turk as follows.

1. Choose and modify the privacy policy: we choose a privacy policy and verify that there is no additional privacy policy link embedded inside the privacy policy. We also modify the policy by removing the companys name (except in the first experiment) to minimize the impact that a brand name might have on results.
2. Break the privacy policy into small segments: we try to break a privacy policy into multiple segments of the same size. However, the privacy policy is always

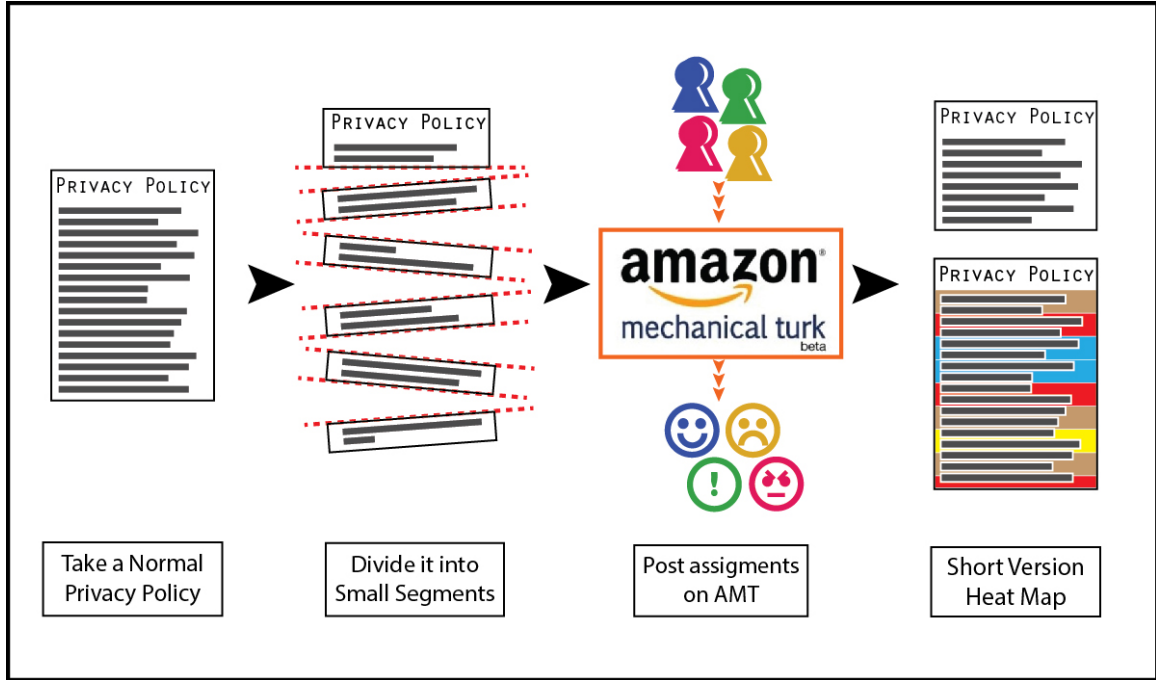


Figure 3.1: A core process of our studies on Amazon Mechanical Turk. We divided privacy policies into segments, posted them on the marketplace and then used results to visualize data

written with long explanation sentences. There are also many sections in one privacy policy. We do not combine the segments belonged to two sections, and we try not to break the sentences. Typically, our segments sizes varied from 50 to 120 words.

3. Load segments to database: excluding the first experiment, we inserted all segments into the MySQL database in order to provide the external question page.
4. Test the HITs: Amazon Mechanical Turk provides the sandbox marketplace that allows developers to test the HITs without paying the real payment. So, we tested our HITs every time before we posted it to the real marketplace.

5. Submit the assignments: we prepaid the amount of rewards and submit assignments to the marketplaces. The price we paid are varied to assignments. We will provide further information in the experiments section.
6. Review and approve the results: Amazon Mechanical Turk provides a reject option for the requesters, but we tried not to reject any of them because Amazon still takes payment even for rejected tasks. We also worried about the requesters account reputation.
7. Summarize and visualize the results: we use Microsoft Excel Pivot Table to find out the basic statistics information as average and variance of each question. Then, we plot the processed data into the graph, examine the results and determine how we can develop further experiments.

3.2 Experiments

3.2.1 The First Experiment - A Feasibility Test with Klout's Privacy Policy

The first experiment is a feasibility test whether crowdsourcing has a potential to evaluate privacy policy or not. We imported seven segments manually picked from 41 segments of Klouts privacy policy and posted 5-point evaluation tasks of comfort and readability on Amazon Mechanical Turk. Crowdsourced workers will see an interface and a 5-point question and they have to select to rate each segment as strongly negative (-2), negative (-1), neutral (0), positive (1) and strongly positive (2). We posted five assignments per segment. We have seven segments and two questions. Therefore, we posted a total of 70 assignments with a \$0.02 reward to workers.

Instructions

Strongly positive	Select this if this part of privacy policy makes you feel extremely safe and relieve to use the service after you read this part.
Positive	Select this if this part of privacy policy makes you relieve to use the service but it still has some unexpected points within the context.
Neutral	Select this if this part of privacy policy is normal. (Nothing is good/bad within this context)
Negative	Select this if this part of privacy policy makes you a little bit surprised but

Judge the sentiment expressed by the following item toward: Klout's Privacy Policy

Segment: Providing additional information beyond what is required at registration is entirely optional, but enables you to better identify yourself and find opportunities in the Klout system.

☐ Strongly negative
 ☐ Negative
 ☐ Neutral
 ☐ Positive
 ☐ Strongly positive

You must ACCEPT the HIT before you can submit the results.

Figure 3.2: A 5-scale evaluation interface we used to ask crowdsourced workers to rate the readability and comfort of individual segments of Klout’s privacy policy.

Hypothesis

We expected the results projected the significant difference of the comfort level and difficulty of each segments good enough to generate a visualization of Klouts privacy policy.

Results

We observed that almost all of results are in the neutral range (-0.5, 0.5) on both factors instead of distributed within (-2, 2) range. There are 6 segments from 7 segments in comfort level and 5 segments from 7 segments in difficulty. (see figure 3.3) Most of them are also in the positive range, which is not useful for our visualization process. Moreover, we saw that many workers rated 'neutral range' 18 from 35 workers in difficulty and 17 from 35 workers in comfort level.

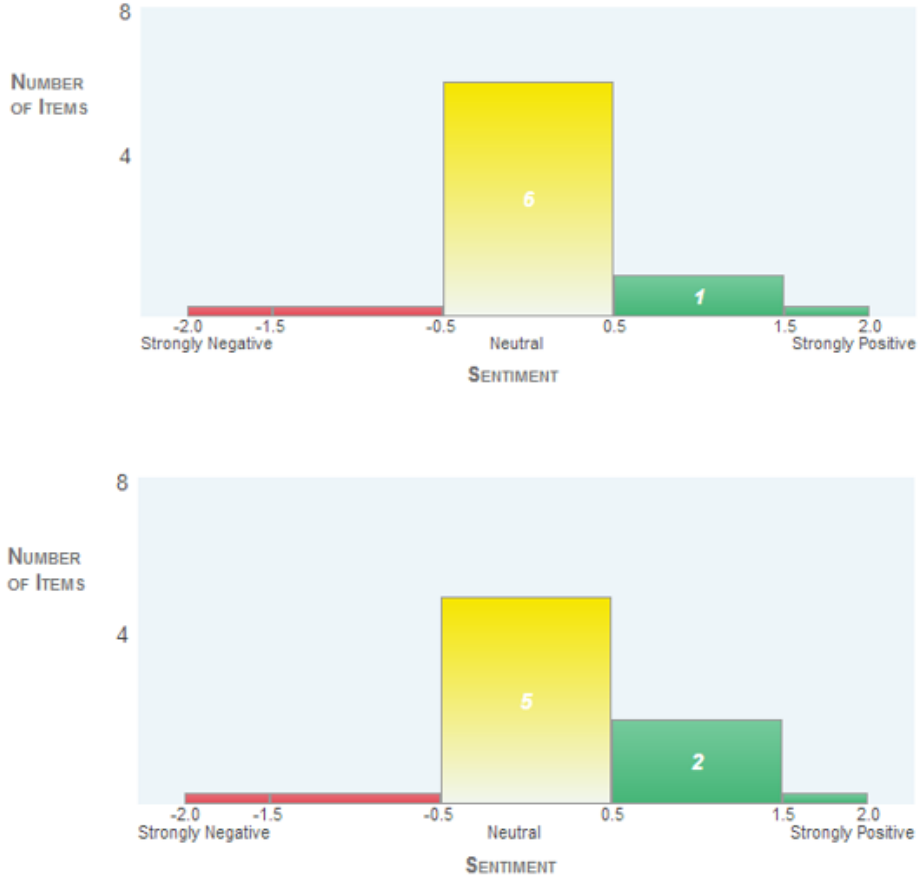


Figure 3.3: Average of readability (top) and average of comfort (bottom) of 7 individual segments of Klout’s privacy policy in $(-2, 2)$ range.

Discussion

The result does not support our hypothesis. Almost all of the segments’ evaluation fell into neutral range $(-0.5, 0.5)$. We have three conjectures on why the first experiment did not work as we expected. First, the workers might be confused because they can read only one segment at a time. Second, the workers might be lazy so they just pick the neutral one in the middle. Third, we might need more workers in order to get better results. Therefore, we conclude that we need to get rid of on these impacts on results prior generating any good visualizations. We move on to

the next experiment.

3.2.2 The Second Experiment - An Improved Interface with Duolingo’s Privacy Policy

In this experiment, we worked on a new interface embedded with a navigation button in order to allow workers to read other segments. We modified our interface apart from the previous study’s single 5-point question. Based on previous studies by Amazon Mechanical Turk, we followed the solution by adding a main idea question to prevent the lazy workers based on previous crowdsourcing work and we ask three questions at the same time [3, 12].

We imported fourteen segments of Duolingo’s privacy policy into .csv file and posted 5-point evaluation tasks of comfort and readability on Amazon Mechanical Turk. Crowdsourced workers have to summarize a main idea of segment they read and then rate 3 questions of 5-point scale from strongly negative (-2) to strongly positive (2). We posted ten assignments per segment. So, we posted total 140 assignments with increased reward as \$0.10 to crowdsourced workers. If the results are not good enough, we will also try to remove some answers according to the quality of the main idea summary and measure the difference.

Evaluations	Negative Values	Positive Values
Comfort	Suprising	Comfortable
Difficulty	Difficult to read	Easy to read
Importance	Unimportant	Important

Table 3.1: Negative and Positive Meanings of Each Evaluation

Hypothesis

We expect to be able to reduce lazy workers impact by forcing workers to summarize the main idea of segment they read. Moreover, we expect to see some significant

The text below shows the privacy policy for Duolingo.com

Please read the **highlighted segment** (segment no. 2/14) and answer the following questions.

Optional: You may use the hide/show buttons to show hidden segments.
You may also use navigation buttons to navigate and read other segments
if you need to understand the context better after you toggle the display.

By using, accessing or participating in the Service, you agree to the terms of this privacy policy (the "Privacy Policy"). Capitalized terms not defined in this Privacy Policy have the meanings set forth in the Terms and Conditions, located at http://____.com/#/terms.

We reserve the right to change our Privacy Policy at any time. If we do this, we will post a notice that we have made changes to this Privacy Policy on the Website for at least 7 days after the changes are posted and will indicate at the bottom of the Privacy Policy the date these terms were last revised.

Any revisions to this Privacy Policy will become effective the earlier of (i) the end of the foregoing 7-day period or (ii) the first time you access or use the Service after any such changes. If you do not agree to abide by this Privacy Policy, you are not authorized to use, access or participate in the Service.

Context Navigation: Up || Highlighted segment || Down

Hide Other Segments Show Other Segments

Please briefly summarize the main idea of the text in the highlighted segment in one or two sentences.

Please input your answer here...

How **important** is the information in the highlighted segment in using the web site? For example, is it something you would want a close friend to tell you before you use the web site?

- ☐ Very important
- ☐ Important
- ☒ **Neutral**
- ☐ Unimportant
- ☐ Very unimportant

How **comfortable** would you be with the policy in the highlighted segment, if you were using this web site?

- ☐ Very comfortable
- ☐ Comfortable
- ☒ **Neutral**
- ☐ Uncomfortable
- ☐ Very uncomfortable

How **understandable** is the text in the highlighted segment?

- ☐ Very easy to understand
- ☐ Easy to understand
- ☒ **Neutral**
- ☐ Hard to understand
- ☐ Very hard to understand

Submit

Figure 3.4: An improved interface we developed for the second experiment. We asked crowdsourced workers to summarize the main idea of the individual segment they read and answer three 5-point questions of comfort, difficulty and importance.

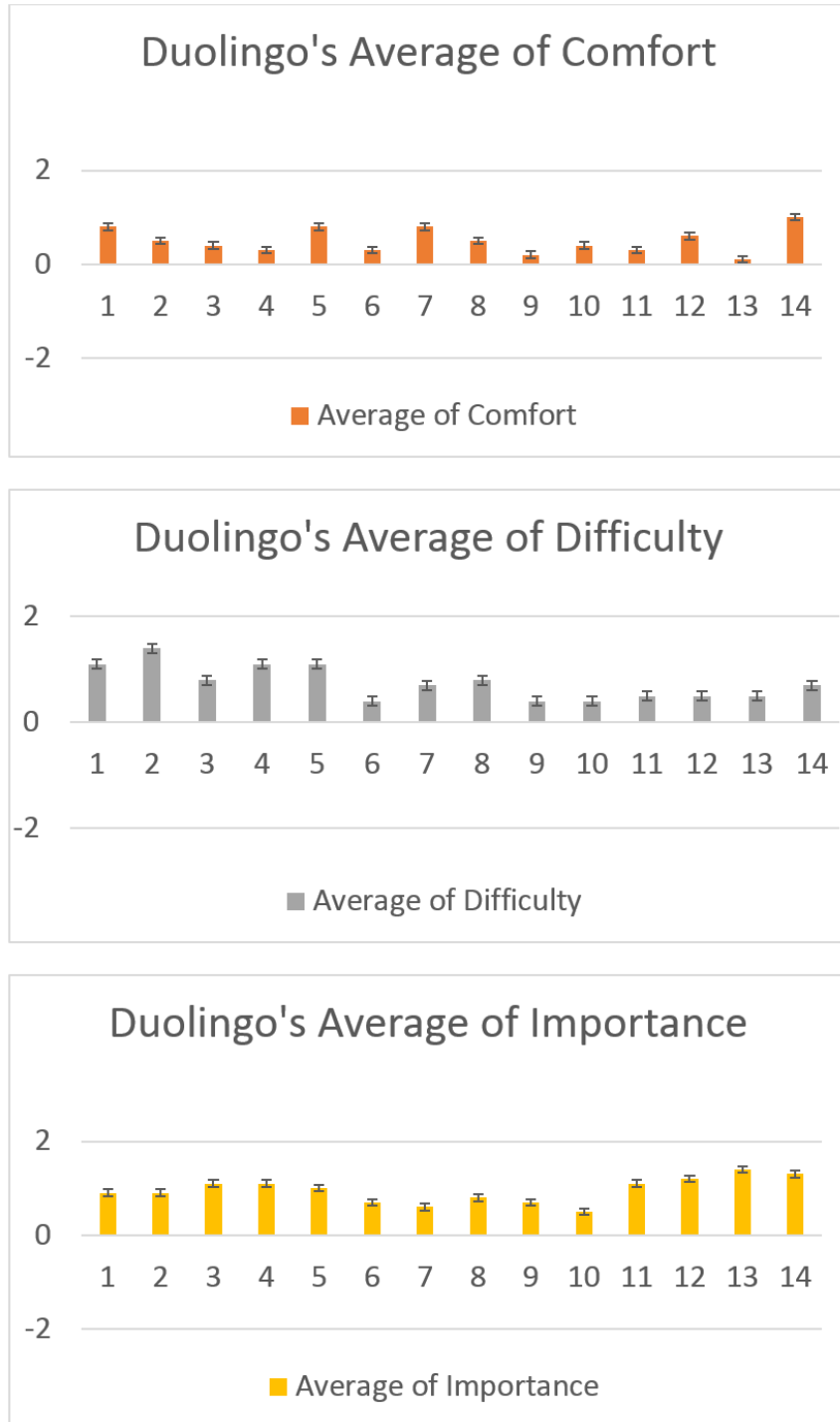


Figure 3.5: The Second Experiment: An average of comfort (top), difficulty (middle) and importance (bottom) of Duolingo's privacy policy in $(-2, 2)$ range.

differences on three variables compared with the first study.

Results

The result once again does not show any significant difference from the first study. Almost all of result still fall into the same range. In this study, we get all individual segments' comfort, difficulty and importance in (0,1) range instead of (-0.5, 0.5) range in the first study.

Discussion

The result does not support our hypothesis. The result seems to be unusable again because of the same range of distribution. We got the same result as the previous experiment again, and it might be worse because all of them fell into range (0, 1) instead of (-2, 2), as we expected. Even though we removed some answers after manually examining the summaries, results still belong to (0, 1) range.

We also tried to answer the questions and use it to compare with the results. Our answers uniformly fall in (-2, 2) range. Therefore, we have two hypotheses. First, Duolingo involves less important information, so people do not rate comfort, difficulty, or importance negatively. Second, we might need more workers to create a better result. According to our hypotheses, we proceed the next study with financial institution privacy policy.

3.2.3 The Third Experiment - A More Sensitive Service with M&T Bank's Privacy Policy

This study used the interface and questions as same as the second study, except this time we used M&T Bank's privacy policy, a financial institution, instead of Duolingo's privacy policy. We changed the privacy policy we used because we thought that more people would think that the M&T Bank privacy policy is more important and would be more likely to raise concerns. We believe that its service involves

important information. For example, we felt that the segment below is important and should make consumers more sensitive about security.

When you send --- an email, we may retain your message and our response for quality assurance purposes. It is important to remember that regular (non-encrypted) email is not secure. If you are sending us any personal or account information, please only do so through our secure email function within --- Web Banking. To find the secure email service, sign on to --- Web Banking, select the "Customer Service" tab and then choose the 'Contact Us' option.

Moreover, we also observe the variance value of each question in the case that we might find something interesting. We conduct this study in order to test our hypothesis, so we did not post all segments for evaluation. We post 10 assignments on 6 different segments from M&T Banks privacy policy.

Hypothesis

We expect a better result compared to the second study. At least, consumers should rate some segments important (more than 1) with some significant differences along a whole privacy policy.

Results

The result once again does not show any significant difference from the first and second studies. Almost all of result still fall into the same range. In this study, we get all individual segments' comfort, difficulty and importance in (0,1) range instead of (-0.5, 0.5) range in the first study.

Discussion

The result does support our hypothesis. As we expected, the result for average importance is little bit high compared to the second study. However, we still cannot expect to use this data to make a good visualization. We also manually removed

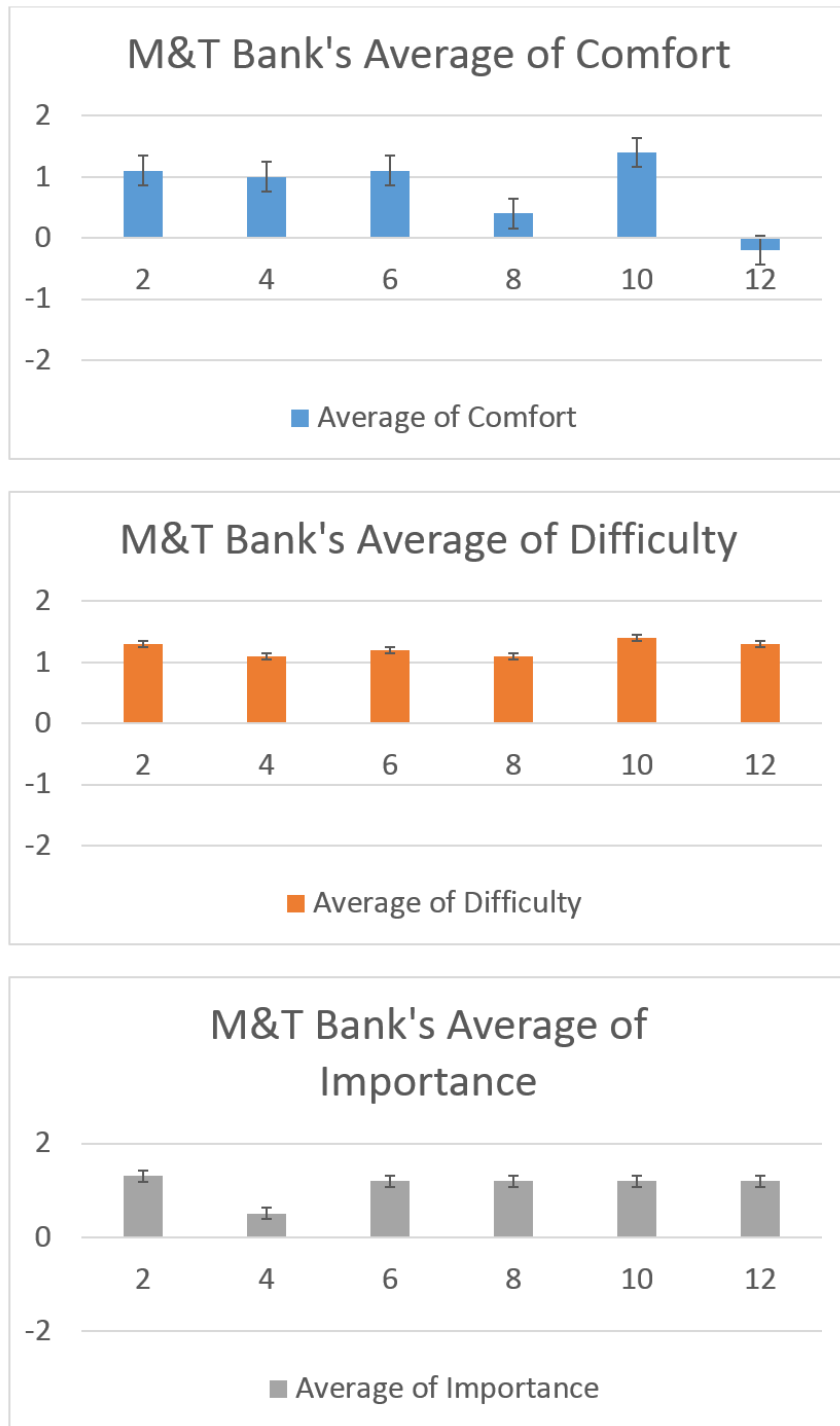


Figure 3.6: The Third Experiment: An average of comfort (top), difficulty (middle) and importance (bottom) of M&T Bank's privacy policy in $(-2, 2)$ range.

some answers regard to the qualities of main idea summaries as same as the second study, but we did not find any significant difference after doing that too. Even though we see one negative value in the average of comfort, we still think it is near the neutral level. Therefore, we continue to the next study with a larger number of workers according to another hypothesis from the second study.

3.2.4 The Fourth Experiment - A Higher Number of Assignments with M&T Bank's Privacy Policy

This study is based on the third study, but we increased the number of max assignments on Amazon Mechanical Turk. Instead of 10 assignments per segment like previous studies, we post 30 assignments of two segments as 60 assignments in total with the same reward. We tried to post the task at the different time routine in order to restrict the same workers as much as possible. Once again, we run this experiment as a feasibility study to test our hypothesis, so we do not post all segments.

Hypothesis

We expect a better variation of the result because the bad results should be mixed with good results or we should be able to find out the consistency of results as same as previous studies.

The both number of max assignments do not have any significant differences. So, we can see that our results are quite consistent.

Discussion

The result does not support our hypothesis. According to the result, we still did not see any significant difference from both cases again. We concluded that the number of workers does not have a huge impact on our results. So we will continue our study with 10 assignments later on. At this point, we doubt that the importance of information might not be only a factor that users will be concerned. We have a

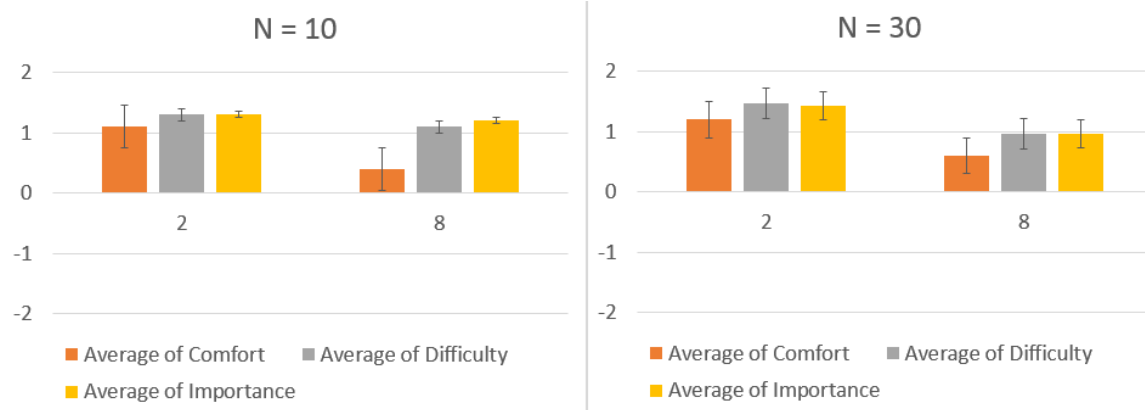


Figure 3.7: The Fourth Experiment: The compared results of $n = 10$ assignments (left) and $n = 30$ assignments (right) of M&T Bank's privacy policy in $(-2, 2)$ range.

hypothesis to test a more difficult privacy policy instead.

3.2.5 The Fifth Experiment - A More Difficult Privacy Policy with USCIS's Privacy Policy

Based on the hypothesis in the fourth study discussion, we conduct this study with USCIS privacy policy, which has legal terms in the contents. We think that USCIS privacy is difficult. We observed a following segment as an example.

"If we store your PII in a record system designed to retrieve information about you by personal identifier (name, personal email address, home mailing address, personal or mobile phone number, etc.), so that we may contact you, we will safeguard the information you provide to us in accordance with the Privacy Act of 1974, as amended (5 U.S.C. 552a). The Act requires all public-facing sites or forms that request PII to prominently and conspicuously display a privacy notice."

This privacy policy embed some legal terms which we believe that it is more difficult than other studies we conducted. We also examine workers time spent on each assignment to observe it as another factor of lazy workers. We post four segments in total with ten assignments per segment as 40 assignments in total.

Hypothesis

We expect a better result compared to the second study. At least, consumers should rate some segments embedded with legal terms difficult (more than 1). We also expect to get some significant differences in results too.

Results

Even though we used a more difficult privacy policy, the result once again does not show any significant differences from the previous studies. Almost all of result still fall into the same range. In this study, we get all individual segments' comfort, difficulty and importance in (0,1) range instead of (-0.5, 0.5) range in the first study.

Discussion

Unfortunately, the result does not support our hypothesis. We do not find any significantly difficult segment (only one negative segment). Moreover, we observe that time spent on each segment is also reasonable for the task expected some extreme value, such as 10 seconds or 30 minutes on one task. According to the time spending value on each assignment, we do not think that Amazon Mechanical Turk workers are lazy workers. So, we have to find out what makes the result look like this with other hypotheses. We have a new hypothesis that workers might be confused due to the long segment we used.

3.2.6 The Sixth Experiment - Smaller Segments of USCIS's Privacy Policy

This study is based on the fifth study with USCISs privacy policy but we conduct this study with smaller size segments. We divided one normal size of segment (< 100 words per segment) into three small segments (< 40 words) and post 10 assignments per segment again on Amazon Mechanical Turk. So, we post 30 assignments in total with the same interface. We provided the rest of segments for user to navigate as

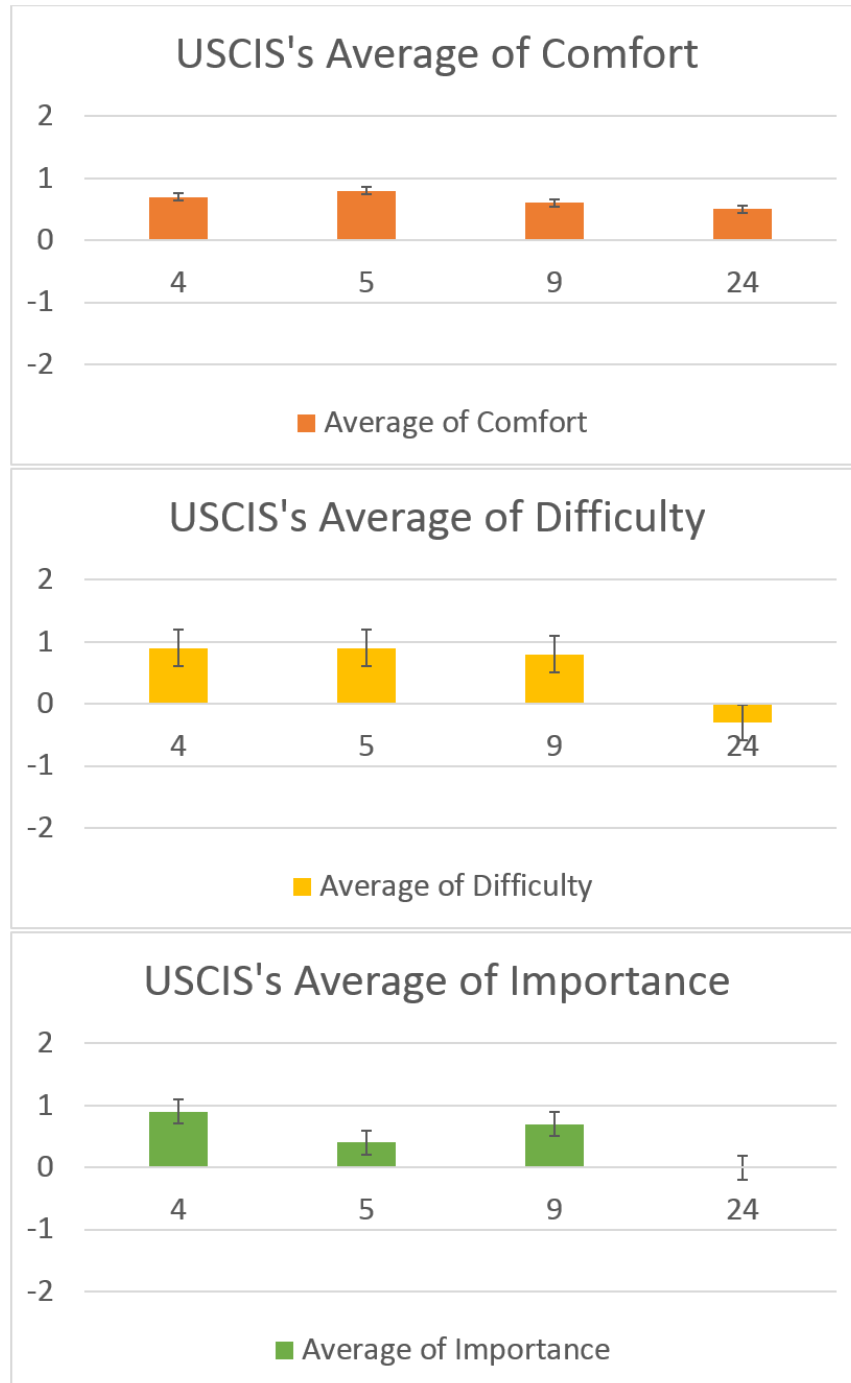


Figure 3.8: The Fifth Experiment: An average of comfort (top), difficulty (middle) and importance (bottom) of USCIS's privacy policy in $(-2, 2)$ range.

well.

Hypothesis

Since shorter segments are different from longer one, these segments might be easier but more confusing in contexts. We expect a significant to see difference between the normal-sized segments and average of three small-sized segments.

Results

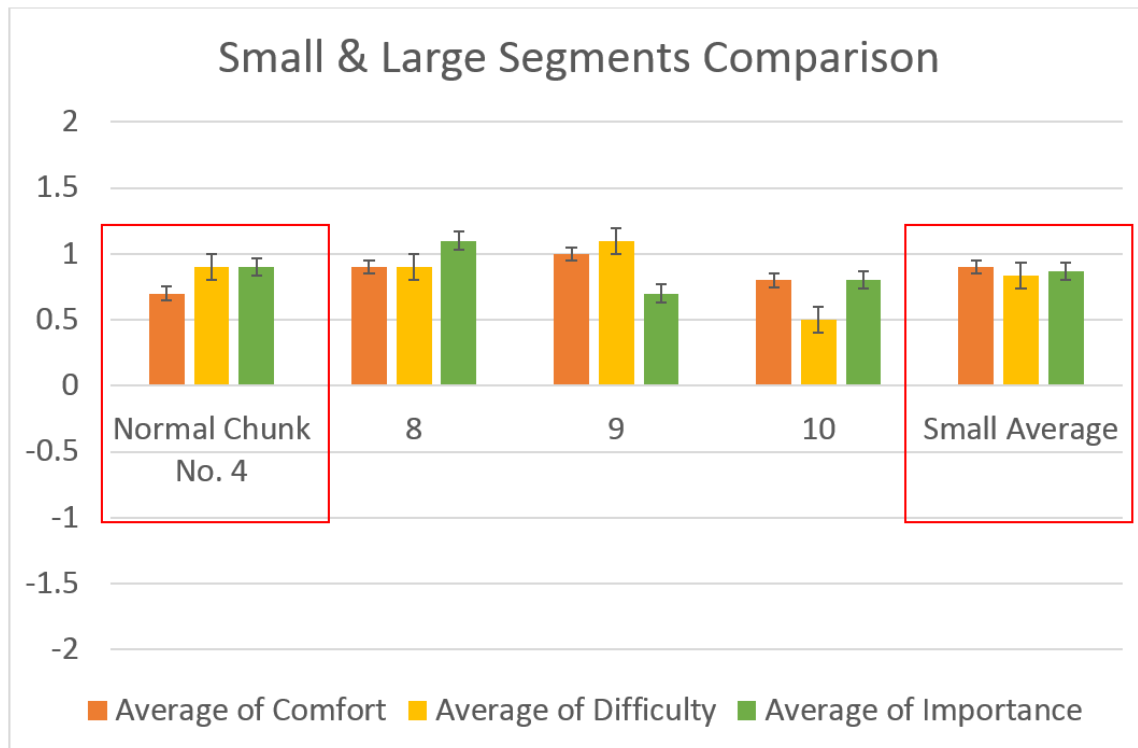


Figure 3.9: The Sixth Experiment: A comparison graph of a normal segment and 3 small segments of USCIS’s privacy policy in $(-2, 2)$ range.

The results do not show any significant difference between normal size segments and small size segments.

Discussion

The result does not support our hypothesis. We have the same problem again. We also do not observe any significant difference from shrinking the segment size. We can conclude that sizes of segments do not have an impact on the results.

3.2.7 The Seventh Experiment - More Surprising Use of Information with Wikimedia's Privacy Policy

After we used four different privacy policies in previous studies, we think that Wikimedia has some surprising segments. Wikimedia mentions the location data collection in the privacy policy in the following segment.

If you consent, we can use GPS (and other technologies commonly used to determine location) to show you more relevant content. We keep information obtained by these technologies confidential, except as provided in this Policy. You can learn more by checking out the list of examples of how we use these technologies in our FAQ."

We use the same interface as previous studies. We post 13 segments with ten assignments per segment. So we posted 130 assignments in total.

Hypothesis

We expected to see a negative value for average of comfort, especially in the segment about location data and the data collection. At the same time, we expect the neutral range in other values (importance and difficulty), which would be the same as previous studies.

Results

Even though we used a privacy policy with more surprising use such as location data collection, the result does not show any significant differences from the previous studies. Almost all of result still fall into the same range. In this study, we get

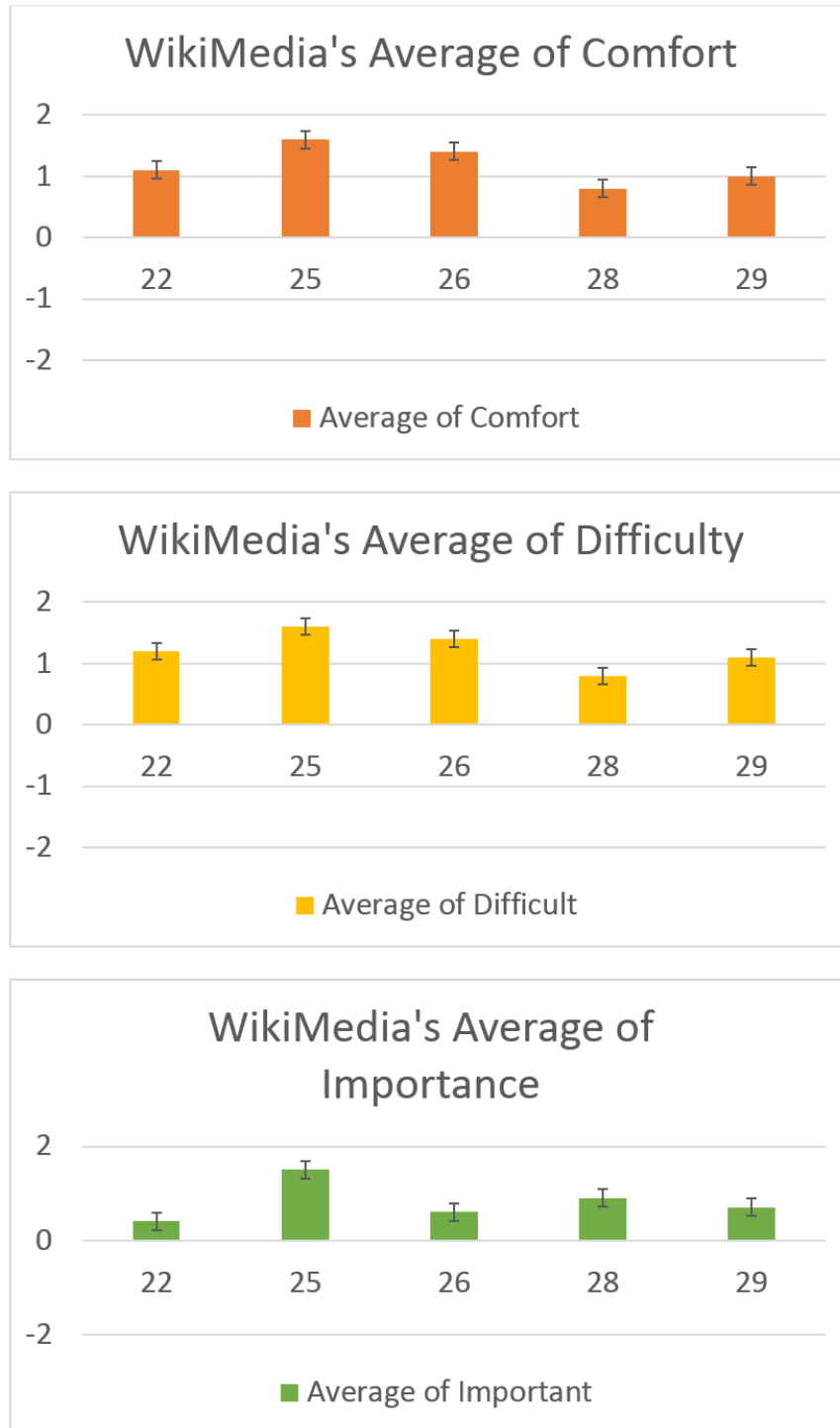


Figure 3.10: The Seventh Experiment: A partial graph of average of comfort (top), difficulty (middle) and importance (bottom) of Wikimedia's privacy policy in $(-2, 2)$ range.

all individual segments' comfort, difficulty and importance in (0,1) range instead of (-0.5, 0.5) range in the first study.

Discussion

The result does not support our hypothesis. Once again, we cannot obtain a good distribution from results. We still cannot use this result to generate the visualization, like a heat map, as we expected. According to seven studies, we have another hypothesis that the user might hesitate to answer the extreme value of -2 or 2 in any questions. So, we continue with another study.

3.2.8 The Eighth Experiment - Binary Scale of Questions with Wikimedia's Privacy Policy

According to the seventh study's discussion hypothesis, for this study, we changed the number of answers in the last three questions into the binary answer. In this study, we also use Wikimedia's privacy policy as same as the seventh study as well.

How **important** is the information in the highlighted segment in using the web site?
For example, is it something you would want a close friend to tell you before you use the web site?

- ☐ *Important*
- ☐ *Unimportant*

How **comfortable** would you be with the policy in the highlighted segment, if you were using this web site?

- ☐ *Comfortable*
- ☐ *Uncomfortable*

How **understandable** is the text in the highlighted segment?

- ☐ *Easy to understand*
- ☐ *Hard to understand*

Figure 3.11: An interface of binary questions we asked crowdsourced workers

We post 13 more segments as same as the seventh study with 10 assignments per segment. We apply the following normalization formula to previous results.

We define binary answer value as only '0' and '1' In order to do a comparison of two types of question, we apply the normalization formula on previous results. We are able to normalize both data into (0, 1) range by following formula:

$$Normalized = \frac{(i - min)}{(max - min)}$$

Hypothesis

Since we normalized both data, we could not use results to generate any visualization. We did not expect to see any significant differences but we expect to see some significant difference between both types of questions.

Results

We do not see any significant differences between both types of questions. Some of differences seem to be large but a majority of them are quite small.

Discussion

The result does not support our hypothesis. Indeed, we see some differences in the difficulty and importance but they are small numbers of them. We cannot conclude that binary question can provide a better result by this.

3.2.9 Result Summary & Discussion

The result we got from 8 studies with 5 different privacy policies were not as usable as we expected, making it hard to create an efficient heat map visualization. We verified workers with a main idea summary and time spent. We also tried to change the number of assignments, different websites, different sizes of segments and different scales of questions. Nonetheless, the results we got from these studies cannot produce the efficient heat map. We need to find an alternative solution for data use or even a new approach to ask workers' opinion.

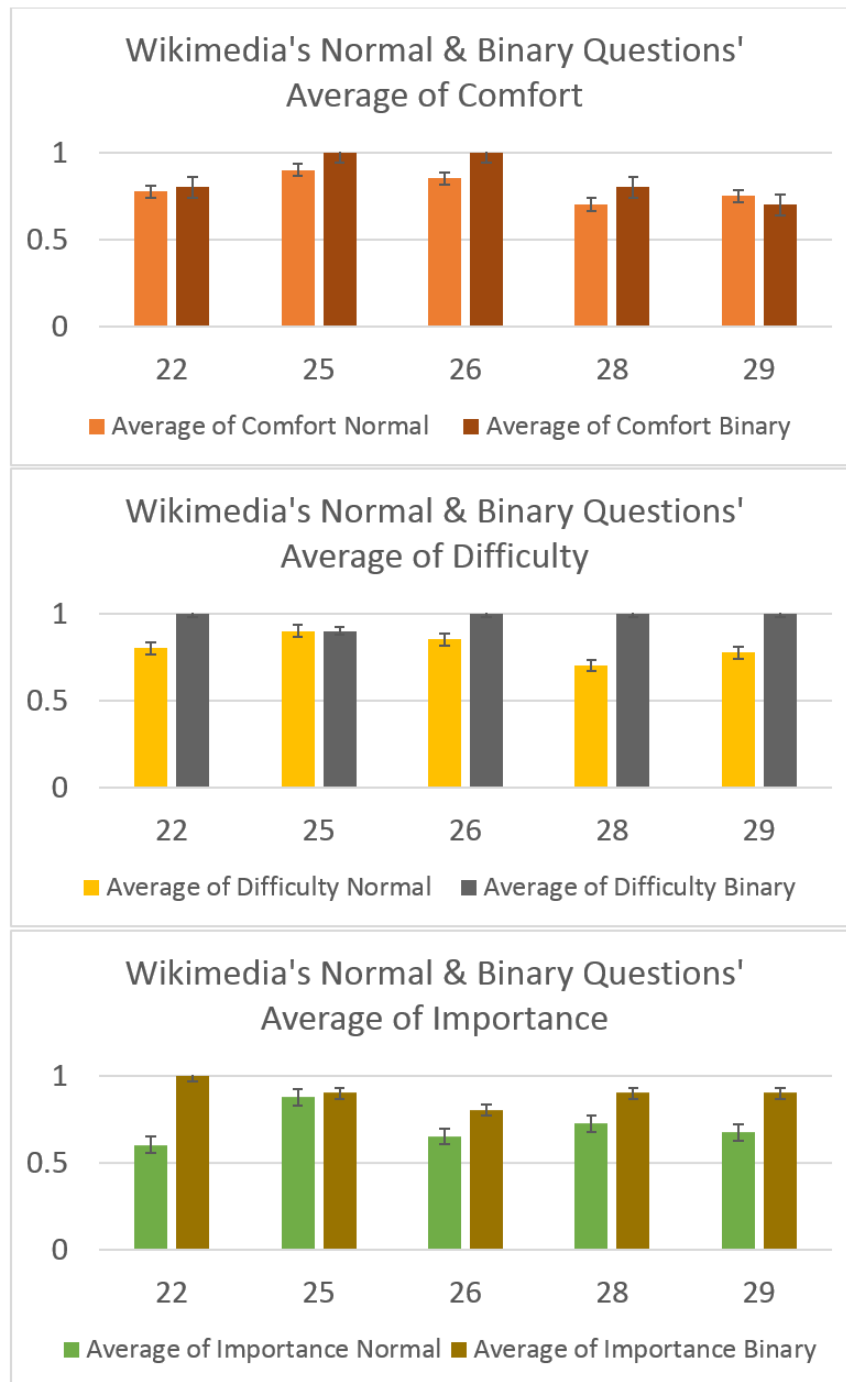


Figure 3.12: The Eighth Experiment: A partial comparison graph of average of comfort (top), difficulty (middle) and importance (bottom) of Wikimedia's privacy policy in normalized (0,1) range.

#	Privacy Policy	Description	Total of Assignments	Lessons Learned
1	Klout	- Asked crowd workers to rate 5-scale (-2,-1, 0, 1, 2) of comfort and readability of 7 segments of privacy policy.	70	- 11 of 14 results fell into a neutral range (-0.5, 0.5).
2	Duolingo	-We suspected lazy workers. - In order to detect lazy workers, we ask crowd workers to summarize privacy policy's main idea. - We also increase reward to \$0.10 per assignment.	140	- Results have little variance in score, and all of them fell into range (0,1). - Main idea question could reduce some bad answers, but it is unusable for whole policy summary.
3	M&T Bank	- We suspected workers ignored unimportant information in services, so this time, we use service involving financial information.	60	- Average of importance is higher than previous study, but results still have little variance. - Variance of comfort, readability and importance do not have any significant difference.
4	M&T Bank	- We added more workers per segment to reduce the potential impact of lazy workers.	60	- No significant difference between n=10 and n=30 on both segments.
5	USCIS	- We used a more difficult privacy policy to see if that could increase variance in answers from crowd workers. - This study also observes time spent on each segment.	40	- Results show no significant variance among different segments. - Time spent on each segment is legitimate.
6	USCIS	- We divided segments into smaller sizes.	30	- No significant difference between normal-sized segment and an average of small-sized segments.
7	Wikimedia	- We used a privacy policy involving more sensitive location data.	130	- Results still have no significant variance.
8	Wikimedia	- We used binary questions instead of a 5-point scale to force users to make a clear decision on comfort, readability, and importance.	130	- Normalized results of binary questions do not have any significant difference compared with normalized results of 5-point questions.

Table 3.2: Summary Table of Eight Experiments

4

Conclusion

In this cyber world, we need privacy policies to aid users' understanding in privacy practices. Even though privacy policies tend to be difficult and time-consuming to read, they are still one of the few tools consumers have. There are many previous works trying to help users in privacy practices, such as P3P, privacy seals, and privacy nutrition labels. However, none of them could replace a whole privacy policy. Our research also aimed to simplify a privacy policy by using crowdsourcing intelligence, such as Soylent and Lin et al.'s works.

We divided privacy policies into small segments and asked crowd workers to read and evaluate each segment. We divided five privacy policies (Klout, Duolingo, M&T Bank, USCIS and Wikimedia) into small segments (less than 100 words). Then, we posted our tasks to Amazon Mechanical Turk, an online crowdsourcing marketplace, to evaluate characteristics, such as the comfort level, difficulty in reading, and importance of segments. We conducted eight studies in order to obtain results that are good enough to visualize privacy policies. In total, we had 660 assignments.

However, our finding shows that people consistently rated these privacy policy segments as important, easy to read, and not surprising, which is not what one

would expect based on previous works. We suspected lazy workers in our results. Therefore, we attempted to design a new interface, post more assignments, observe time spent, shrink the size of segments, and change scales of question in order to resolve this issue within five different privacy policies. Although we can conclude that not all of them are lazy workers from the time spent and main idea summary, none of the results have shown any significant variance yet. We cannot generate any useful heat map from our results. In addition, the main idea summary cannot be used because those summaries are too varied in terms of writing and require more work to summarize all of them in the same way.

Our research fails to find good results enough to generate visualized privacy policies by asking crowdsourced workers to evaluate each individual segments of privacy policies. We cannot find any significant variance in privacy policies. In order to achieve our objective, we need to modify crowdsourcing assignments into different approaches and find new ways to visualize the results. We would like to ask workers to compare and choose which segment is more surprising, more difficult and more important from two different segments. Hence, we are planning to try other methods such as the Elo rating algorithm [15, 22]. We will also modify the visualization of the results by flagging the top five segments that ranked lowest in comfort, highest in difficulty and highest in importance. We also plan to verify the accuracy of the results by comparing answers with privacy experts and evaluating the skimming time.

Bibliography

- [1] Amazon.com, Inc. Amazon mechanical turk requester. Website. Amazon.com, Inc. Accessed: 2013-10-016. [Online]. Available: <https://requester.mturk.com/tour>.
- [2] P. Benassi, “Truste: an online privacy seal program,” *Communications of the ACM*, vol. 42, no. 2, pp. 56–59, 1999.
- [3] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and K. Panovich, “Soylent: a word processor with a crowd inside,” in *Proceedings of the 23rd annual ACM symposium on User interface software and technology*. ACM, 2010, pp. 313–322.
- [4] R. B. Cialdini, “Harnessing the science of persuasion,” *Harvard Business Review*, vol. 79, no. 9, pp. 72–81, 2001.
- [5] Consumer Action, “Consumer action ”do not track” survey results,” 2013.
- [6] L. F. Cranor, “P3p: Making privacy policies more useful,” *Security & Privacy, IEEE*, vol. 1, no. 6, pp. 50–55, 2003.
- [7] —, “The economics of privacy: Necessary but not sufficient: Standardized mechanisms for privacy notice and choice,” *J. on Telecomm. & High Tech. L.*, vol. 10, pp. 273–445, 2012.
- [8] L. F. Cranor, S. Egelman, S. Sheng, A. M. McDonald, and A. Chowdhury, “P3p deployment on websites,” *Electronic Commerce Research and Applications*, vol. 7, no. 3, pp. 274–293, 2008.
- [9] M. J. Culnan and P. K. Armstrong, “Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation,” *Organization Science*, vol. 10, no. 1, pp. 104–115, 1999.

- [10] Daily Crowdsourcing. What is crowdsourcing? Website. Daily Crowdsourcing. Accessed: 2013-10-02. [Online]. Available: <http://dailycrowdsourcing.com/crowdsourcing-basics/what-is-crowdsourcing>.
- [11] W. Davis. (2012, June) Judge rejects amazon's bid to dismiss privacy lawsuit. Website. MediaPost. Accessed: 2013-10-13. [Online]. Available: <http://www.mediapost.com/publications/article/176077/judge-reject-amazons-bid-to-dismiss-privacy-laws.html>.
- [12] J. S. Downs, M. B. Holbrook, S. Sheng, and L. F. Cranor, "Are your participants gaming the system?: screening mechanical turk workers," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 2399–2402.
- [13] Duolingo. (2012, April) Duolingo: Privacy policy. Website. Duolingo. Accessed: 2013-11-06. [Online]. Available: <http://www.duolingo.com/privacy>.
- [14] R. Flesch, *The Art of Readable Writing*. John Wiley & Sons, 1994.
- [15] GOBASE.org. Introduction to the elo rating sytem. Website. GOBASE.org. Accessed: 2013-10-09. [Online]. Available: <http://gobase.org/studying/articles/elo/>.
- [16] Harris Interactive, "Privacy leadership initiative (pli) privacy notices research final results," *Retrieved July*, vol. 13, p. 2006, 2001.
- [17] J. Howe. (2006, Juner) Crowdsourcing: A definition. Website. WIRED. Accessed: 2013-10-02. [Online]. Available: <http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing.a.html>.
- [18] C. Jensen and C. Potts, "Privacy policies as decision-making tools: An evaluation of online privacy notices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '04. New York, NY, USA: ACM, 2004, pp. 471–478. [Online]. Available: <http://doi.acm.org/10.1145/985692.985752>.
- [19] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder, "A nutrition label for privacy," in *Proceedings of the 5th Symposium on Usable Privacy and Security*. ACM, 2009, p. 4.
- [20] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," 1975.

- [21] Klout. (2013, September) Privacy policy. Website. Klout. Accessed: 2013-11-06. [Online]. Available: <http://klout.com/corp/privacy>.
- [22] J. Lasek, Z. Szlávik, and S. Bhulai, “The predictive power of ranking systems in association football,” *International Journal of Applied Pattern Recognition*, vol. 1, no. 1, pp. 27–46, 2013.
- [23] P. G. Leon, L. F. Cranor, A. M. McDonald, and R. McGuire, “Token attempt: the misrepresentation of website privacy policies through the misuse of p3p compact policy tokens,” in *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*. ACM, 2010, pp. 93–104.
- [24] J. Lin, N. Sadeh, S. Amini, J. Lindqvist, J. I. Hong, and J. Zhang, “Expectation and purpose: understanding users’ mental models of mobile app privacy through crowdsourcing,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 501–510.
- [25] A. M. McDonald and L. F. Cranor, “Cost of reading privacy policies, the,” *ISJLP*, vol. 4, p. 543, 2008.
- [26] Microsoft Corporation, “Data privacy day privacy survey 2013,” 2013.
- [27] Miniwatts Marketing Group. (2012, June) Internet usage statistics. Website. Miniwatts Marketing Group. Accessed: 2013-10-30. [Online]. Available: <http://www.internetworldstats.com/stats.htm>.
- [28] T. Moores, “Do consumers understand the role of privacy seals in e-commerce?” *Communications of the ACM*, vol. 48, no. 3, pp. 86–91, 2005.
- [29] M&T Bank. (2012, June) M&t bank: Privacy policy. Website. M&T Bank. Accessed: 2013-10-09. [Online]. Available: <https://www.mtb.com/customerservice/Pages/PrivacyPolicy.aspx>.
- [30] I. Pollach, “What’s wrong with online privacy policies?” *Communications of the ACM*, vol. 50, no. 9, pp. 103–108, 2007.
- [31] A. Raskin. (2010) Privacy icons: Alpha release. Website. Aza On Design. Accessed: 2013-10-08. [Online]. Available: <http://www.azarask.in/blog/post/privacy-icons/>.
- [32] D. J. Solove, “Privacy self-management and the consent dilemma,” *Harv. L. Rev.*, vol. 126, pp. 1879–2139, 2013.

- [33] J. Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti, “The effect of online privacy information on purchasing behavior: An experimental study,” *Information Systems Research*, vol. 22, no. 2, pp. 254–268, 2011.
- [34] U.S. Citizenship and Immigration Services. (2012, September) Uscis - privacy policy. Website. Department of Homeland Security. Accessed: 2013-10-09. [Online]. Available: www.uscis.gov/Privacy_Policies.
- [35] Wikimedia. (2012, September) Privacy policy - meta. Website. Wikimedia. Accessed: 2013-10-09. [Online]. Available: http://meta.wikimedia.org/wiki/Privacy_policy/BannerTestA.

Appendix A

Question Interfaces

Instructions

Strongly positive	Select this if this part of privacy policy makes you feel extremely safe and relieve to use the service after you read this part.
Positive	Select this if this part of privacy policy makes you relieve to use the service but it still has some unexpected points within the context.
Neutral	Select this if this part of privacy policy is normal. (Nothing is good/bad within this context)
Negative	Select this if this part of privacy policy makes you a little bit surprised but

Judge the sentiment expressed by the following item toward: Klout's Privacy Policy

Segment: Providing additional information beyond what is required at registration is entirely optional, but enables you to better identify yourself and find opportunities in the Klout system.

○ — ○ — ○ — ○ — ○
Strongly Negative Negative Neutral Positive Strongly positive

You must ACCEPT the HIT before you can submit the results.

Figure A.1: Interface Provided by Amazon Mechanical Turk. Crowdsourced workers need to read a given individual segments and evaluate its readability or comfort in 5-point scale.

The text below shows the privacy policy for Duolingo.com

Please read the **highlighted segment** (segment no. 2/14) and answer the following questions.

Optional: You may use the hide/show buttons to show hidden segments.
You may also use navigation buttons to navigate and read other segments
if you need to understand the context better after you toggle the display.

By using, accessing or participating in the Service, you agree to the terms of this privacy policy (the "Privacy Policy"). Capitalized terms not defined in this Privacy Policy have the meanings set forth in the Terms and Conditions, located at http://____.com/#/terms.

We reserve the right to change our Privacy Policy at any time. If we do this, we will post a notice that we have made changes to this Privacy Policy on the Website for at least 7 days after the changes are posted and will indicate at the bottom of the Privacy Policy the date these terms were last revised.

Any revisions to this Privacy Policy will become effective the earlier of (i) the end of the foregoing 7-day period or (ii) the first time you access or use the Service after any such changes. If you do not agree to abide by this Privacy Policy, you are not authorized to use, access or participate in the Service.

Context Navigation: Up || Highlighted segment || Down

Hide Other Segments Show Other Segments

Please briefly summarize the main idea of the text in the **highlighted segment** in one or two sentences.

Please input your answer here...

How **important** is the information in the highlighted segment in using the web site? For example, is it something you would want a close friend to tell you before you use the web site?

- ☐ Very important
- ☐ Important
- ☒ **Neutral**
- ☐ Unimportant
- ☐ Very unimportant

How **comfortable** would you be with the policy in the highlighted segment, if you were using this web site?

- ☐ Very comfortable
- ☐ Comfortable
- ☒ **Neutral**
- ☐ Uncomfortable
- ☐ Very uncomfortable

How **understandable** is the text in the highlighted segment?

- ☐ Very easy to understand
- ☐ Easy to understand
- ☒ **Neutral**
- ☐ Hard to understand
- ☐ Very hard to understand

Submit

Figure A.2: Improved Interface with Main Idea Summary and Three 5-point Questions. We use this improved interface to conduct 6 experiments from second experiment to seventh experiment.

The text below shows the privacy policy for Wikimedia

Please read the **highlighted segment** (segment no. 5/51) and answer the following questions.

Optional: You may use the hide/show buttons to show hidden segments.
You may also use navigation buttons to navigate and read other segments
if you need to understand the context better after you toggle the display.

Like most things on the Internet, anything you share may be copied and redistributed throughout the Internet by other people. Please do not contribute any information that you are uncomfortable making permanently public, like the picture of you in that terrible outfit your mom forced you to wear when you were eight.

Account Information & Registration
Want to create an account? Great! Don't want to create an account? No problem! You are not required to create an account to read or contribute to a Wikimedia Site. However, if you contribute without signing in, your contribution will be publicly attributed to the IP address associated with your device.

If you want to create a standard account, we do not require you to submit any personal information to do so. Really. No name, no email address, no date of birth, no credit card information, nothing. Just a username that you make up and a password. Your username will be publicly visible, so please think carefully before you use your real name as your username. Your password is only used to verify that the account is yours.

Context Navigation: || ||

Please briefly summarize the main idea of the text **in the highlighted segment** in one or two sentences.

**REQUIRED: Bad quality summary might be rejected!!*

Please input your answer here...

How **important** is the information in the highlighted segment in using the web site? For example, is it something you would want a close friend to tell you before you use the web site?

- ☐ Important
☐ Unimportant

How **comfortable** would you be with the policy in the highlighted segment, if you were using this web site?

- ☐ Comfortable
☐ Uncomfortable

How **understandable** is the text in the highlighted segment?

- ☐ Easy to understand
☐ Hard to understand

Figure A.3: Binary Question Interface Based on Second Interface for the Eighth Experiment. We changed only a scale of question into binary scale instead of 5-point scale.

Appendix B

Experiments Data

Segments	Avg. of Comfort	Avg. of Difficulty
Providing additional information beyond what is required at registration is entirely optional, but enables you to better identify yourself and find opportunities in the Klout system.	0.8	0.4
If you are not a user of the Services and wish to opt out, please click here: http://klout.com/corp/optout . If you are a current user of the Services, you can opt-out by logging into your account and then accessing the link above.	0.4	1
Klout does not analyze non-public data from private networks (in other words, networks that are password-protected) unless you authorize us to access such data.	0.4	0.8
Our Services analyze data that is made available publicly through the Internet in order to measure influence and create a score ("Klout Score") for Internet users, including, but not limited to, our registered users.	0.4	0.4
Personal Information that we may collect in such instances may include your full user name, password, email address, city, time zone, telephone number, and other information that you decide to provide us with, or that you decide to include in your public profile.	0.2	-0.4
If you desire to have access to certain restricted sections of the Services and/or to have access to or use of certain parts of the Services, you will be required to become a registered user, and to submit certain Personal Information to Klout.	0.2	0.2
Klout presents some of this data on the Site and our mobile application(s), such data may also be available on other websites and Klout may share this data with our third party partners (as described in more detail below).	0	-0.2

Table B.1: The First Experiment Data: The Comfort and Difficulty of Individual Segments of Klout's Privacy Policy in $(-2, 2)$ Range

Segment ID	Avg. of Comfort	Avg. of Difficulty	Avg. of Important
1	0.8	1.1	0.9
2	0.5	1.4	0.9
3	0.4	0.8	1.1
4	0.3	0.8	1.1
5	0.8	1.1	1
6	0.3	0.4	0.7
7	0.8	0.7	0.6
8	0.5	0.8	0.8
9	0.2	0.4	0.7
10	0.4	0.4	0.5
11	0.3	0.5	1.1
12	0.6	0.5	1.2
13	0.1	0.5	1.4
14	1	0.7	1.3

Table B.2: The Second Experiment Data: Comfort, Difficulty and Importance of Individual Segments of Duolingo’s Privacy Policy in $(-2, 2)$ Range

Segment ID	Avg. of Comfort	Avg. of Difficulty	Avg. of Important
2	1.1	1.3	1.3
4	1	1.1	0.5
6	1.1	1.2	1.2
8	0.4	1.1	1.2
10	1.4	1.4	1.2
12	-0.2	1.3	1.2

Table B.3: The Third Experiment Data: Comfort, Difficulty and Importance of Individual Segments of M&T’s Privacy Policy in $(-2, 2)$ Range

Segment ID	Avg. of Comfort	Avg. of Difficulty	Avg. of Importance
2 ($n = 10$)	1.1	1.3	1.3
2 ($n = 30$)	1.2	1.47	1.43
Difference	0.1	0.17	0.13
8 ($n = 10$)	0.4	1.1	1.2
8 ($n = 30$)	0.6	0.97	0.97
Difference	0.2	0.13	0.23

Table B.4: The Fourth Experiment Data : Comparison Table between 10 Assignments and 30 Assignments of M&T Banks Privacy Policy

ID	Avg. of Comfort	Avg. of Difficulty	Avg. of Important	Avg. of Time Spent
4	0.7	0.9	0.9	1:57
5	0.8	0.9	0.4	2:12
9	0.6	0.8	0.7	1:51
24	0.5	-0.3	0	1:57

Table B.5: The Fifth Experiment Data: Comfort, Difficulty and Importance of Individual Segments of USCIS's Privacy Policy in $(-2, 2)$ Range and Average of Time Spent

Segment ID	Avg. of Comfort	Avg. of Difficulty	Avg. of Important
Normal Size	0.7	0.9	0.9
Small Size No. 8	0.9	0.9	1.1
Small Size No. 9	1	1.1	0.7
Small Size No. 10	0.8	0.5	0.8
Avg. of Small Segments	0.9	0.83	0.87
Difference	0.2	0.07	0.03

Table B.6: The Sixth Experiment Data: Comfort, Difficulty and Importance of Normal Size Segment and Small Size Segments of USCIS's Privacy Policy in $(-2, 2)$ Range

Segment ID	Avg. of Comfort	Avg. of Difficulty	Avg. of Important
21	1.6	1.5	1.1
22	1.1	1.2	0.4
23	1.1	1	0.3
24	1.3	1.1	0.4
25	1.6	1.5	1.5
26	1.4	1.4	0.6
27	0.9	1.6	1.1
28	0.8	0.8	0.9
29	1	1.1	0.7
30	1	1	0.9
31	1.1	1.4	1.2
32	0.8	1	0.8
33	1.3	1	1.4

Table B.7: The Seventh Experiment Data: Comfort, Difficulty and Importance of Individual Segments of Wikimedia’s Privacy Policy in (-2, 2) Range

#	Avg. of Comfort			Avg. of Difficulty			Avg. of Importance		
Segment ID	5-point	Bi.	Diff.	5-point	Bi.	Diff.	5-point	Bi.	Diff.
21	0.9	0.9	0	0.875	0.9	0.025	0.0775	0.9	0.125
22	0.775	0.8	0.025	0.8	1	2	0.6	1	0.4
23	0.775	8	0.025	0.75	1	0.25	0.575	0.9	0.325
24	0.825	0.8	0.025	0.775	0.9	0.125	0.6	1	0.4
25	0.09	1	0.1	0.09	0.9	0	0.875	0.9	0.025
26	0.85	1	0.15	0.85	1	0.15	0.65	0.8	0.15
27	0.725	0.9	0.175	0.9	1	0.1	0.775	0.9	0.125
28	0.7	0.8	0.1	0.7	1	0.3	0.725	0.9	0.175
29	0.75	0.7	0.05	0.775	1	0.225	0.675	0.9	0.225
30	0.75	0.8	0.05	0.75	0.9	0.15	0.725	0.9	0.175
31	0.775	0.7	0.075	0.85	1	0.15	0.8	1	0.2
32	0.7	0.7	0	0.75	1	0.25	0.7	0.8	0.1
33	0.825	0.7	0.125	0.75	0.8	0.05	0.85	0.9	0.05
Avg. of Difference			0.09			0.198			0.248

Table B.8: The Eighth Experiment Data : Comparison Table between 5-point questions and binary questions difference of Wikimedias Privacy Policy