NTNU
Norwegian University of
Science and Technology

Next generation privacy policy

Ole Kristian Lillebo

Master of Science in Computer Science
Submission date: June 2011
Supervisor: Lillian Røstad, IDI
Co-supervisor: Inger Anne Tøndel, SINTEF IKT

Norwegian University of Science and Technology
Department of Computer and Information Science

Abstract

Privacy policies are commonly used by service providers to notify users what information is collected, how it will be used and with whom it will be shared. These policies are however known to be notoriously long and hard to understand, and several studies have shown that very few users actually read them. Alternative solutions that accurately communicates the most important parts of the policy in a way that is more enjoyable to read, is therefore needed to aid the users in making informed decisions on whether or not to share information with a provider.

By following a design science strategy we first explore current solutions, and based on an initial evaluation we find the Nutrition Label to be the current approach best suited to base further work on. Through an assess and refine cycle we first evaluate the Nutrition Label based on usability literature, and propose a set of design criteria which is used as a basis for developing an alternative solution, entitled the Privacy Table. By following an iterative design process, we evaluate the Privacy Table in terms of accuracy, time-to-response and likeability through a pre-test, a laboratory experiment with 15 participants, and finally through an Internet experiment with 24 participants, where each iteration results in a re-designed version of the Privacy Table.

While we don't find clear evidence for any difference between the formats, we find indications for that they perform similarly in terms of accuracy and enjoyability. We discover several issues regarding the Nutrition Label where some are related to the terminology used, which could indicate that it would need modifications in order to be usable among non-native English speakers. We also suggest that future research on the Nutrition Label should focus on its usability rather than further expansion, and that it should be considered to base it on a more simplified underlying technology than the P3P language. Finally we find that a merged version of the Privacy Table and the Nutrition Label could be advantageous to use in relation with current and future privacy enhancing technologies, as a top layer to communicate the most important privacy practices.

Preface

This report presents the work with my Master Thesis written during the tenth and final semester of a Master Degree in Computer Science at The Norwegian University of Science and Technology (NTNU). The thesis was written in cooperation with, and as a part of the COPE project at Sintef ICT.

I would like to thank my co-supervisor at Sintef ICT, Inger Anne Tøndel, for the guidance and motivation throughout the semester. Your help has been invaluable for the completion of this thesis.

I would also like to thank the students from the course TTM4135 who participated in my laboratory experiment, as well as the ones who participated in the Internet experiment. Finally, I would like to thank my friends and family for your support and patience through this process.

Trondheim, 14. June 2011

Ole Kristian Lillebo

Contents

CHAPTER 1

Introduction

1.1 Motivation

Privacy online has become increasingly important as the trail of information we leave behind us is rapidly growing, and the potential misuse of this information is well recognized. To minimize the intrusion into a person's privacy caused by the collection, storage and use of personal data, the field of data protection has emerged. An important part of data protection is to communicate the rules regarding the use of this personal data to the user, and the privacy policy serve as communication channel in this context.

By communicating a providers privacy practices, the privacy policy has the potential of enabling the user to make more informed decisions on whether or not to share information with a provider. Research has however shown that the textual policies have failed to achieve this. As very few people actually read policies, questions have been raised regarding whether they have any effect of the user at all. There is also the question whether companies really are interested in further development of policies to better display their privacy practices, and initiatives from the research community is therefore of high importance to ensure a further development in the field. Several technological approaches have also been proposed, but they have often been criticized and suffer from lack of recognition and adoption.

A new solution for presenting privacy policies is therefore needed. This solution should capture the parts of the policy which is crucial for the user to understand, and also the parts which the user is interested in. An important aspect regarding such a solution is its visual presentation. By presenting the content in a way which is more enjoyable to read, we believe more people would read the notices, which in turn would increase the general understanding of online privacy practices. By also enabling for better comparison of privacy practices across sites, the solution has the potential of enabling users to make more informed decisions based on their privacy preferences, in situations where they have the option to choose between different service providers.

1.2 Background

The background for this master thesis is two-sided. Its foundation was laid through an in-depth study conducted in the preceding semester with the purpose of laying a theoretical platform on the subject, and its problem was derived from the current work of the privacy group at Sintef ICT, for whom the thesis was written in cooperation with.

1.2.1 In-depth study

The in-depth study laid a theoretical platform in terms of knowledge and competence in areas related to the privacy policy, and should be regarded as a pre-study for this thesis. Among topics investigated were common issues with today's textual policies and a look at the concepts of privacy concerns and trust. The complex structure of laws and regulations surrounding the content of the policies was also studied, as well as current approaches in the fields of privacy enhancing technologies (PETs) and privacy seals.

We concluded the in-depth study with a statement that a new and innovative approach for presenting privacy policies is severely limited by laws, regulations as well as other aspects, and that future approaches should focus on the concept of trust and how the policy content could be better presented in a way that could make the users more aware of their privacy online. We will refer to the in-depth study as [Lillebo, 2010] throughout the report.

1.2.2 Sintef ICT

The thesis was written in cooperation with, and as a part of the COPE project at Sintef ICT. The COPE project has the purpose of enhancing end-users' privacy by providing technology to clarify the consequences of information sharing, so that the gap between privacy concerns and actual behavior can be reduced. This goal will be achieved by developing a web browser plug-in allowing privacy preference specification, perform policy compliance assessments and display the result through an intuitive graphical user interface.

Given the broad field of the COPE project, several possible topics were discussed through the planning stages of the master thesis. We can roughly divide these topics in two main categories: the language or technology used in the underlying system, and the graphical user interface on top. As the underlying system of the proposed system is based on machine learning to generate privacy preferences, an area in which the author of this thesis has little or no knowledge in, combined with a personal interest of design and user interfaces, the latter were chosen as the thesis topic.

The choice of developing a user interface in order to better communicate the policy practices, and therefore closing the gap between privacy concerns and actual behavior, was also in line with the suggested further work from the in-depth study mentioned above.

1.3 Problem definition

In accordance to the above presentation of background and motivation for the thesis, the overarching problem definition of the project was defined to find an alternative way of presenting a privacy policy, which could be more usable for the users, and aid in quickly and accurately finding the information they are looking for. To be useful for the COPE project, this solution should be suitable for use with a privacy agent, and be able to present important parts of the policy to the user in a simple but concrete way, based on the users' preferences.

In order to evaluate the effectiveness of the alternative solution, we proposed the following evaluation criteria:

The solution should:

1. Make the general policy content more understandable,
2. Make the policy easier and more enjoyable to read
3. Highlight the most important privacy practices, while preserving details and nuances
4. Enable for easier comparison of privacy practices across sites
5. Be useful for the COPE project as well as current and future Privacy Enhancing Technologies (PETs)

The following paragraphs describe each criterion in detail.

1. With making the content more understandable we mean that users should be able to quickly understand a company's privacy practices, and be able to make informed decisions regarding their personal information. It is also important that the users should understand that they can control or limit the collection and use of personal information.

2. By presenting the privacy policy in a more attractive format one could also make the experience of reading and comparing policies more enjoyable, potentially increasing policy reading rates and thus improving the general comprehensibility of online privacy practices.

3. As textual policies widely differ across sites, both in length and content, a challenge for an alternative presentation is how to handle the trade-off between keeping

a simple design while preserving the details and nuances of different policies. A new policy design can either serve as a replacement for or as a complimentary solution to the textual policy.

4. It is also important to make the user understand that privacy practices can differ across sites. By improving policy comparability, the user could make a more informed decision in cases where he/she has to choose between similar sites.

5. The solution should also be relevant and useful for the proposed privacy agent in the COPE project, as well as for other current and future Privacy Enhancing Technologies (PETs). Previous PET technologies have suffered from lack of adaption, and it is therefore important to ensure that the solution is capable of supporting both current and future privacy agents, while being simple enough to enable easy implantation on top of a complex underlying system.

1.4 Scope

It is important to make a distinction between the content of the privacy policy, and its visual presentation. As we found in the in-depth study, the global patchwork of laws and regulations that defines the privacy policy content, any new solution could never cover all these details and nuances. The new solution should therefore act as an additional solution, or an aid in better communicating the content of the entire policy text to the user.

Based on this, the focus in this master thesis is therefore on how the content is presented, and not on the content itself. While design criterion 4 stated that the policy should preserve the details and nuances of different policies, this is only on a general basis and conducting an analysis of the content in today's privacy policies is out of the scope for this paper.

Criterion 5 also stated that the solution should be useful for the COPE project and current and future PETs. This is also on a general basis, and it is out of the scope for this project to specifically implement a solution on top of any privacy policy language or privacy agent.

1.5 Report structure

Figure 1.1 presents the structure of this master thesis. Following the introduction in **Chapter 1**, we elaborate on the methodology in **Chapter 2**. We then present current solutions for presenting privacy policies through a pre-study in **Chapter 3**, before presenting the design process in **Chapter 4**. The evaluation method is elaborated in **Chapter 5**, and the results are presented in **Chapter 6**. We then discuss the findings in **Chapter 7**, before ending the master thesis by concluding our work and suggesting which areas future research should focus on in **Chapter 8**.

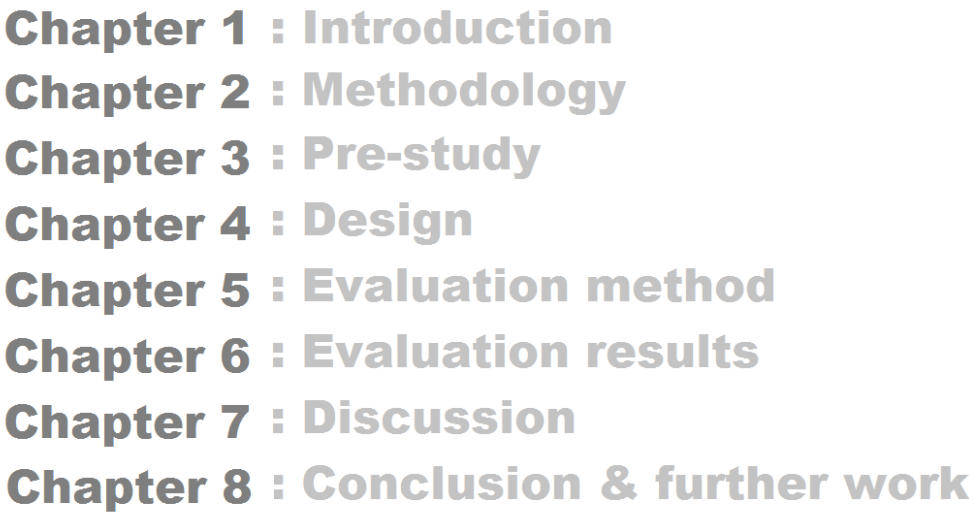


Figure 1.1: Report structure

CHAPTER 2

Methodology - design science

In order to answer the problem definition and develop an alternative solution for presenting privacy policies, we chose a design science research stratgey. The following chapter first give an an introduction to the theory behind the methodology, before presenting the methodology framework and a set of guidelines for design science proposed by Hevner et al. [2004]. Where not referenced, all theory and figures regarding design science is taken from Hevner et al. [2004], and for a full description of the concepts we refer to the same paper.

2.1 Design and behavioral science

The overarching research strategy for this thesis is design science. According to Oates [2006], a design and creation research strategy focuses on developing new IT products, also called artifacts. An artifact can be a constructs, model, method or an instantiation, and the explicit intention of design science is improving the functional performance of the artifact [Oates, 2006].

Hevner et al. [2004] argues that design is both a process (set of activities) and a product (artifact), and March and Smith [1995] identify two design processes: build and evaluate. Purposeful artifacts are built to address heretofore unsolved problems and they are evaluated with respect to the utility provided in solving those problems. The evaluation of the artifact then provides feedback information and a better understanding of the problem, in order to improve both the quality of the product and the design process. This build-and-evaluate loop is typically iterated a number of times before the final design artifact is generated. According to March and Smith [1995], progress is achieved in design science when existing technologies are replaced by ones that are more effective.

An IT artifact, implemented in an organizational context, is often the object of study in Information System (IS) behavioral-science research. While design-science is fundamentally a problem solving paradigm, behavioral science seeks to develop and justify theories (principles and laws) that explain or predicts organizational and

human phenomena surrounding the analysis, design, implementation, management and use of information systems [Hevner et al., 2004].

As already mentioned, the object of research in information systems is to acquire knowledge that enables the development of solutions to heretofore unsolved business problem [Hevner et al., 2004]. While behavioral-science approaches this goal through development and justification of theories that predicts the phenomena that occur, design-science approaches the goal through the construction of innovative artifacts aimed at changing the phenomena that occurs. This theory challenges design-science researchers to create artifacts that enable organizations to overcome the acceptance problems predicted. Hevner et al. [2004] further argues that there is opportunities for IS research to make significant contributions by engaging the complimentary research cycle between design-science and behavioral-science to address fundamental problems faced in the productive application of information technology.

2.2 Conceptual framework

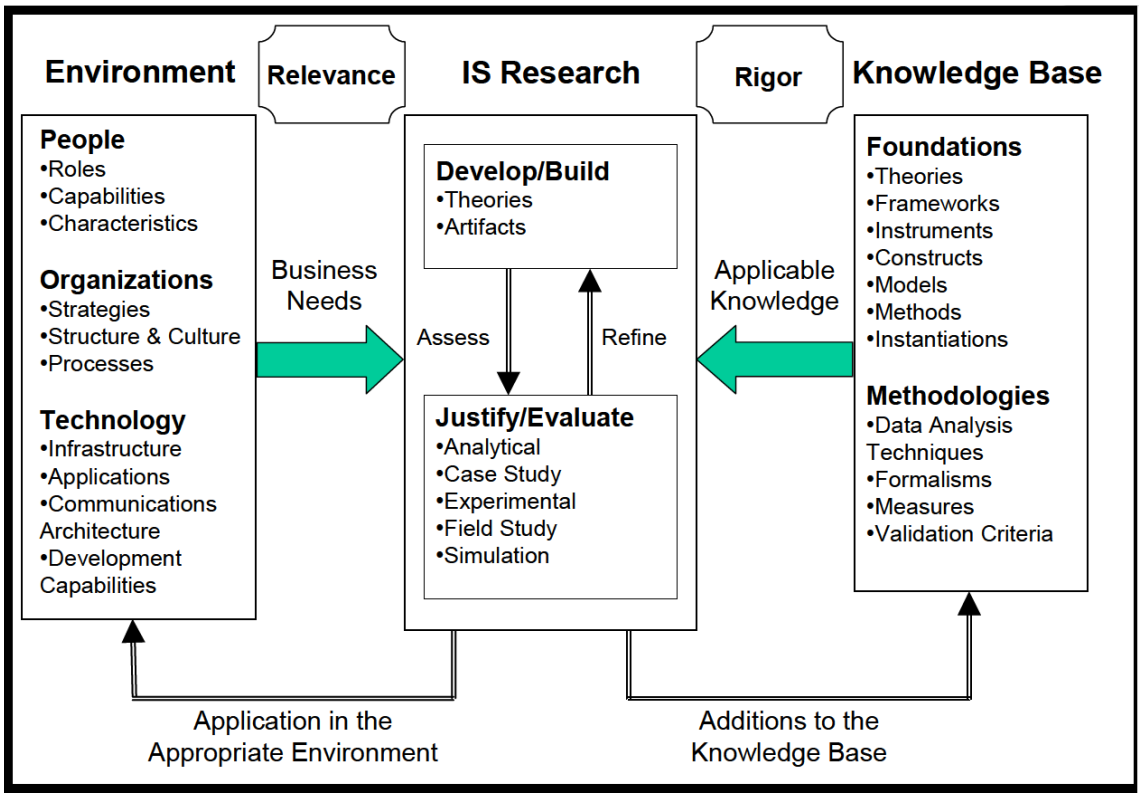


Figure 2.1: Information Systems Research Framework as proposed by Hevner et al. Figure reused from Hevner et al (2004).

Based on this, Hevner et al. [2004] proposed a conceptual framework for understanding information systems research, and a set of guidelines for conducting and evaluating good design-science research. This framework and guidelines has the purpose of informing researchers how to conduct, evaluate, and present design science

research, and serves as a basis for the work throughout this report.

Figure 2.1 presents the proposed research framework for combining behavioral and design-science paradigms. We will discuss the framework through the proposed design guidelines and relate the various sections to our research.

2.3 Guidelines

An important aspect in the framework is the knowledge base. The knowledge base provides the raw materials from and through which IS research is accomplished. The knowledge base consist of foundations such as theories, frameworks, models or instantiations used in the development and build phase, and methodologies which provides guideline through the justification and evaluation phase of the research study [Hevner et al., 2004].

Through the in-depth study on privacy practices leading up to this master thesis we investigated relevant theories regarding privacy online, the current situation of privacy related tools and technologies, and also what roles companies and businesses has in this continuous development. Together with relevant methodologies and an initial pre-study, these findings formed the knowledge base for both the development and build phase and the justification/evaluation phase throughout the master thesis.

2.3.1 Guideline 1: Design as an artifact

A key element in design-science is the artifact. These artifacts are rarely full-grown information systems, but rather innovations that define the ideas or products which can be accomplished through design and implementation. For such an artifact to be a significant result, it must be a serious question to whether it is possible to construct it, and the outcome must be important to the IS community and contribute to the knowledge base [Hevner et al., 2004].

A new and innovative approach of presenting a privacy policy serves as the artifact for our research. As mentioned in the introduction, few approaches have been made in this area and the ones that excist has suffered from low adoption and lack of recognition. As the majority of webpages today display a regular textual policy, which have been found to be read by less than 10% [Govani and Pashley, 2005], we argue that an innovating approach for presenting the privacy policy would be worthwhile and contribute knowledge to the IS community.

2.3.2 Guideline 2: Problem relevance

The second guideline states that the objective of design-science research is to develop solutions to important and relevant business problems. This is demonstrated by

the left-hand side of figure 2.1. These problems, or business needs, are defined by the goals, tasks and opportunities as they are perceived by the people in the organizations, and framing research activities to address these needs assures research relevance.

The obvious business need for our artifact is its role in the COPE project by Sintef ICT. The proposed web-based plugin depends on the display of its output through an intuitive graphical user interface. This graphical interface also defines and sets the boundaries for which areas of privacy are presented to the user.

We also believe there is be a general business need for an innovative approach of presenting privacy policies. Through the in-depth study which led to this thesis, we discovered that companies and businesses might have several interests in systems to better communicate privacy practices to their customers. Trust is an important factor in this regard, and small factors could influence whether a customer would feel secure or not by using the service provided by the company. It has for example been found that the mere professionalism of a site might contribute more to a consumers perception of the safety of a site than that the site displays a privacy policy. The widespread use of privacy seals demonstrates that companies are willing to pay a premium to demonstrate to its customers that it handles its privacy issues seriously.

2.3.3 Guideline 3: Design evaluation

The third guideline states that the utility, quality and efficacy of the designed artifact must be rigorously demonstrated via well-executed evaluation methods [Hevner et al., 2004]. The business environment forms the requirements for which the evaluation of the artifact is evaluated by, and the evaluation also includes the integration of the artifact within the technical infrastructure. IT artifacts can be evaluated in terms of functionality, accuracy, performance, reliability, usability among other relevant attributes, and are evaluated by using methods from the knowledge base.

Such evaluation methods can be observational, testing, descriptive, analytical or experimental. As further described in the evaluation chapter (Chapter 5 on page 55), the latter were relevant for the evaluation of our artifact. A dynamic analysis can study artifacts in use for dynamic qualities (such as accuracy and performance), and a controlled experiment enables the study of an artifact in a controlled environment (for qualities such as usability).

Hevner et al. [2004] highlights that the design phase is iterative and incremental, and the evaluation provides feedback to the construction phase. A design is only complete when it satisfies the requirements of the problem it was meant to solve. Through three evaluation processes, the pre-test, the laboratory experiment and the Internet experiment, we iteratively improved the design based on received feedback. The final design was then evaluated against the initial problem requirements and validation criteria to investigate whether we had proposed a satisfactory solution.

2.3.4 Guideline 4: Research contributions

Guideline 4 states that effective design research must provide clear contributions in the areas of the design artifact, design construction knowledge and/or design evaluation knowledge [Hevner et al., 2004]. Most often the contribution is the artifact itself, and exercising the artifact in the appropriate environment produces significant value to the IS community. Other contributions can be improving or extending the existing foundations in the knowledge base, or providing contributions to design science research in the forms of components such as measures and evaluation metrics.

Criteria for assessing contribution include that the artifact must be implementable, and that it demonstrates a clear contribution to the business environment, solving an important unsolved problem. The contribution of the artifact design is discussed in section 7.6 on page 151.

2.3.5 Guideline 5: Research rigor

As relevance is relevant for ensuring that the artifact solves a problem addressed by a business need, rigor is relevant for addressing the way in which research is conducted. Hevner et al. [2004] argues that it is possible and necessary for all IS research paradigms to be both relevant and rigorous. Rigorous methods must be applied in both the construction and evaluation of the design artifact, and rigor is derived from the effective use of the knowledge base - theoretical foundations and research methodologies. Success depends on the researcher's skilled selection of appropriate technologies to develop an artifact and selection of appropriate means to evaluate it.

As previously explained, our initial knowledge base consisted of the findings from an in-depth study of current online privacy practices, from we which together with relevant methodologies and an initial pre-study formed an initial criterion list for selecting which current solution best solved the outlined problem definition need (see Chapter 3 on page 15). A thorough analysis based on usability literature was then performed on the best solution, in order to evaluate in which areas the design could be improved (see Chapter 4 on page 35). By using theoretical foundations outlined in the in-depth study and methodologies derived from usability literature to evaluate current solutions, and to construct a new artifact, we believe rigor was ensured throughout the construction phase.

The selection of appropriate evaluation technologies was mainly based on a study of relevant research where similar variables were tested in similar artifacts. As explained in the Evaluation chapter (Chapter 5 page 55), recreating previous experiments was an important decision in order to ensure validity and generalizability of the results. The outcome and success of the evaluation method selection is discussed in section 7.6 on page 151.

2.3.6 Guideline 6: Design as a search process.

Simon [1996] describes the design process as a generate/test cycle, where alternative designs are generated and then tested against requirements. This design cycle describes our iterative process which was mentioned above. We began by testing the best current privacy policy solution against a requirement list, before generating an alternative design based on these findings, and then testing the alternative design in a pre-test, before generating an improved design based on this, and so on.

2.3.7 Guideline 7: Communication of Research

The final guideline suggests that design-science must be presented effectively to both technology-oriented as well as management-oriented audiences. Technology-oriented audience need sufficient details to be able to construct the artifact and used within an organizational context. It is also important for such audiences to understand how the artifact was created and evaluated, which establishes repeatability of the research project and builds the knowledge base for further research. For management -oriented audiences it is also important to focus on the importance of the problem and the effectiveness of the solution approach realized in the artifact.

We have therefore thoroughly documented the design and evaluation process throughout the thesis, and included all relevant info regarding both processes, as well as detailed results, in the appendices.

2.4 Design process

Figures 2.2, 2.3 and 2.4 illustrates the design process throughout the subsequent chapters.

Following this chapter we first evaluate current solutions for presenting privacy policies through a pre-study in **Chapter 3**. Based on an evaluation of these findings, we find that the **Nutrition Label** is the current solution which best suits or initial criteria list outlined in the introduction, and we therefore base further work of the new solution on its design.

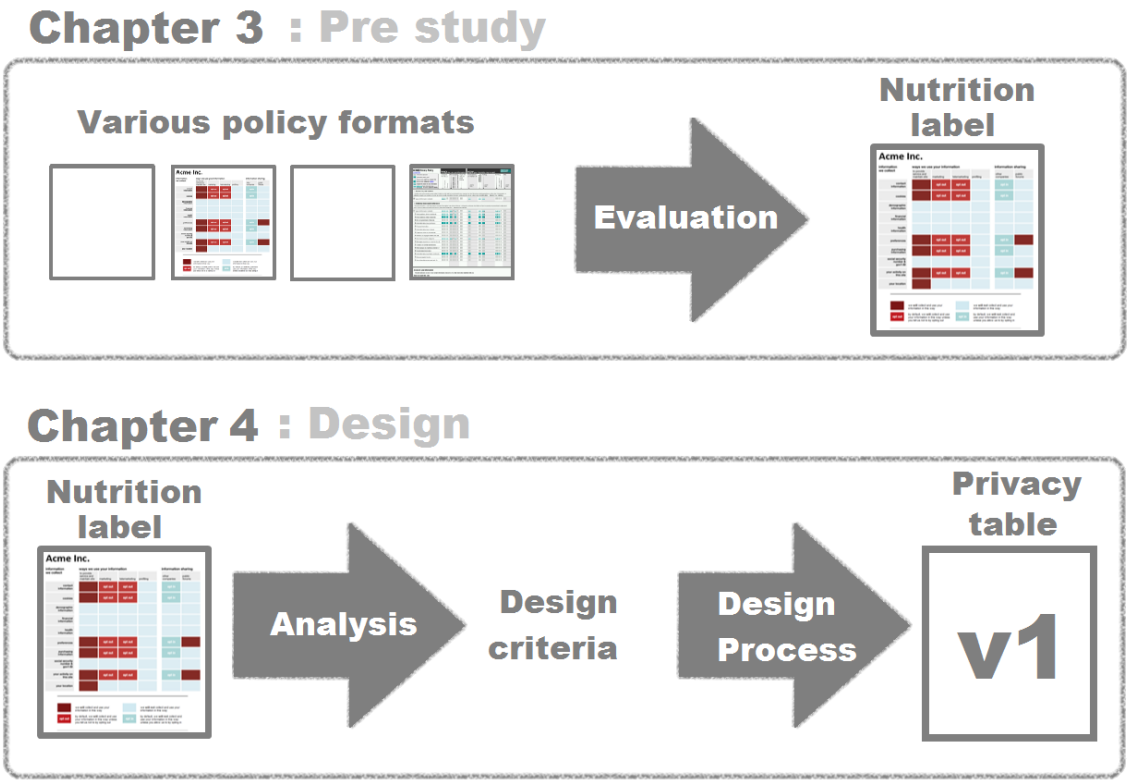


Figure 2.2: Overview of chapter 3 and 4

Through the design phase in **Chapter 4** we use usability literature to analyze the Nutrition Label, which results in a list of design criteria. This list of criteria is then used to create an alternative solution, entitled the **Privacy Table**, which is evaluated through a iterative design cycle as described in guideline 7. In addition to a pre-test, we conduct a laboratory experiment and a Internet experiment, which all result in a redesigned version of the Privacy Table (**Chapter 5**).

The results from the evaluation are presented in **Chapter 6**, before they are discusses in **Chapter 7**. Through the discussion we evaluate whether the proposed solution has solved the intitial problem, and whether the contribution we have made is relevant for both the COPE project and to the privacy community as a whole. Based on this we present a merged solution between the Nutrition Label and the

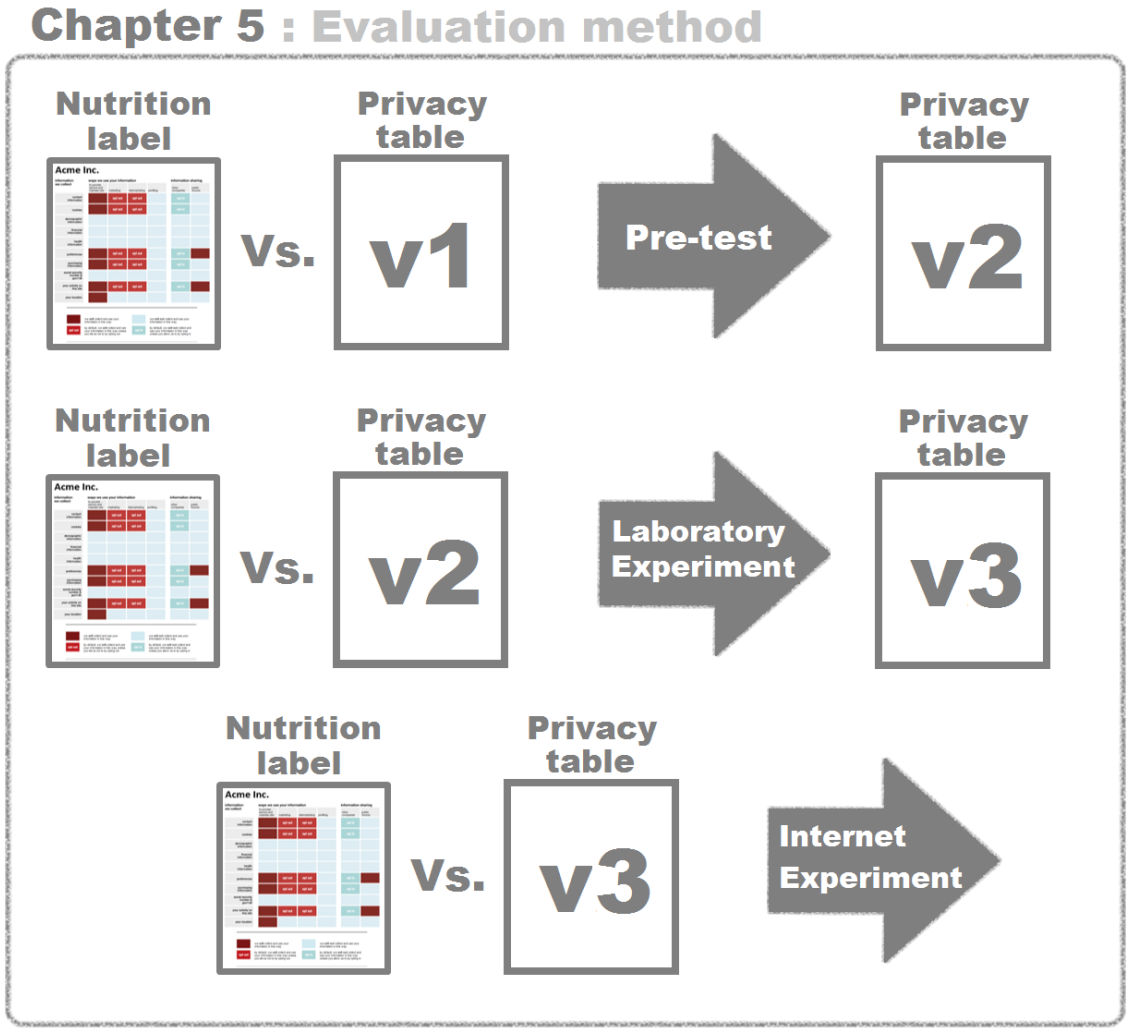


Figure 2.3: Overview of chapter 5

Privacy Table, which we believe captures the best features of both formats. We end this report by concluding our work and suggesting which areas future research should focus on in **Chapter 8**.

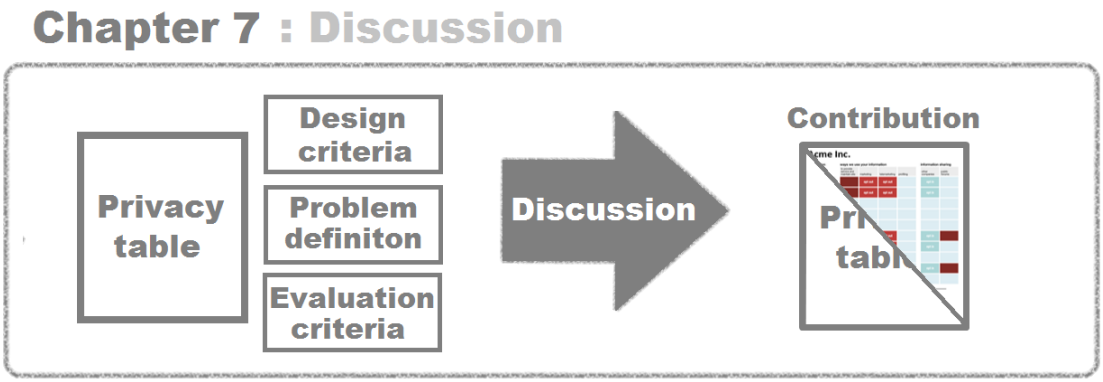


Figure 2.4: Overview of chapter 7

CHAPTER 3

Existing solutions

The following chapter presents the findings from the initial pre-study, which served the purpose of investigating current approaches for presenting privacy policies. The findings have been categorized into three main sections: layered policies, pictograms and standardized solutions.

3.1 Layered policies

The issues regarding full text policies was thoroughly elaborated through our in-depth pre study [Lillebo, 2010], and this section therefore focuses on the alternative layered solution.

Layered privacy notices first appeared on US business websites in 2003, and companies began to publish multilayered notices in different languages in 2005. The European Union (EU) provided the following recommendations for layered notices through Article 29 common position of December 2004 [Hunton and Williams, 2005]:

- Layer 1 - The short notice: the very minimum, e.g. contact details, and the purposes of processing.
- Layer 2 - The condensed notice: covering the basics in less than a page, ideally using subheadings, and covering Scope; Personal information collected; Uses and sharing; Choices (including any access options); Important information; How to contact us
- Layer 3 - The full notice

According to [Hunton and Williams, 2005], which proposed ten steps for creating multi-layered notices, these layered notices would be a "win" for organizations, consumers, as well as for regulators. It was elaborated that companies would "win" because multilayered notices easily built consumer trust, and that it had been found

that consumers preferred template-based condensed notices to longer text-based versions. Among other advantages with multilayered notices was easy translation for policies in different languages, and that the resulting notice was usable for consumers because they"like information that is clear, graphically appealing, and easy to compare" [Hunton and Williams, 2005].

Among companies which currently provide a layered notice is Microsoft Corp, as seen in figure 3.1. Their layered notice consist of five blocks of information, which all provides links to the corresponding information in the full text policy. The layered Microsoft policy can therefore be seen on as a interactive table of contents.

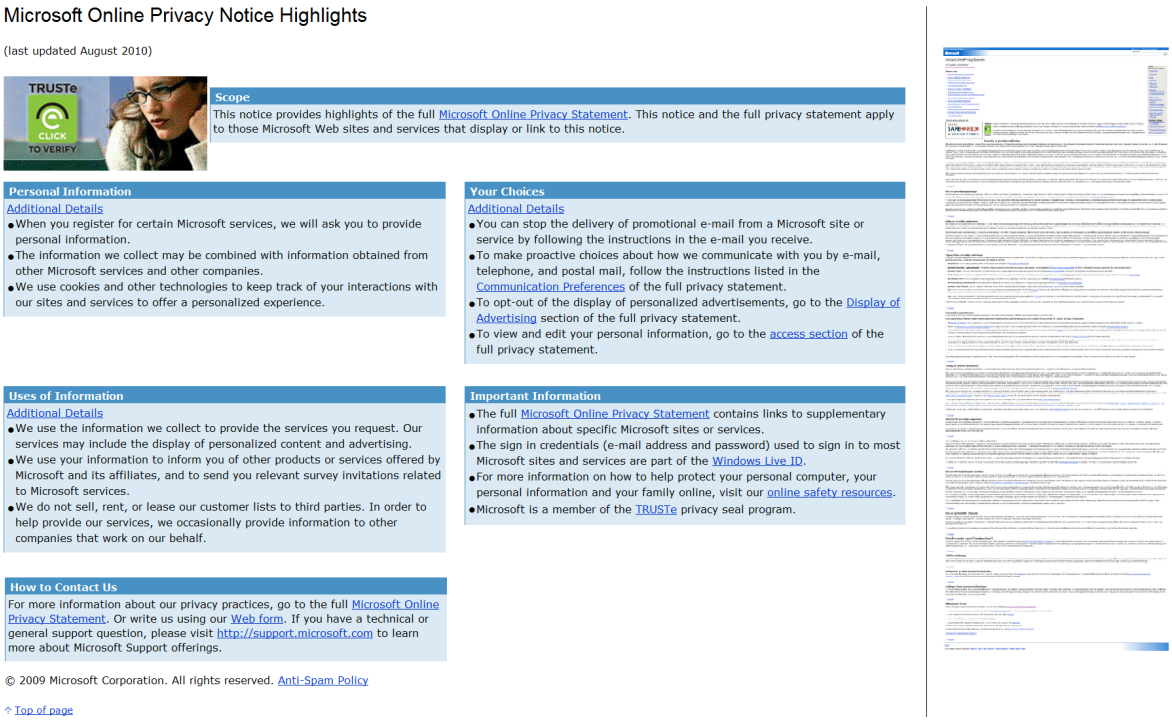


Figure 3.1: Screenshot of the Microsoft layered privacy notice (left) and the corresponding full text policy (right)

Another variation for a layered notice is a suggestion from the Privacy Seal providor TRUSTe [Pinnick, 2011]. According to Pinnick [2011], there had been general agreement on a Mozilla hosted Privacy Icon Workshop that a short notice needed to be much simpler than previous attempts in order to be effective. One suggested way of simplifying the notice was to remove things that the users probably already knew, such as the types of data collected (i.e. name, age, financial or purchase information), and instead focus more on the data practices and uses which were invisible to users (such as secondary use, data sharing, third party tracking, data retention) [Pinnick, 2011].

The resulting suggestion for a layered notice can be seen in figure 3.2, where data using and sharing is in the center of attention. The overall structure is similar to the Microsoft policy, with links to the corresponding content in the full text-policy.

As for the Microsoft notice, the first layer provides an organized view of the most important content, and has direct links to the corresponding content in the second layer.

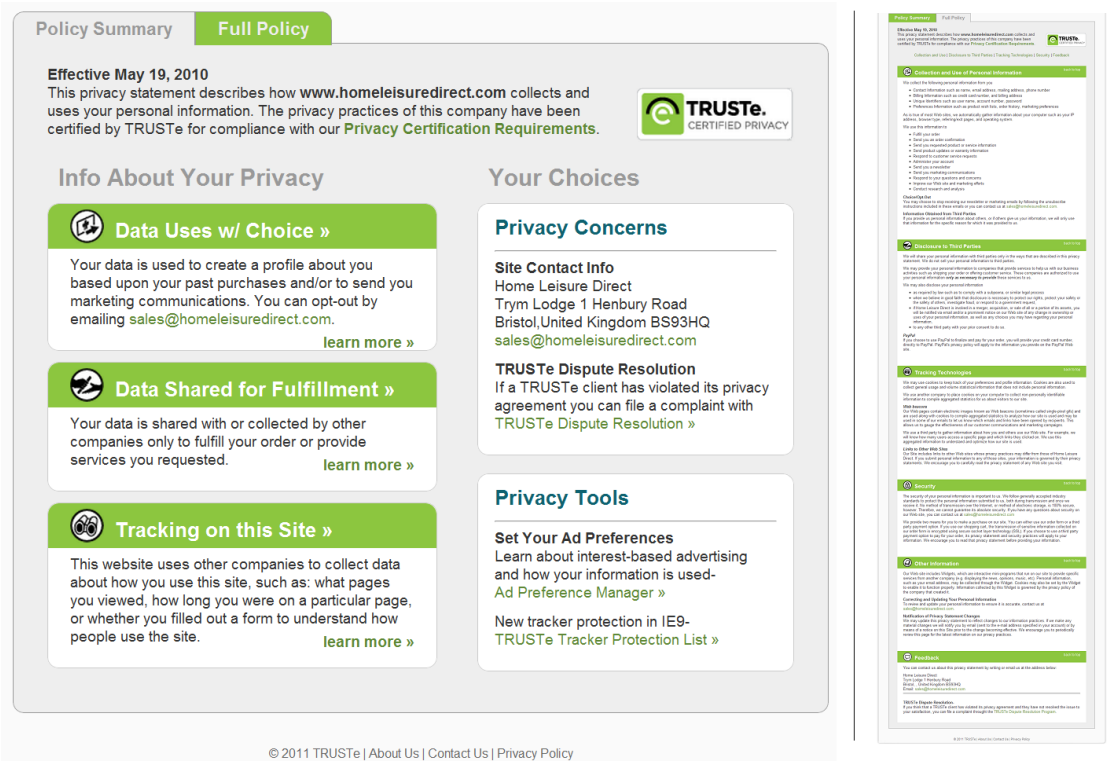


Figure 3.2: Screenshot of the suggestion of a layered notice by TRUSTe (left) and the corresponding full text policy (right) [Pinnick, 2011]

While these layered notices can be categorized as being a standardized format (see section 3.3 on page 25) Kelley et al. [2010] argues that the only standard components are a tabular page layout and mandatory text for the section headers. Other design details and the text of each section are left to the discretion of each company. Kelley et al. [2010] also adds that the amount of information to include in a layered notice is left up to each company, with layered notices requiring consumers to click through to the full text policy.

Kelley et al. [2010] also measured the performance of a layered policy in their study on online privacy formats, and it was found to be scoring similarly to a textual policy in terms of accuracy, whilst being more enjoyable to use. It was also found that only 25 of 78 participants ever clicked through to the full policy text from the layered overview. The layered notice was however outperformed by the standardized formats (see section 3.3 on page 25) in terms of accuracy for single information finding questions, as well as for questions requiring the respondent to compare two policies.

3.2 Pictograms and icons

The concept regarding privacy seals was elaborated in the in-depth study [Lillebo, 2010],and the idea behind privacy seals is that a retailer can pay a fee to a seal-of-approval program administrated by seal granting authority, either a independent company or a industry group, and in return display a logo that certifies that the retailer will follow a set of standards to protect consumer privacy [Jensen and Potts, 2004]. The TRUSTe Privacy Seal icon for communicating that a website is safe for kids use is seen in figure 3.3.

Figure 3.3: The Privacy Seal icon for the TRUSTe granting authority. Copyright TRUSTe

Trust is an important concept in this regard, and by displaying a privacy seal a website or a company sends out a signal to its users that it values its privacy and that it intends to abide by its privacy policy [Tang et al., 2005]. The seal authority will provide a set of guidelines with requirements for what the companies' privacy policy must address to obtain the certification. In addition, the seal authority will provide a voluntary enforcement mechanism to assure that the site abides by its privacy practices, but the authority has no control over whether the site actually follows its privacy policy or not [Rifon et al., 2005].

We started the presentation of pictograms with the introduction of Privacy Seals because it is a good example of how privacy practices can be communicated through a pictogram. In theory, through displaying a single icon, the user is ensured that the entire privacy policy meets his or hers privacy preferences.

Several different approaches for expressing privacy-related statements through icons exists, and we can roughly distinguish current solutions in four main categories [Hansen, 2009]:

1. Pictograms representing a trust indicator or a legal evaluation, e.g. Privacy Seals as described above
2. Pictograms representing how well a policy statement matches the privacy preferences of a user
3. Multiple pictograms representing the content of a textual policy in terms of what will be collected, how it will be used and with whom it will be shared
4. Multiple pictograms as a part of a larger or advanced system to automatically represent a textual policy

3.2.1 Pictograms for preference matching (2)

An attempt in the second category in the list above is the Privacy Bird, which is a privacy preference agent based on the P3P language (described in [W3C, 2007] and introduced in section 3.3.2). The user is given feedback through an icon (see figure 3.4) for whether a site matches a set of pre-defined privacy settings. While preference matching and the P3P language is out of the scope of this paper, the representation of such information through a graphical interface is highly relevant. While a Privacy Seal only communicates whether a site follows a set of standards to protect consumer privacy, the Privacy Bird icon communicates whether the sites privacy practices matches the consumers own privacy preferences.

Figure 3.4: The Privacy Bird icon shown depending on the match between privacy preferences and the policy content. Copyright Privacy Bird

While it is no longer regarded as a preference matching policy tool, we also shortly introduce the Privacy Finder search engine, as it is related to the Privacy Bird functionality.

The Privacy Finder was originally based on the Privacy Bird, and displayed privacy bird icons next to regular search results [PrivacyFinder, 2011]. Given the discontinued development of the Privacy Bird (the latest supported operative system is Windows XP running Internet Explorer 5.01, 5.5, or 6.0. [1]), the latest version of the Privacy Finder orders results based on the retrieved sites privacy practices (see figure 3.5). A icon of four green boxes indicates that the website provides a good privacy policy, and the number of green boxes that are missing are proportional to the number of preference conflicts between the website's privacy policy and a list of preset privacy preferences [PrivacyFinder, 2011]. The absence of the privacy meter means that a valid computer-readable privacy policy (the P3P policy) could not be located.

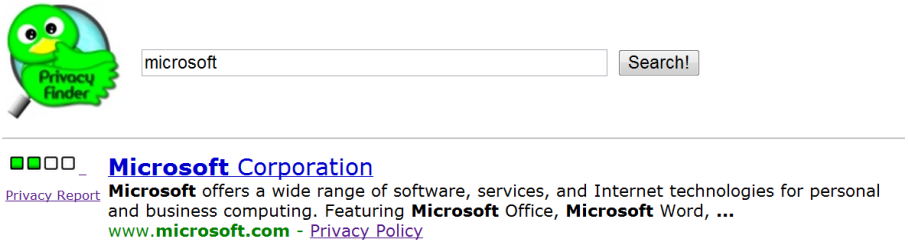


Figure 3.5: The Privacy Finder search result for "Microsoft"

By hovering over the icon a summarized list of the sites privacy practices is displayed, and by clicking the "Privacy Report" link below the icon, a standardized policy in the shape of a Nutrition Label (see section 3.3.3 on page 29) is displayed to the user.

[1] Ref. http://www.privacybird.org/ May 25th 2011

3.2.2 Multiple pictograms for representing the content of a textual policy (3)

Privacy Seals and the Privacy Bird are examples of single privacy pictograms. There have also been several attempts on multi-icon representations, where the overall purpose is to present the content of a textual policy in terms of icons. In this context, the Creative Commons icon set regarding copyright law has been an inspiratory source [Creative-Commons, 2011].

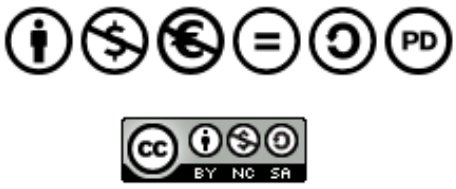


Figure 3.6: Subset of the Creative Commons icons, and an example license below. From left to right: Attribution, Noncommercial (US), Noncommercial (EU), Non Derivative, Share Alike and Public domain. Copyright [Creative-Commons, 2011]

Figure 3.6 displays a subset of the icon-set, where a combination of these is used to describe various degrees of restrictions. An example is given in the same figure (below), where the combination of symbols indicate an "Attribution-Noncommercial-ShareAlike" license, which lets "others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms" [Creative-Commons, 2011].

As mentioned, the Creative Commons icons has been a inspiratory source for several other similar attempts, for example the proposal by Mary Rundle in figure 3.8 which was presented at the United Nations Internet Governance Forum (IGF) in Athens in October 2006 [Rundle, 2006]. A more sophisticated version of a similar icon-set was suggested by Matthias Mehldau [Mehldau, 2007], and a small subset of it can be seen in figure 3.7.

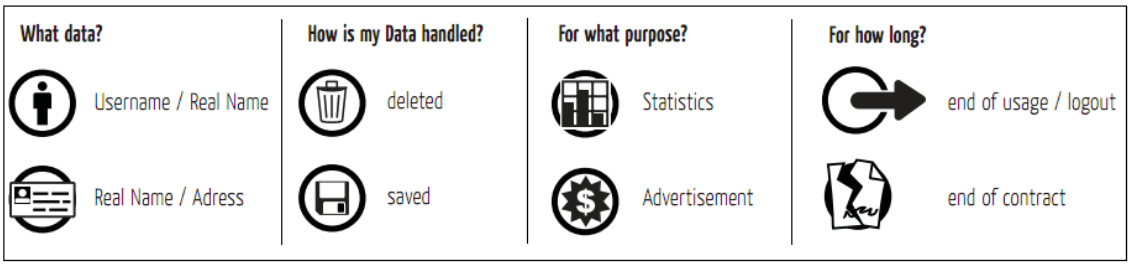


Figure 3.7: Subset of the proposal for a Iconset for data-privacy declarations, by Matthias Mehldau. [Mehldau, 2006]

	You agree not to use this data for marketing purposes.
	You agree not to trade or sell this data.
	You agree to submit to a third-party audit program on data use; if government has requested access to my data, you agree to involve my governmental ombudsman.
	You agree to make available to me the data that you have on me without my having to pay for it/at a minimal charge.
	You allow me to address inaccuracies in the data and request its removal.
	You agree to take reasonable steps to keep my data secure.
	You agree to arrange with X organization to help resolve any disputes we have over your treatment of this data. [The seal / name of the entity follows.]

Figure 3.8: Proposal for "Creative Commons-like icons" on privacy by [Rundle, 2006]

Several other suggestions for similar icon-sets has also been made, for example by Helton [2009] as seen in figure 3.9, and by Fischer-Hubner et al. [2010] in figure 3.10 as a part of the PrimeLife project [2].

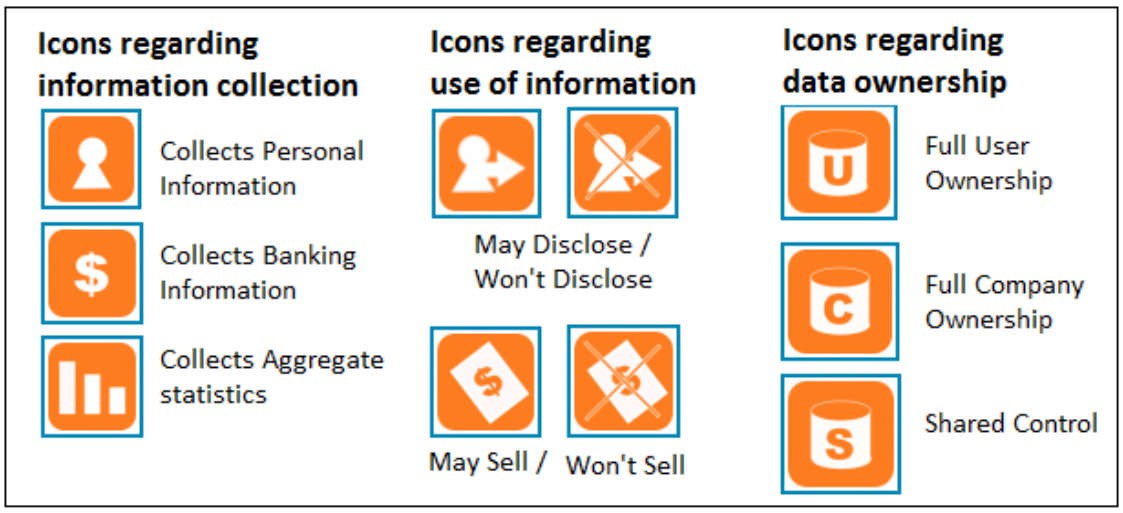


Figure 3.9: Mockup of the icons suggested by [Helton, 2009]

[2]PRIME and PrimeLife is two research projects funded by the European Commissions 7th Framework Programme, with a purpose of resolving core privacy and trust issues and ensuring that the community at large adopts privacy technologies.

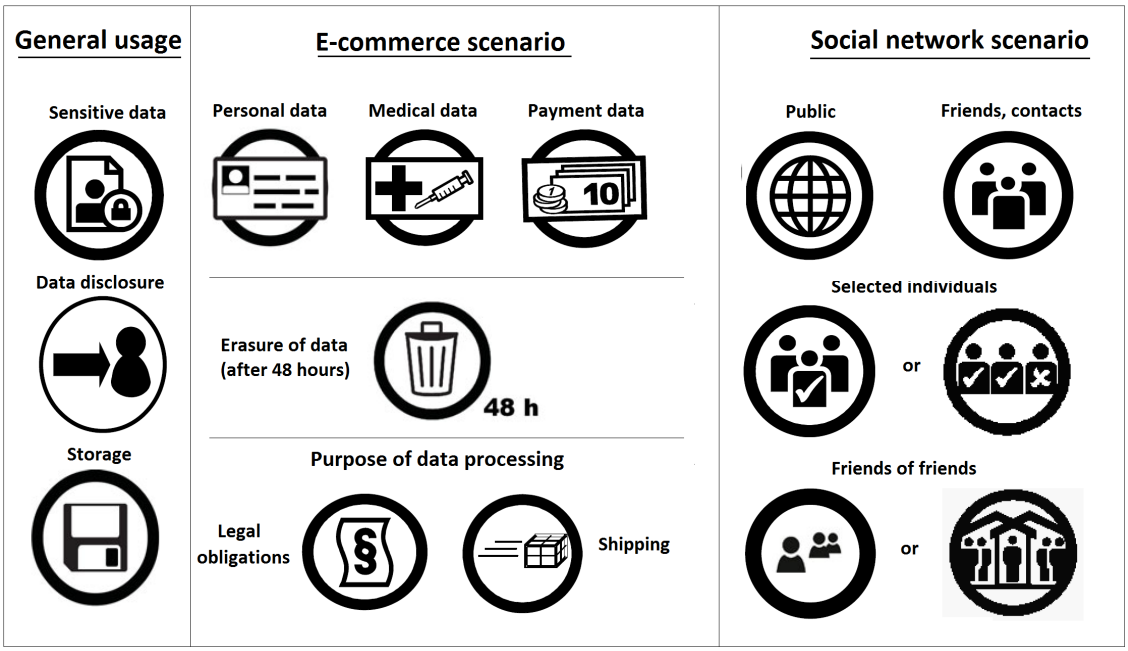


Figure 3.10: Icons developed for the FP7 project PrimeLife by Fischer-Hubner et al. [2010]

3.2.3 Multiple pictograms as a part of a larger system (4)

While the idea behind the privacy icons which is seen in figure 3.11 is similar to those approaches described above, Gomez et al. [2009] also included a proposal for matching the icons to the content of a policy. The process of analyzing textual policies was described as: "Policies were evaluated for Types of Data Collected, General Data Collection Practices, and Data Sharing Practices. Each policy received an evaluative code of YES, NO, or UNCLEAR for each category. YES and NO codes were only assigned if the distinction could clearly be made based on the wording of the site's privacy policy. UNCLEAR was given if the given information was not specified or was to too nuanced or vague to be determined"[Gomez et al., 2009].

Another suggested approach which takes the underlying system in account is a proposal for "Privacy Commons" by Bickerstaff [2009]. The "Privacy Commons" is based on the structure of the "Creative Commons" which was mentioned earlier in this section. The three-layer structure for the Creative Commons copyright licenses is demonstrated in figure 3.12.

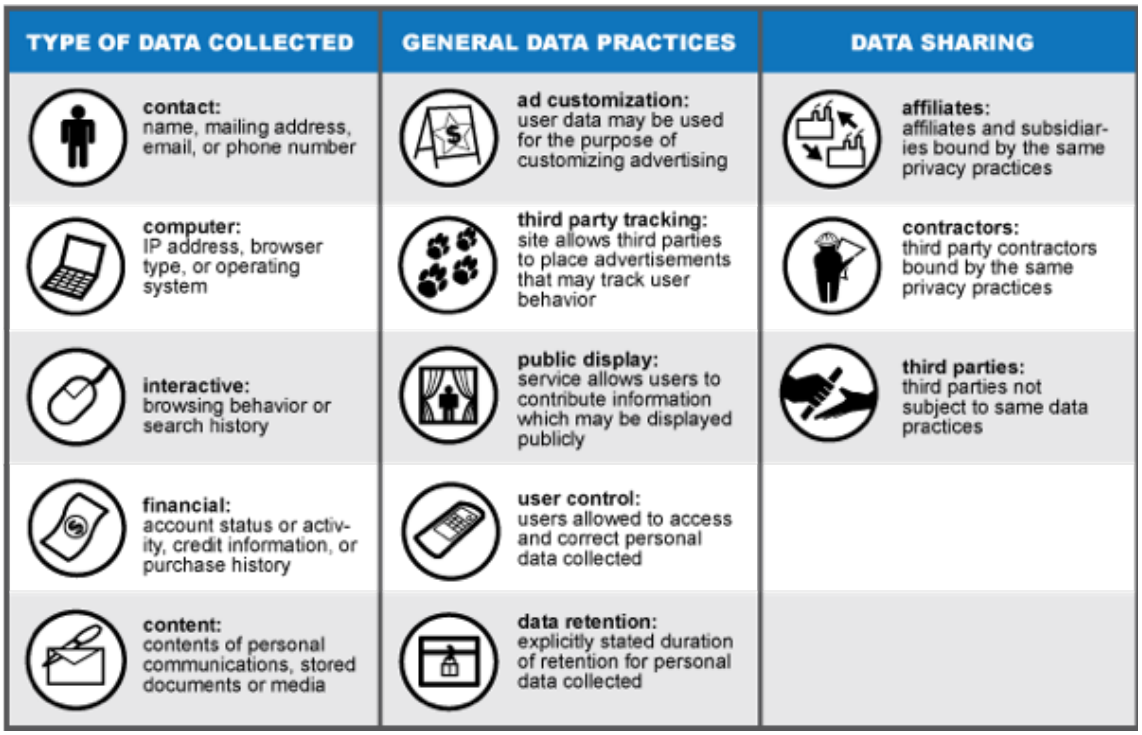


Figure 3.11: Proposal for privacy pictograms by Gomez et al.[2009] (Part of the KnowPrivacy project)

According to Creative-Commons [2011], each license begins as a traditional legal tool, in the form of a full legal text. The next layer consists of a human readable part, called the Commons Deed. The Commons Deed is a handy reference for licensors and licensees, summarizing and expressing some of the most important terms and conditions. The final layer, the machine readable layer, is written in a format which enables search engines or other software to understand the license.

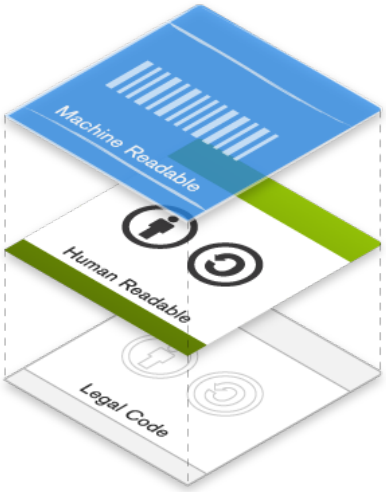


Figure 3.12: The three-layer structure for the Creative Commons copyright licences. Copyright [Creative-Commons, 2011]

The basic concept of the proposed "Privacy Commons" is simply a privacy version with the full policy text at the bottom layer, a human readable layer consisting of icon-sets as described above, and a machine readable version on top to be used with search engines and other software. Bickerstaff [2009] argues that such a system must be simple to use from users perspective, should lead to a general acceptance among websites and needs a reasonable level of granularity.

3.2.4 Related research and evaluation of pictograms

Related research on Privacy Seals and the Privacy Bird was discussed in the in-depth study[Lillebo, 2010], and the major findings were that while Privacy Seals has seen some adoption, they suffer from lack of recognition among users, and a general lack of monitoring and enforcement mechanisms from the seal-providers. Being based on the P3P language, the Privacy Bird suffers from little adoption, which in turn results in a majority of sites being classified as unsecure (yellow icon), even if the websites privacy practices matches the user's preferences.

The research regarding pictograms are still in the early stages of development, and the various solutions mentioned in the preceding sections has gained little outreach [Holtz et al., 2011]. Most research on privacy icons have been related to the PrimeLife project, where one test with about 20 students from Sweden and China was conducted at Karlstad University (KAU) in Sweden. Another evaluation of the PrimeLife icons an online survey, interviewed a total of 70 participants from ten different countries [Holtz et al., 2011].

With some exceptions, the findings from these evaluations were indicating that the development of privacy icons have been well worth the effort. An important finding was that cultural differences might result in different different interpretation and understanding of specific icons [Holtz et al., 2011]. It was also found that large icon sets should be reduced to that extent and complexity that interested users will be able to understand and to deal with.

Through the presentation of the results from these evaluations, Holtz et al. [2011] presents the following criteria list for designing icons:

- Privacy icons should allow for quick comprehension by all possible groups of users regardless of their cultural or social background.
- The different constructions of privacy and individual freedom should not hamper grasping the meaning of icons.
- Social factors like education and age must not restrict their user-friendliness.
- It should be possible to understand the icons within different legal frameworks.
- Icons should be understandable by an international target group and not be limited to a country or region.
- Icons should have a circular shape, as a triangular shape is widely associated with warning symbols.
- Icons should just be colored black and white, because colors like red, orange or yellow often have a warning function. This is also relevant due to some users being color-blind.

- Icons should be designed in a fashion that enables a thorough depiction of information.

Holtz et al. [2011] concludes that special attention should be given to the possibilities of combining the icon approach with machine-readable service policies and user preferences, as suggested by the "Privacy Commons" approach in section 3.2.3.

3.3 Standardized format

The final category of solutions for presenting privacy policies is standardized formats. With standardized format we mean a structure with pre-defined sections for content, i.e. every policy looks the same - only the content differs.

3.3.1 The Kleimann Communication Group prototype

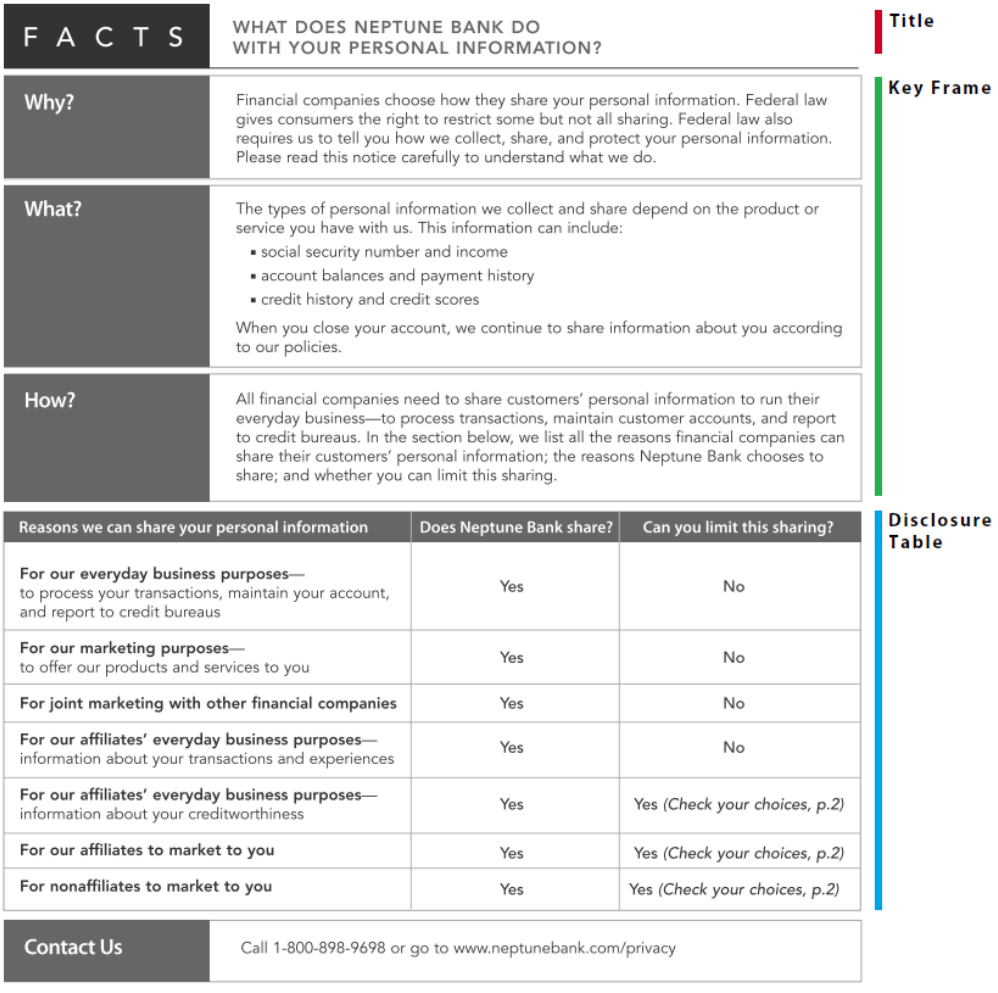
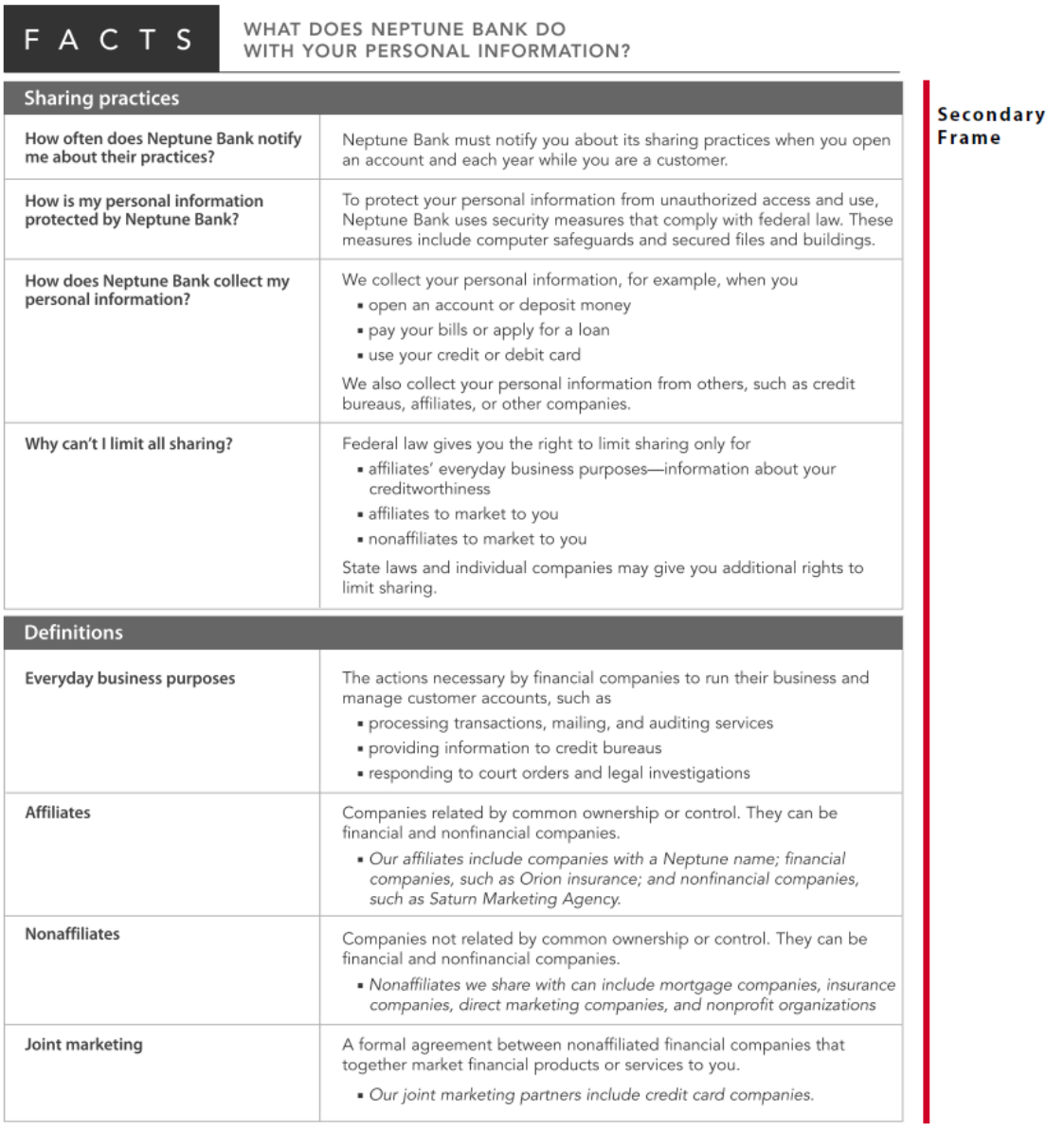


Figure 3.13: Page 1 of the Kleimann Communication Gr. prototype [KCG, 2006]

One of the first propiosals for a standardized format, which has been used as a basis for more recent solutions, is the prototype seen in figure 3.13, 3.14 and 3.15 which was developed through a study on paper-based financial notices by the Kleimann Communication Group [KCG, 2006].



FACTS WHAT DOES NEPTUNE BANK DO WITH YOUR PERSONAL INFORMATION?

Sharing practices	
How often does Neptune Bank notify me about their practices?	Neptune Bank must notify you about its sharing practices when you open an account and each year while you are a customer.
How is my personal information protected by Neptune Bank?	To protect your personal information from unauthorized access and use, Neptune Bank uses security measures that comply with federal law. These measures include computer safeguards and secured files and buildings.
How does Neptune Bank 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 others, such as credit bureaus, affiliates, or other companies.
Why can't I limit all sharing?	Federal law gives you the right to limit sharing only for ▪ affiliates' everyday business purposes—information about your creditworthiness ▪ affiliates to market to you ▪ nonaffiliates to market to you State laws and individual companies may give you additional rights to limit sharing.

Definitions	
Everyday business purposes	The actions necessary by financial companies to run their business and manage customer accounts, such as ▪ processing transactions, mailing, and auditing services ▪ providing information to credit bureaus ▪ responding to court orders and legal investigations
Affiliates	Companies related by common ownership or control. They can be financial and nonfinancial companies. ▪ *Our affiliates include companies with a Neptune name; financial companies, such as Orion insurance; and nonfinancial companies, such as Saturn Marketing Agency.*
Nonaffiliates	Companies not related by common ownership or control. They can be financial and nonfinancial companies. ▪ *Nonaffiliates we share with can include mortgage companies, insurance companies, direct marketing companies, and nonprofit organizations*
Joint marketing	A formal agreement between nonaffiliated financial companies that together market financial products or services to you. ▪ *Our joint marketing partners include credit card companies.*

Figure 3.14: Page 2 of the Kleimann Communication Group prototype [KCG, 2006]

The Kleimann Communication Group found that a standardized format helped consumers recognize the policy and understand the information in it [KCG, 2006]. By using a standard table design and subheadings in combination with a simplified content, readability was dramatically improved compared to full-length text notices. It was found that consumers recognized the policy and the information in it, and the proposed design also allowed for easier comparison of sharing practices across finan-

cial institutions, as the users knew where to look for differences.

Th first page of the prototype can be seen in figure 3.15, and holds a title, a key frame and a disclosure table, while a secondary frame is placed on page 2 (figure 3.14). The prototype also has a third page (figure 3.15), which consists of contact informatio,n and an opt-out form.

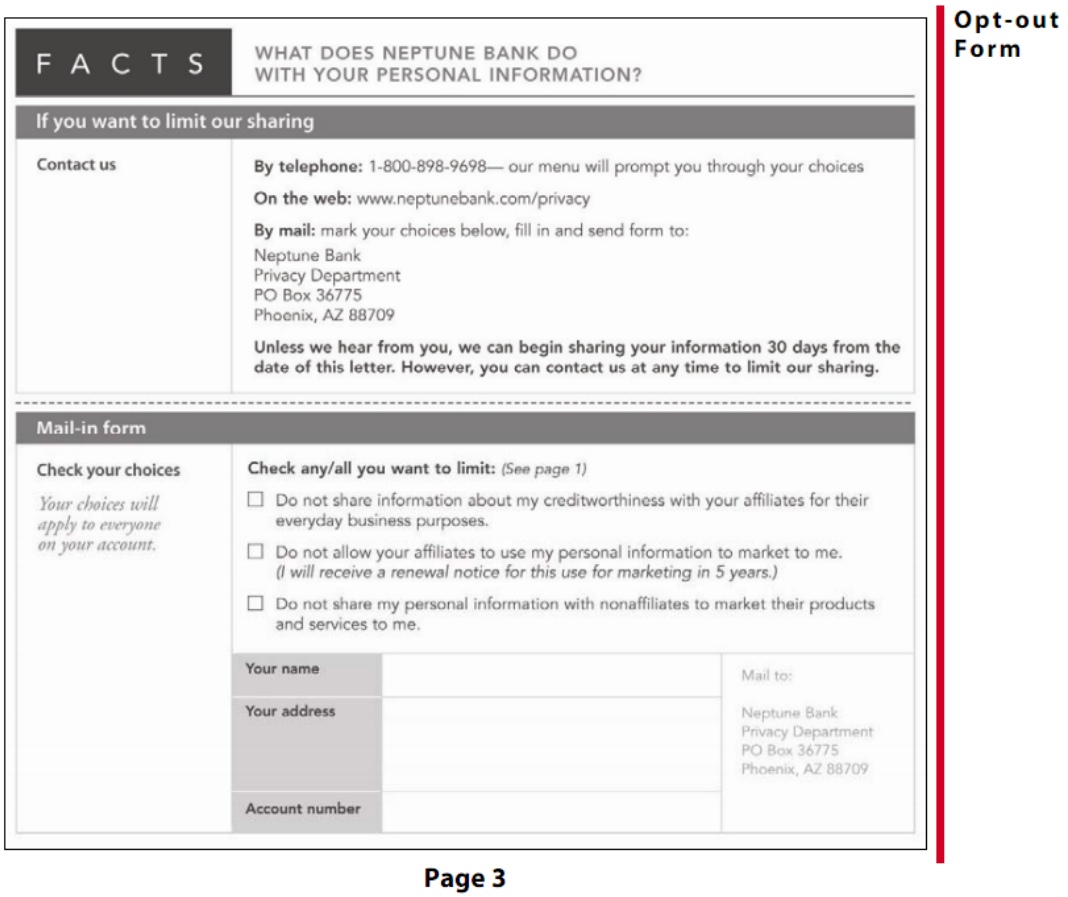


Figure 3.15: Page 3 of the Kleimann Communication Group prototype [KCG, 2006]

The heart of the standardized format and the key component in ensuring comparability is the disclosure table. By listing seven basic reasons for why personal information can be shared, answered by a simple yes or no, consumers can easily compare sharing practices across institutions[KCG, 2006]. A second column also identifies whether the consumer can opt-out of a particular sharing. It was found that this simple yes/no design increased participants' understanding of sharing practices in contrast a more textual approach where the participants became overwhelmed of all the text[KCG, 2006].

The Kleimann group also found that having a good and simple design was important, and underlined the importance of finding a balance between using as few words as possible and providing enough information so consumers could understand. Being

a prototype for financial privacy notices, the Kleimann group also had to comply with a series of requirements regarding content of the policy and solved this by having the key context information on the front page, while putting more detailed information not essential to the consumer on the second page. This way, the consumer can understand the basic concepts of financial sharing practices while not being overwhelmed by too much information [KCG, 2006].

Although intended for paper-based financial policy notices, the findings from the Kleimann report are also relevant for the design of non-financial internet policies, and have been used in several subsequent proposals for a standardized format.

3.3.2 P3P Expandable grid

Before presenting the P3P expandable grid, we make a short introduction to the P3P language. P3P is a platform designed by W3C to enable websites to encode their data-collection and data-use practices in a machine-readable format, known as P3P policies [Antón et al., 2007]. The P3P policy itself is written in XML language, where each policy contains information regarding the legal entity responsible for the privacy policy, whether the user has access to the stored data. In addition the policy has at least one statement on what data is collected, how it is used with whom it will be shared and for how long it will be stored [Wang and Kobsa, 2008]. For further information regarding P3P we refer to our in-depth study [Lillebo, 2010] and Antón et al. [2007].

One of the early attempts on creating a table format for online policies was the P3P expandable grid system, proposed by Reeder et al. in 2008. Based on the concept of the expandable grid, a visualization technique for displaying a policy, the P3P Expandable Grid automatically reads and displays policies written in the P3P language. The policy is presented to the user in a 2-dimensional matrix similar to the disclosure table proposed by Kleimann, but rather than a simple yes or no, each square is colored according to the P3P outcome.

Due to the complex nature of P3P policies, the resulting prototype as seen in figure 3.16 did however become too complicated and did not perform well in tests ([Reeder et al., 2008], [Kelley, 2009]). It was found that it was generally too visually busy and that users did not understand where to look for information in the P3P data hierarchy system. Confusing icons and unfamiliar terminology also caused problems and some users did not understand that the grid was interactive and that it could be clicked to be expanded [Reeder et al., 2008].

Kelley [2009] pointed out five other major problems through their study of the P3P Expandable grid. Among them were unclear labels, unfamiliar terminology, too many symbols and multiple statements that might be related to the same types of information. They also found that the hide unused information button in the top right corner only condensed rows, not columns.

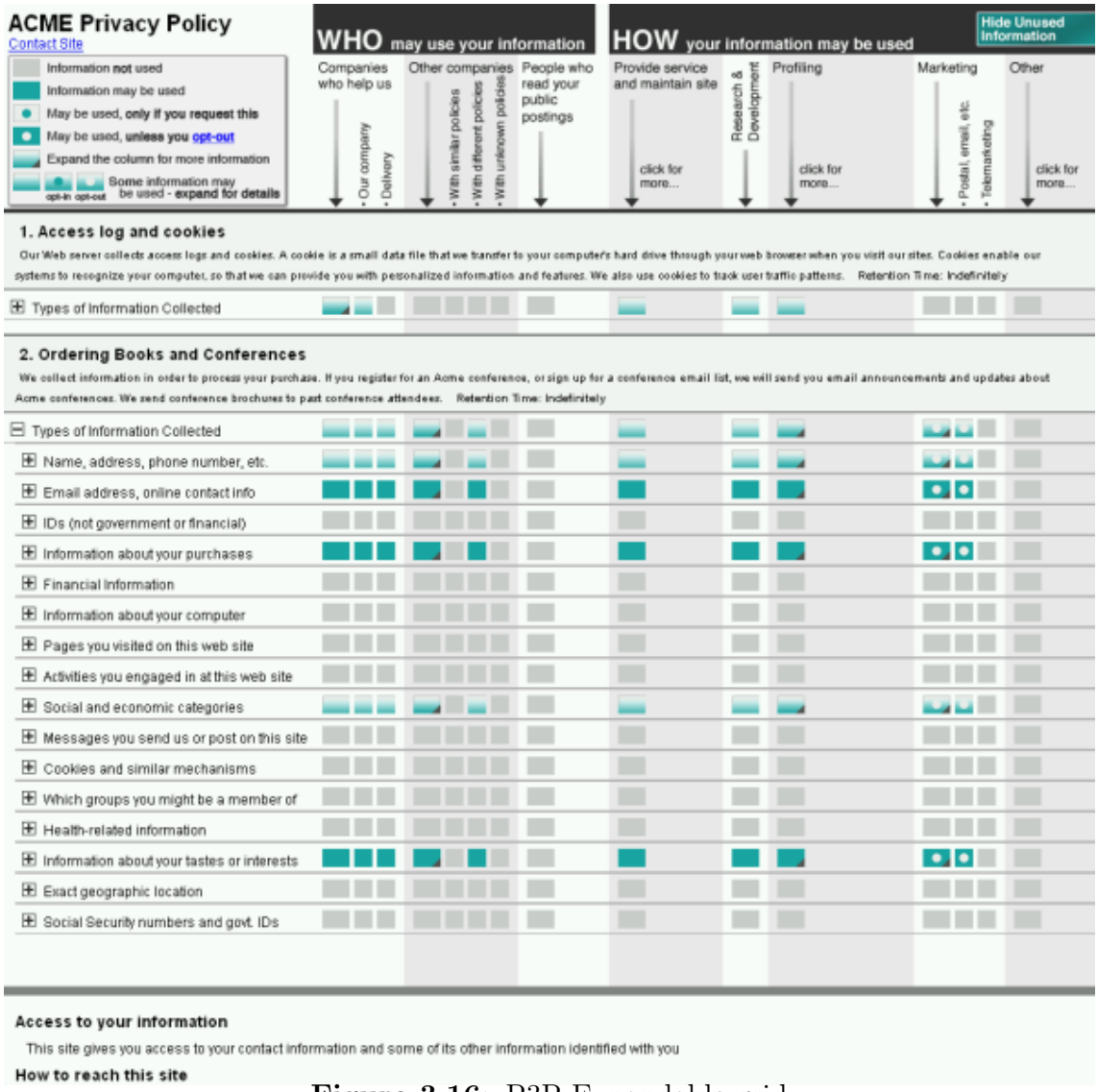


Figure 3.16: P3P Expandable grid

Reeder et al. [2008] suggested several improvements for handling the usability problems of the prototype. Among them was including a visual starting point such as a search bar, using shorter labels and providing plenty of explanation for icons. They also suggested including a more simplified policy version, summarizing the data practices only of the highest concerns to the consumers. Finally they suggested either emphasizing or removing the interactivity, as several users did not understand they could interact with the grid.

3.3.3 Nutrition Label

Based on the findings from the study on P3P expandable grids and the findings from the Kleimann report, Kelley et al. [2009] initiated the work on presenting the privacy policy in the shape of a Nutrition Label.

Their work was based on the nutrition facts label as seen in the figure to the right, which has become iconic in the US after being mandated by the Nutrition Labeling and Education Act of 1990 [Drichoutis et al., 2006]. Research on the food Nutrition Label found that by reducing the amount of information displayed and equalizing labels across products, the Nutrition Labels helped consumers make more informed decisions through allowing for comparison between small sets of items ([Seymore et al., 2004], [Drichoutis et al., 2006]).

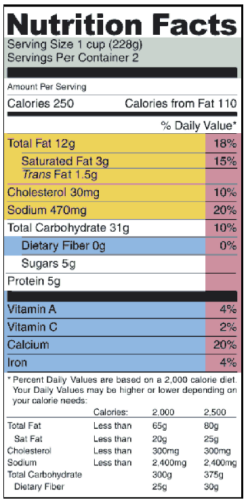


Figure 3.17: The Food and Drug Administrations Nutrition Facts panel as regulated by the NLEA. Source: www.fda.gov

While several consumers stated they would prefer more information than given in the simple notices, studies found that including more information would actually not be beneficial ([Drichoutis et al., 2006], [Beard et al., 2007]). Other studies found that the labels were most useful for the consumers already interested in the nutrition information ([Drichoutis et al., 2006], [Beard et al., 2007], [J.S. et al., 2009]).

Based on these findings and the mentioned problems with the P3P expandable grid, Kelley et al. [2009] abstracted the following three general design principles:

1. A box around the label identifies the boundaries of the information and defines the areas that are regulated or should be trusted.
2. Bold rules to separate information and clearly designated sections grouped by similarity
3. Clear and boldfaced title communicates the content and purpose of the label, assisting in recognition.

Based on these principles and a intention of making policies more enjoyable and easier to read, they proposed a prototype for a Simplified Label as seen to the left in figure 3.18. Together with the implementation of the three main principles listed above, the biggest change from the P3P Expandable Grid was the reduction of complexity. The Simplified Label also reintroduced quantifiable fields by including binary Yes/No statements, which was found to be readily understood by focus groups in the [KCG, 2006] report.

Through an iterative design process, Kelley et al. [2009] proposed a second version, called the Simplified Grid as seen to the right in figure 3.18. Given the simplicity of the Simplified Label, the authors felt it sacrificed too much detail and therefore reintroduced the grid structure, without adding too much complexity. The Simplified Grid was considered a "happy medium" between the best aspects of the P3P Expandable Grid and the initial label[Kelley et al., 2009]. One of the most significant changes from the P3P expandable grid was the reduction from 10 to 4 symbols to represent sharing practices.

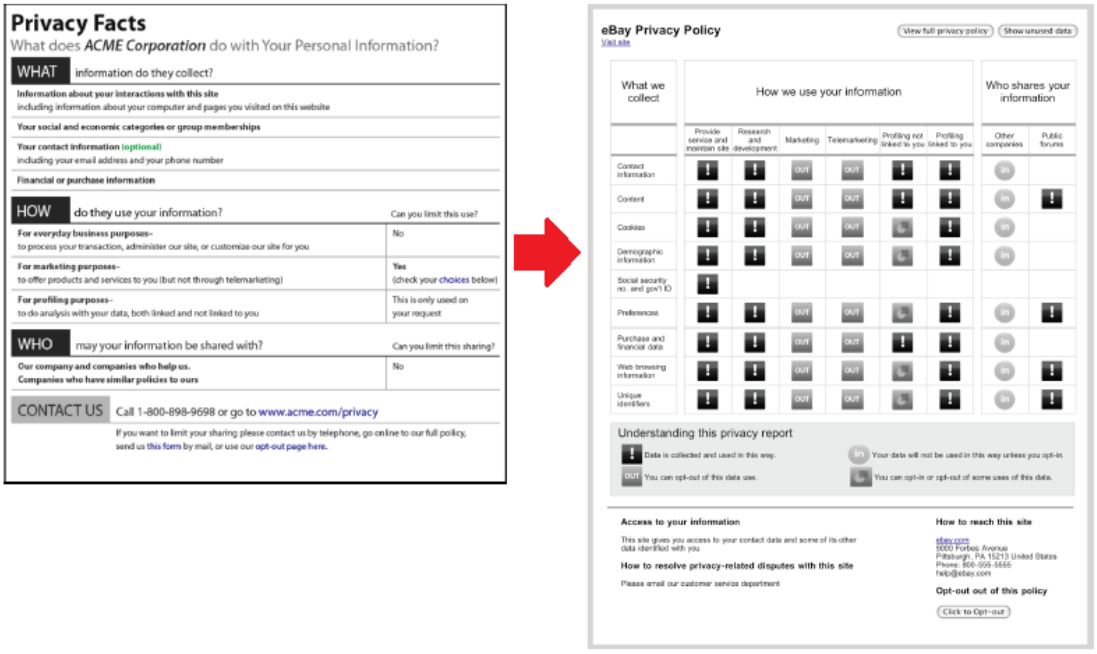


Figure 3.18: The Simplified Label (left) and the Simplified Grid (right)

The third and final version from the Kelley et al. [2009] report was the label seen in the figure 3.19. Among changes were better facilitating of comparison between policies, coloring of the various symbols and changes to the table structure. Kelley et al. [2009] highlighted three important design factors for the development of the label: the ability to find information, the understanding that there are differences between privacy policies and control over one's information, and the simple time-based costs of reading privacy policies.

Through a laboratory study where the Nutrition Label was evaluated against a natural language policy, the proposed privacy label allowed participants to find information more quickly and accurately, and provided a more enjoyable in-

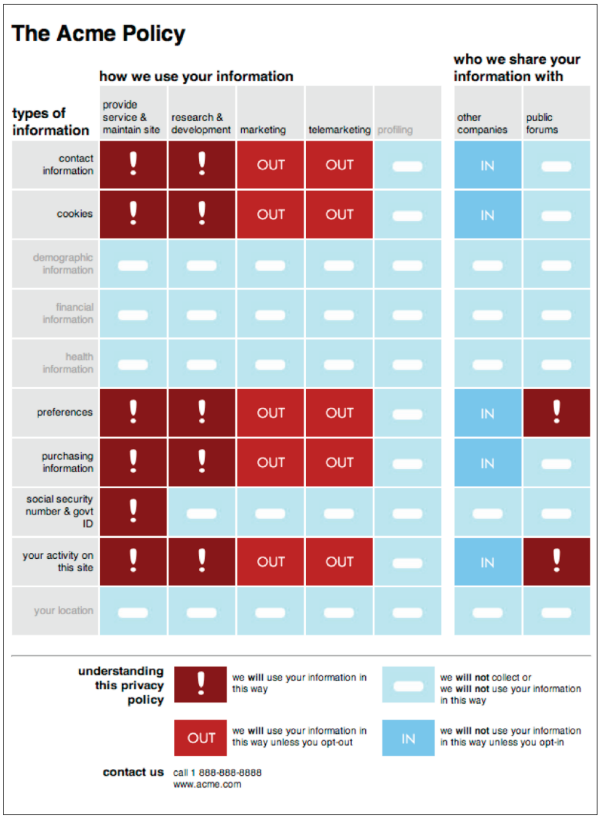


Figure 3.19: The Nutrition Label from Kelley et al. [2009]

formation seeking experience than the textual policy.

Further development on the Nutrition Label was done through a later study by Kelley et al. [2010] which compared the Nutrition Label to several other formats, including the standardized short text and full text policies. As seen in figure 3.20 (left), minor design changes were done to the symbols in the label, as well as removal of the "research and development" column. A shortened version of the Nutrition Label, titled "Standardized Short Table" was also created for the comparison (see figure 3.20 (right).

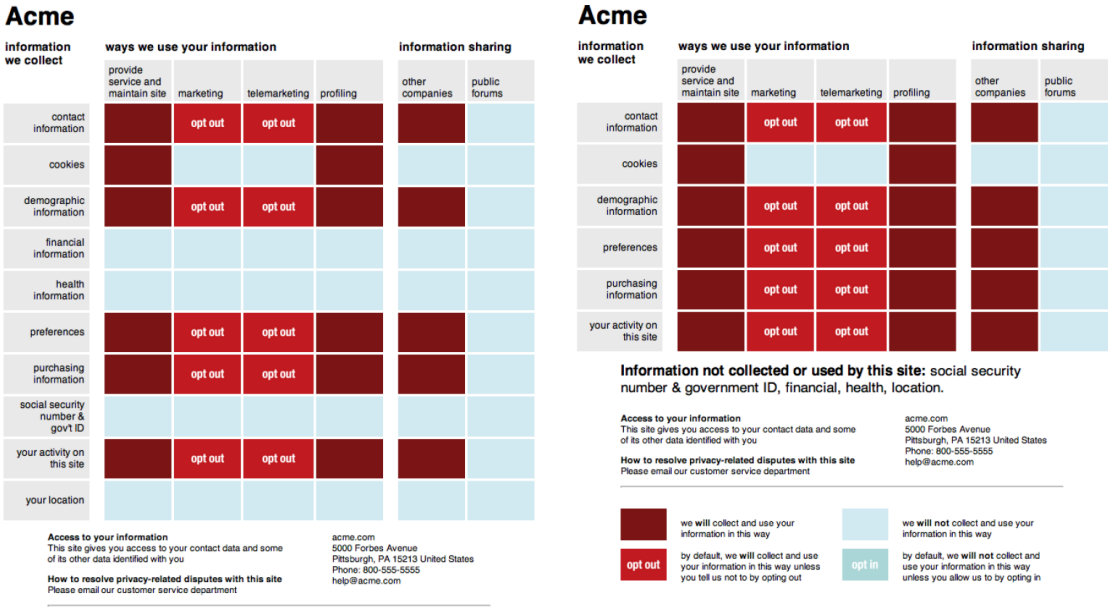


Figure 3.20: The Nutrition Label (left) and the shortened version (right) used in the evaluation by Kelley et al. [2010]

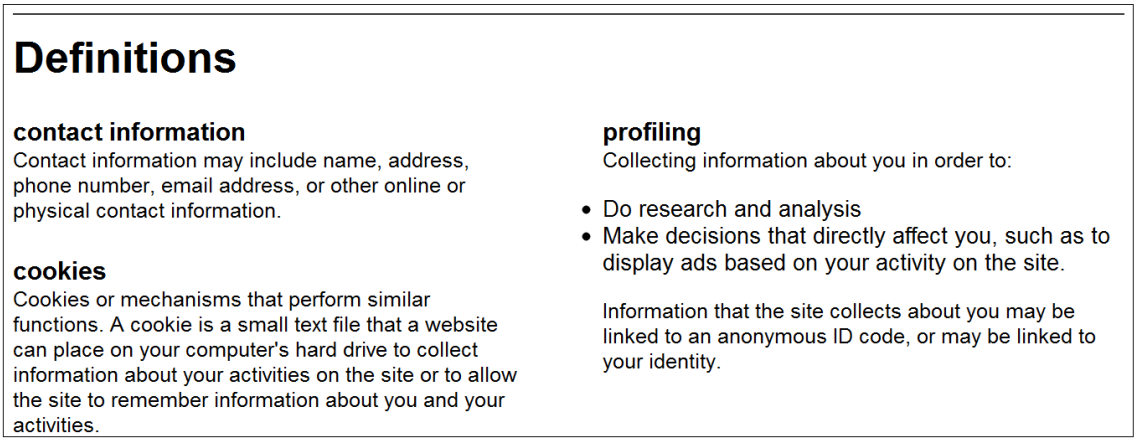


Figure 3.21: A subset of the definitions section for the Nutrition Label

Through an online user study of 764 participants, Kelley et al. [2010] found that standardized privacy policy presentations can have significant positive effects on accuracy and speed of information finding and on reader enjoyment of privacy policies.

Kelley et al. [2010] suggested that future work should concentrate on facilitating comparisons and framing the policy with contextual information.

Through a recent presentation, the authors of the Nutrition Label suggestion how such contextual information could be implemented with the remaining structure [Kelley et al., 2011]. A sample of the definition page which has doubled the size of the Nutrition Label, is seen in 3.21.

The Nutrition Label has also been implemented as a part of the Privacy Finder search engine which was described in section 3.2.1. By clicking the "Privacy Report" link next to the search result, a full sized Nutrition Label representing the privacy practices of that site is displayed, as seen in figure 3.22. This policy is automatically created based on the P3P policy, and thus only supports websites with provides such a P3P policy [PrivacyFinder, 2011].

As seen in the same figure, the authors has also added a section presenting general information such as company name and address, links to opt-out pages and the textual policy, as well as information regarding data retention and access to the information. Privacy Seal (section 3.2) information has also been incorporated.

According to Kelley [2011], future work regarding the work on a standardized label will be to conduct another study where the various components of the Nutrition Label is removed to show how much each of the specific design patterns impacts the overall usability metrics. For instance, one version will remove the standardized terms we used in the label, and replace them with the different terms the actual policies began with. This work is a part of the thesis work of Patrick Gage Kelley [Kelley, 2011], and is due to be conducted during the summer of 2011.

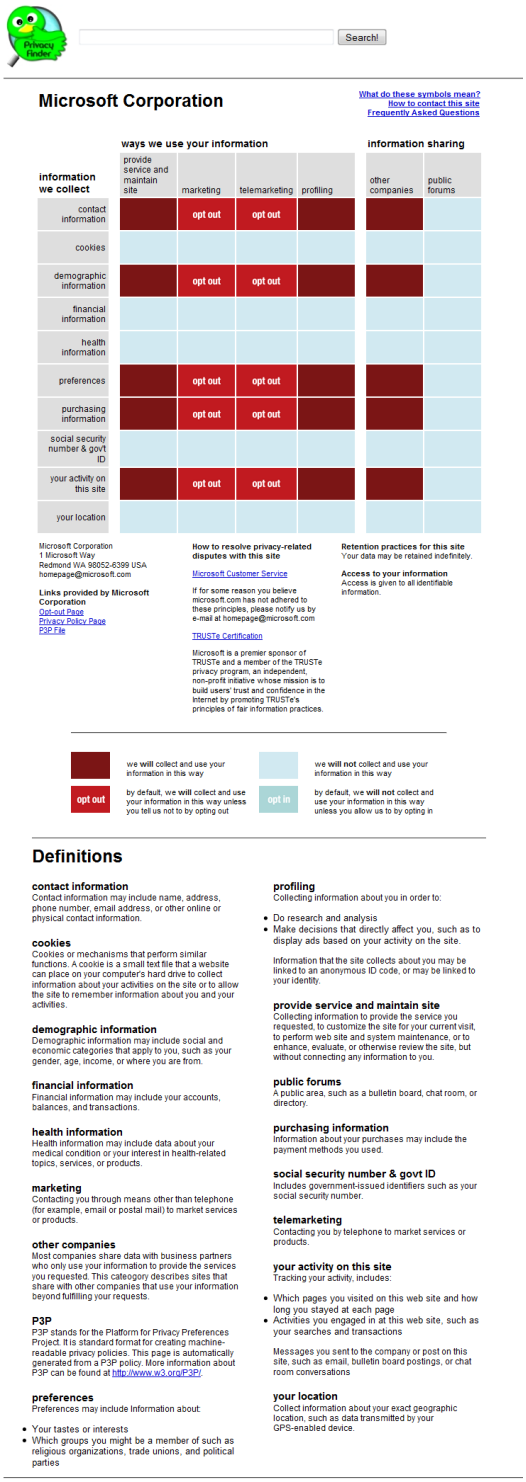


Figure 3.22: The Microsoft Corp. policy as retrieved from the privacy finder search engine

CHAPTER 4

Design

The original intention of the pre-study chapter was to elaborate on current solutions for presenting privacy policies, and based on this evaluate these against our list of criteria, in order to choose a solution to base further work on. The outcome of the pre-study chapter was however somewhat unexpected, and did not result in a series of candidates which we could compare. In general, we found that while layered notices has seen some adoption, pictograms are still in the early phases of research, and with the exception of the Nutrition Label which has developed to a implemented solution, little research has been conducted on standardized solutions.

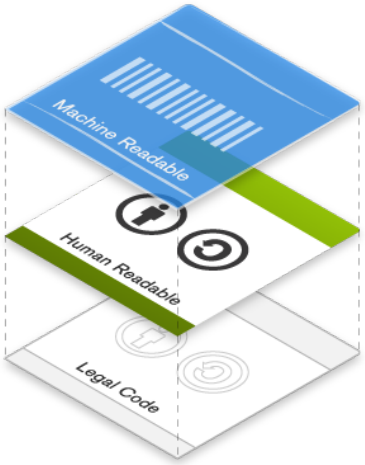


Figure 4.1: The three-layer structure for the Creative Commons copyright licences.

What we did see from the pre-study was that all three categories of solutions followed a similar pattern, which closely resembled the three-layer structure which was proposed by Creative Commons. The concept of the three-layer structure was elaborated in section 3.2.3 on page 22, and consisted of a machine and a human readable layer, in addition to the legal code as seen in figure 4.1.

To better present this theory, we have summarized the findings from Chapter 3 in figure 4.2, where each area is represented by a relevant solution: the TRUSTe layered notice, the KnowPrivacy pictograms with the underlying conversion system, and finally the standardized Nutrition Label. Based on the theory by Creative Commons we can separate the three layers, and argue that future research on privacy notices should focus on developing a single solution for each layer. In other words, one solution is needed to convert a full text policy to machine readable code, a second solution is needed to narrow the most important parts of the policy, and a third solution is needed to visually present this information to the user.

[Holtz et al., 2011] also supported this strategy for future work on the icon approach, but we are uncertain where a preference based system could fit in the layers. Either

way it would require a more advanced structure behind the human-readble layer which is out of the scope for this paper.

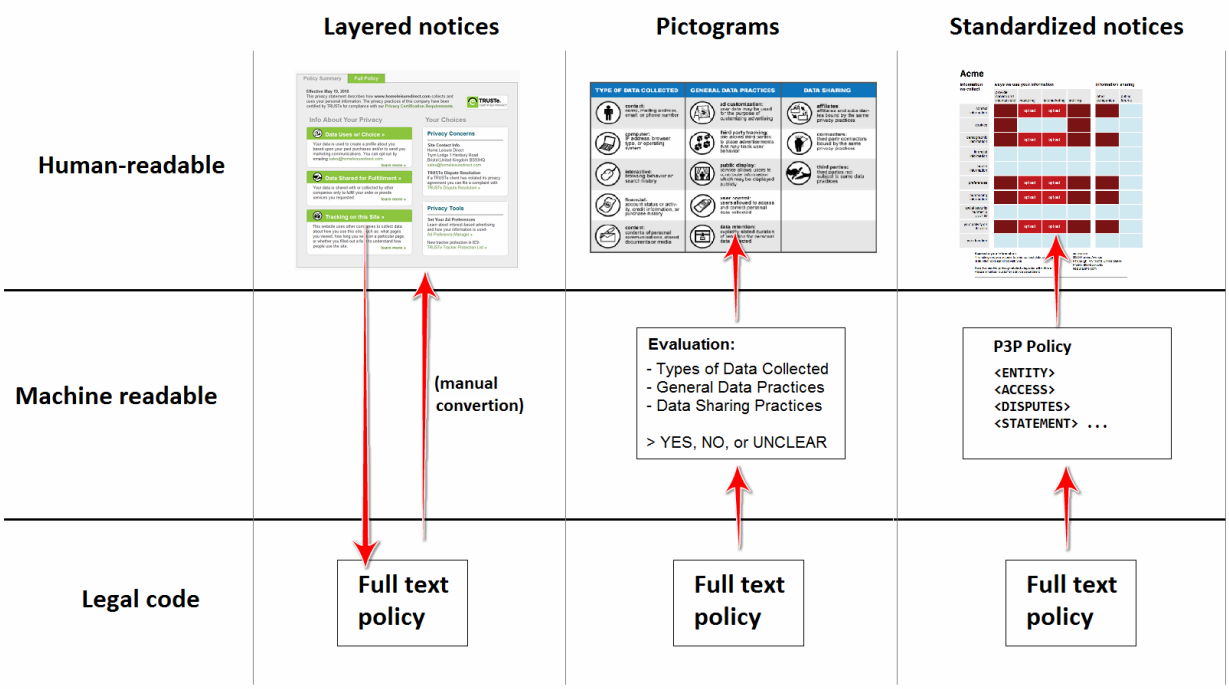


Figure 4.2: Overview of the findings from Chapter 2, organized in the Creative Commons three-layer structure [Creative-Commons, 2011]

As the scope of this thesis was on the visual presentation of the policy, the human-readable layer, we also argue that the further work on such a solution in this report is relevant for the research communities of both layered notices, pictograms and standardized notices.

Tracing back to the problem definition for the thesis, we outlined the following criteria for a new solution:

1. Make the general policy content more understandable,
2. Make the policy easier and more enjoyable to read
3. Highlight the most important privacy practices, while preserving details and nuances
4. Enable for easier comparison of privacy practices across sites
5. Be useful for the COPE project as well as current and future Privacy Enhancing Technologies (PETs)

Based on this we developed the following list of questions to guide the evaluation process.

1. How easy is it to find the information the user is looking for?

2. How easy and user-friendly is the solution to use?
3. Does the solution require experience or can a novice quickly understand the format?
4. How well does the solution manage to cover the details of a textual policy?
5. How well does the solution allow for policy comparison?
6. Can the solution be used in current or future Privacy Enhancing Technologies (PETs)?

Our planned approach was to evaluate each solution based on these questions, but given the similarities between the visual solutions, and with little or no research findings to base our evaluation on, we decided to base our further work on the Nutrition Label as a result of the following discussion.

Primarily, we observed that while little progress has been made on layered notices since their introduction in the mid-2000s, the research on pictograms is still in the early phases. The Nutrition Label on the other hand, has through several research projects reached a prototype implementation stage and further research and development is also planned. Kelley et al. [2010] also found that the standardized solution (the Nutrition Label) performed significantly better compared to a layered solution in terms of accuracy for single policy information finding and comparison questions. The same study also found an indication for that the standardized label was more enjoyable to use than the layered version.

We also believe that a good solution should find a balance between simple icons and detailed text, and that the Nutrition Label represents such a middle-way. Given this middle way between text and icons we also believe the Nutrition Label is best suited for use in current or future Privacy Enhancing Technologies, and as it has already been implemented on top of the P3P language, a conversion to an alternative underlying system should be possible.

While we found the Nutrition Label to be the best suited solution to base further work on, the initial analysis also revealed several issues regarding the presentation of the label. Primarily, being novice users, we found it generally hard to comprehend the format. The first impression was that the label lacked a clear visual starting point for reading the policy, that there was a lot of colors and symbols to keep control of, and that the matrix system was hard to understand.

Secondly, we observed that while the Nutrition Label was a significant development from the P3P expandable grid, changes to the label since then has mainly been minor design changes. In their conclusion on the study of online policy formats, [Kelley et al., 2010] suggested that future work on standardized solutions should concentrate on framing the policy with contextual information, and provide better education to help consumers make better decisions. As a result of this, the latest version of the Nutrition Label as presented by the Privacy Finder has doubled its size with contextual information. We therefore question whether the best development of the Nutrition Label is to expand the label with additional information to solve evolving issues, or to rather focus on improving the main structure of the design

itself.

In fact, with one exception (the standardized short table in [Kelley et al., 2010]), the studies on the Nutrition Label have only measured its performance versus textual policies or similar textual alternatives. It is therefore hard to know whether the findings indicate that the Nutrition Label design is the best way of presenting a privacy policy, or whether the good results are due to it being the only approach of its kind.

Therefore the further work on a policy format was concentrated on developing a competitive design, based on usability literature and an in-depth analysis of the issues regarding the Nutrition Label.

The following sections describes the Nutrition Label analysis in details, starting with an overview of current real life Nutrition Label policies as retrieved from the Privacy Finder search engine (see section 3.2.1 on page 19). We then outline a list of design criteria for an alternative solution, before presenting the design process and the resulting solution entitled the "Privacy Table".

4.1 P3P / Nutrition Label enabled sites

To get an understanding of how the average real life nutrition policy looks like, we initiated a search for Nutrition Labels among the top 100 most visited US websites[1], by using the Privacy Finder search engine (described in section 3.2.1 on page 19).

The results, as presented in table 4.1, showed that only 22 out of 100 sites provided a P3P policy (and thus a Nutrition Label). Three of these used the same Microsoft Corp. policy (Microsoft.com, Live.com and Windows.com) and two used the same Yahoo! Inc. policy (Yahoo.com and Flickr.com). 15 additional sites gave positive result for a label, but upon clicking the link we received an Internal Server Error for two sites (marked −**) and a "The Privacy Policy for http://www.WEBSITE.com/ contains errors!" for 13 sites (marked −*). We excluded sub-domain policies such as http://learn.linkedin.com.

It is important to highlight that the low number of retrieved policies (22 of 100) is a direct consequence of the low implementation of P3P policies among the top 100 websites. There have been no indications from the creators of the Nutrition Label on whether they plan to base future implementations on P3P policies, or on an alternative language. Either way it is out of the scope of this paper, and we will only use these results to analyze the appearance of the 22 retrieved policies.

It is also worth mentioning that the purpose of the Privacy Finder search engine is to range websites depending on their privacy practices, and the functionality of displaying the policies in the Nutrition Label format is an additional feature. Given it is a prototype, the retrieved policies presented in the following sections might

[1] As published by http://www.quantcast.com/top-sites (results retrieved February 20th, 2011)

also be inaccurate or belong to a subdomain of the searched site (e.g. searching for "LinkedIn" returns a link to a policy for http://learn.linkedin.com, which claims to just collect cookies and site usage).

●	Has a P3P policy / Nutrition Label
–*	Errors in the P3P policy
–**	Internal server error while retrieving the label
–	Does not have a P3P policy / Nutrition Label

1.	google.com	–	34.	windows.com	●	67.	wikia.com	–
2.	facebook.com	–*	35.	myspace.com	–	68.	netflix.com	–*
3.	youtube.com	–	36.	photobucket	–	69.	hp.com	●
4.	yahoo.com	●	37.	match.com	●	70.	localpages.com	–
5.	twitter.com	–	38.	pandora.com	●	71.	washingtonpost	–
6.	msn.com	–*	39.	comcast.net	–	72.	mtv.com	–
7.	wikipedia.org	–	40.	tumblr.com	–	73.	yelp.com	–
8.	amazon.com	–*	41.	att.com	●	74.	drudgereport	–
9.	microsoft.com	●	42.	irs.gov	●	75.	wunderground	●
10.	live.com	●	43.	hulu.com	–	76.	reddit.com	–
11.	ebay.com	●	44.	imdb.com	–*	77.	intuit.com	–*
12.	answers.com	–	45.	cnn.com	–	78.	usps.com	●
13.	blogspot.com	–	46.	manta.com	–	79.	wellsfargo.com	–
14.	ask.com	–	47.	whitepages.com	–	80.	turbotax.com	–*
15.	ehow.com	–	48.	overstock.com	●	81.	twitpic.com	–
16.	blogger.com	–*	49.	webmd.com	●	82.	careerbuilder	●
17.	bing.com	–*	50.	apple.com	–	83.	hrblock.com	–
18.	aol.com	–	51.	flickr.com	●	84.	jcpenney.com	–
19.	wordpress.com	–	52.	cnet.com	●	85.	bestbuy.com	–
20.	craigslist.org	–	53.	chase.com	–	86.	simplyhired.com	–
21.	adobe.com	–	54.	metrolyrics.com	–	87.	city-data.com	●
22.	about.com	●	55.	bankofamerica	–	88.	squidoo.com	–
23.	weather.com	–*	56.	target.com	●	89.	nih.gov	–
24.	salesforce.com	–	57.	bbc.co.uk	–	90.	ups.com	–
25.	foxnews.com	–	58.	yellowpages	–	91.	norton.com	–
26.	huffingtonpost	–	59.	legacy.com	–	92.	metacafe.com	–
27.	linkedin.com	–	60.	dailymotion	–*	93.	chacha.com	–
28.	paypal.com	–	61.	nytimes.com	–	94.	weatherbug.com	–**
29.	mapquest.com	–*	62.	monster.com	–	95.	vimeo.com	–
30.	godaddy.com	–**	63.	comcast.com	–	96.	searchassist.com	–
31.	go.com	–*	64.	people.com	–	97.	evite.com	–
32.	reference.com	–	65.	tmz.com	–	98.	ancestry.com	●
33.	walmart.com	●	66.	bizrate.com	–	99.	time.com	–
						10(	coolmath-games	–

Table 4.1: Sites with Nutrition Label among the to 100 most visited US websites

While the findings does not say anything about how the future adoption of Nutriton Labels will be, they do provide an idea of how the average Nutrition Label based on real life P3P policies looks like. The retrieved Microsoft Corp. policy in its full length can be seen in figure 4.3.

All 19 retrieved Nutrition Labels for the 22 sites can be found in Appendix A. Given that all the contextual definitions, symbols legends and the textual information below the label were more or less the same for all policies, Appendix A only displays a screenshot of the symbol area itself.

Due to the amount of space required to display the Nutrition Label in its full length, retrieved policies in the remainder of this chapter will be presented in a compact format as demonstrated in the eBay Inc. policy in figure 4.4.

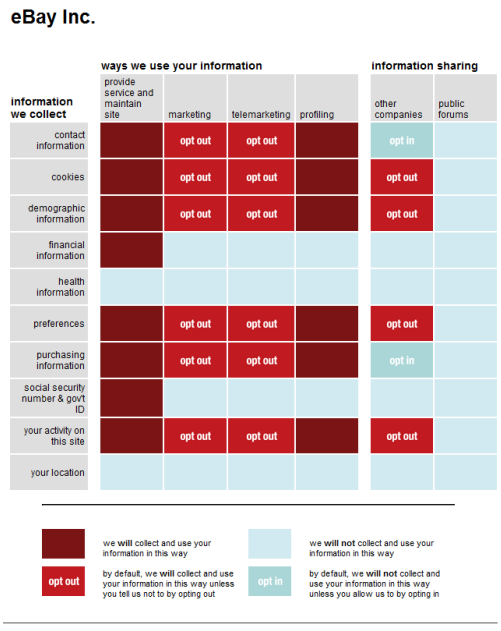


Figure 4.4: The compact eBay Inc. policy

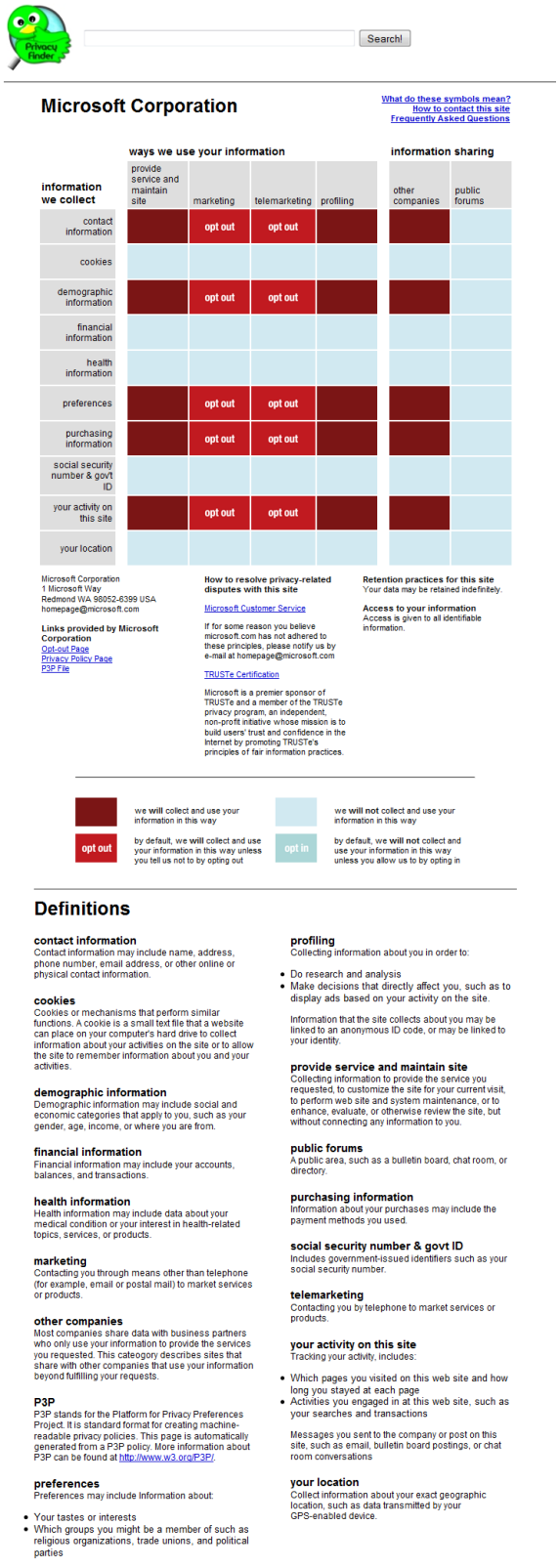


Figure 4.3: The Microsoft Corp. policy as retrieved from the privacy finder search engine

4.2 Analysis of issues regarding the NL

As mentioned in the introduction to this chapter, the main strategy for improving the Nutrition Label has been to add new parts to label structure. Kelley et al. [2010] highlighted that the future focus should concentrate on framing the policy with contextual information, and provide better education to help consumers make better decisions. While adding new parts acts as contextual aid to the user, it also clutters up the design, and the latest version of the Nutrition Label has grown to over triple its original size as seen in the figures on the previous page. This new size has also disabled the Nutrition Label from being printed on a regular A4 sheet of paper, and it also not optimized for print-outs across two pages.

Through the next sections we present findings which we believe supports that the future focus should be on re-evaluating the core design of the label itself, rather than solving evolving issues by constantly expanding the label.

4.2.1 Too complex and visually busy

Our first impression with the Nutrition Label was that we were presented with too much information at once, and that it was hard to understand where and how one should start reading the notice. We felt that there was a lack of starting point in the notice, and that the learning curve to grasp the different categories and symbols was too steep.

Consider the eBay policy in figure 4.5. We have identified 5 zones in the policy, sorted by the area size:

1. The symbol matrix (red)
2. The additional information (white)
3. The symbol legends (blue)
4. The row and column titles (yellow)
5. The row and column headers (green)

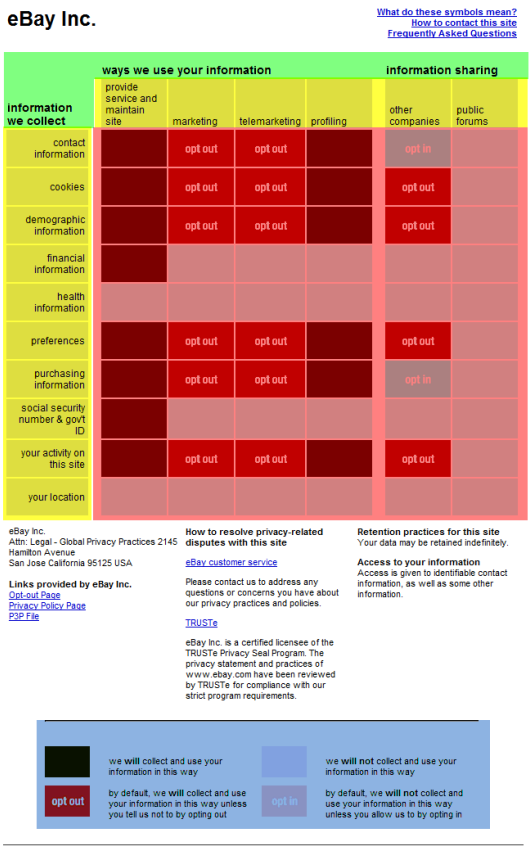


Figure 4.5: The eBay Inc. policy retrieved from the Privacy Finder search engine

The current implementation of the Nutrition Label has a strong focus on the symbol matrix which is very colorful and eye-catching, while the column and row

titles have been made less significant. This might be advantageous for the experienced user who is familiar with the structure, but for a first time user we think the label would be more readable with a more dominant area for the titles and headers (green and yellow area), and a clear starting point for reading the label.

We also feel there is too much information to keep control of while reading the notice. Consider the following user story:

As a prospective user of Acme Inc's services, John would like to see whether they will collect his purchasing records, and if so, for which purposes it will be used

For this task, the user would need to:

1. Locate "information we collect",
2. Look down to find "purchasing information",
3. Look right for the first row element
4. Look up to first read the column title, followed by the column header
5. Scroll to the bottom of the page to find the symbol legend.
6. Repeat steps 4 and 5 for the remaining row elements

The main problem is that one has to relate to two axis with corresponding row and column headers and titles, while at the same time keeping control of the legends for the symbols. In the latest version of the Nutrition Label, it also necesarry to scroll past the additional information field (white zone) to find the symbol legend. An experienced user would obviously be familiar with the symbols and the structure, but for an untrained user we believe the number of axes which one needs to keep control of makes the label challending to read.

It has also found that the use of white space can improve comprehension and make content more , and according to Lin [2004] the use of white space between paragraphs can increase comprehension by almost 20%. White space de-clutters a page by giving the items room to breathe, and together with a good layout it can also influence user satisfaction and experience [Chaparro et al., 2005]. As seen in the eBay policy in figure 4.5, there is minimal use of white space in the Nutrition Label. While the light blue icons which indicate that information is not collected or used in any way might provide some space, we believe that smaller symbols and more white space between the cells would improve readability.

We therefore suggest the following changes to improve the usability of the label:

- Providing a clear starting point for reading the label
- Making the row and column titles and headers more dominant
- Making it clearer which titles belongs to which columns and rows
- Decreasing the size and the amount of symbols the matrix
- Adding more white space
- Reducing the number of axis in the label

4.2.2 Unnecessary "third dimension"

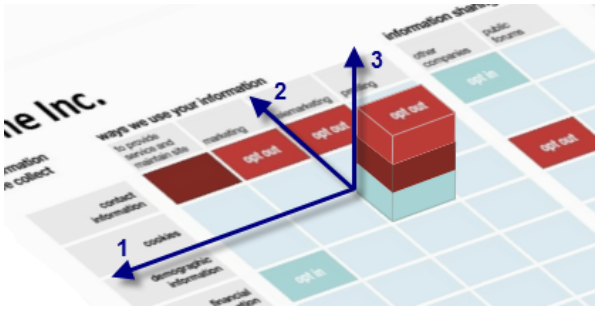


Figure 4.6: The three axes in the Nutrition Label

Given the two-dimensional structure of the Nutrition Label, each row in the matrix corresponds to a data type, and each column to a way of using or sharing this information. As illustrated in figure 4.6, the Nutrition Label also has a third axis to represent how each data type will be used under each category. We will hereafter refer to this as the Nutrition Labels "third dimension".

While using the privacyfinder.org search engine to retrieve Nutrition Labels, we noticed that real life labels based on P3P policies differed from the fictive ones commonly used in previous research. While the fictive Acme policy in figure 4.7 used in the study by Kelley et al. [2010] utilizes the potential of this third dimension to display different sharing options for different information types within a column (e.g. contact information is collected and used for marketing by default, while location data is not used for this purpose.), the real life Microsoft Corp. policy in figure 5.8 has similar sharing practices for all data types in each column.

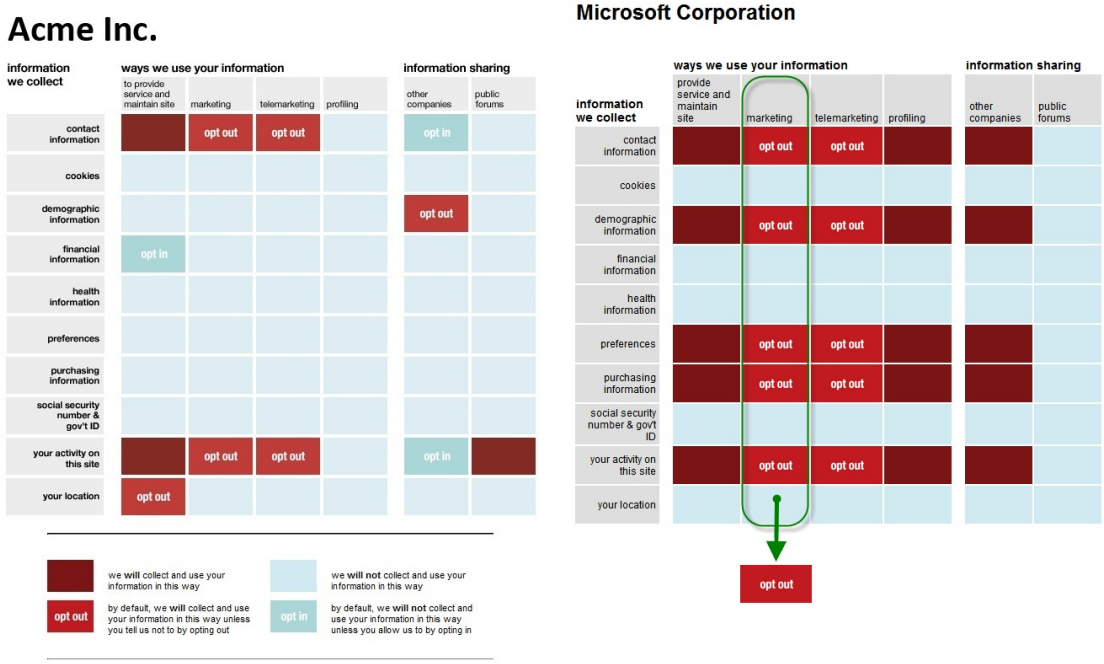


Figure 4.7: The fictive Acme policy (left), The Microsoft Corp. policy (right)

Further examples can be seen in the policies in figure 4.8, where all policies have in common that every column could be summarized by a single symbol, rather than individual ones for each information type. Out of the 22 Nutrition Labels we retrieved in our search, 19 labels had this in common (all policies can be seen in Appendix A).

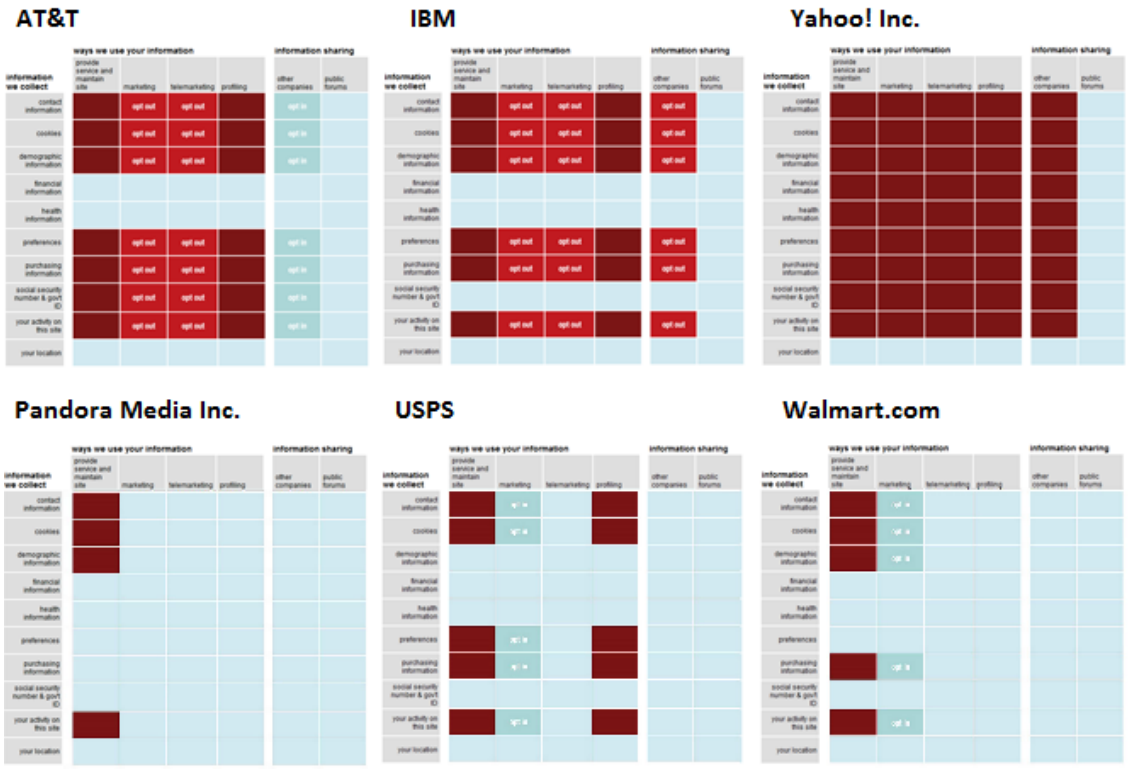


Figure 4.8

The only three policies which utilized the "third dimension" functionality can be seen in figure 4.9. In addition to the AT&T's policy (figure 4.8), the eBay Inc. policy(figure 4.9) was also the only policy to make use of all three sharing symbols.

Figure 4.9

It is also worth mentioning that the average Nutrition Label policy in our search had very few elements, with minimal use of opt-in and opt-out symbols. Several policies had almost no symbols, as exampled by the The Match.com and and Overstock.com policies in figure 4.10.

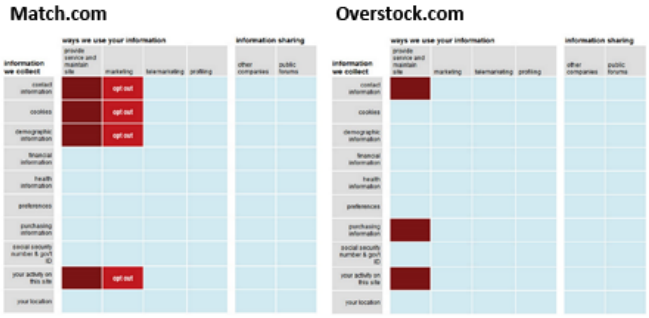


Figure 4.10

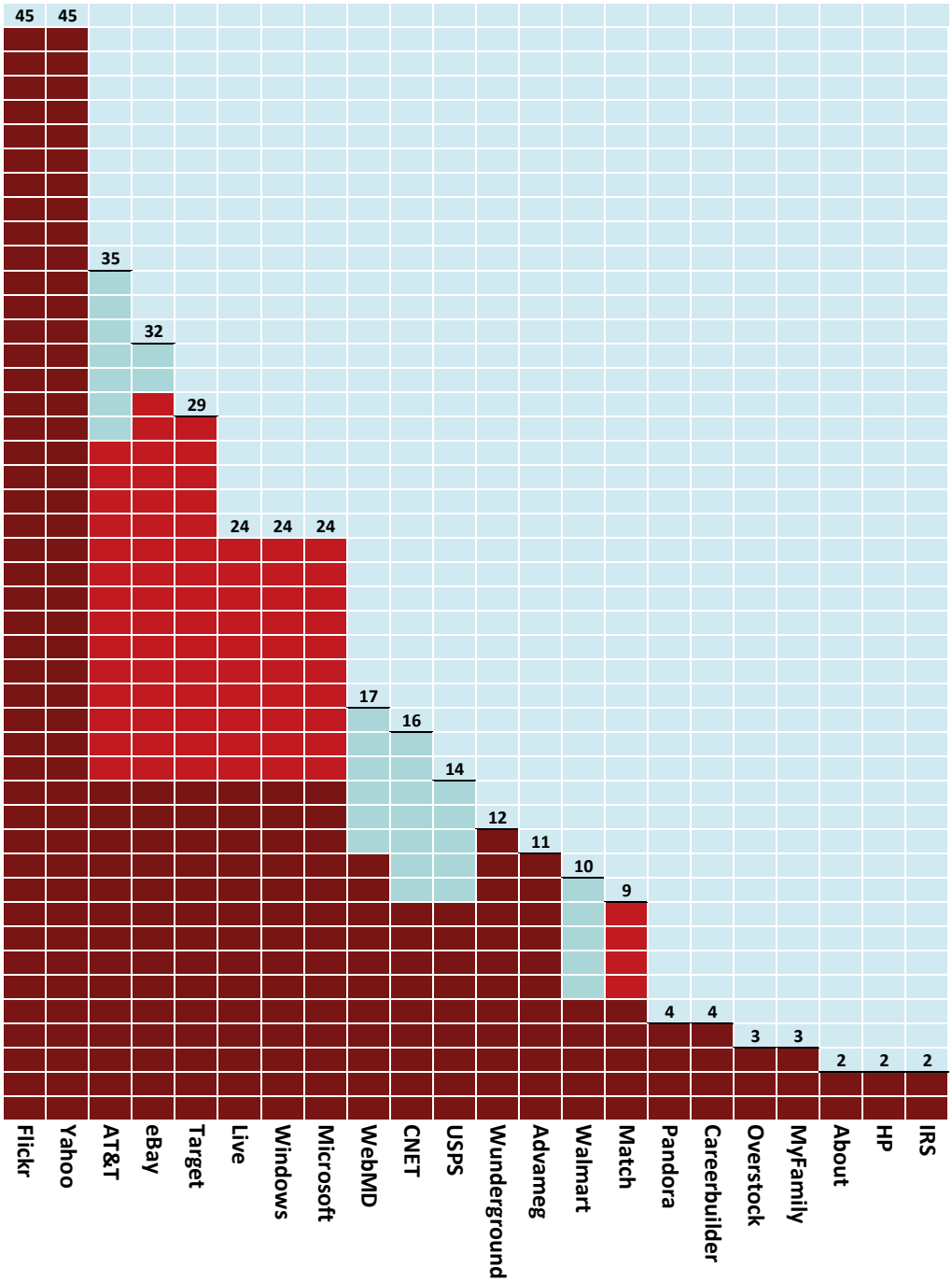


Figure 4.11: A summary of symbols used in the 22 retrieved Nutrition Labels

Figure 4.11 summarizes the symbols used in each of the 22 retrieved policies, where each column represents a policy. While the Nutrition Label has space for 60 symbols, the mean amount of symbols among the 22 policies was only 17. By excluding the two policies with 45 symbols belonging to the Yahoo group (the Yahoo! Inc. and Flickr policies), and the IRS and HP policies with 2 symbols each, the mean is down to just 15.

There can be several reasons for why the majority of the Nutrition Labels retrieved from the Privacy Finder search engine has a very low rate of symbol-use. As mentioned several times already, they are automatically created based on a P3P policy, meaning they might not be accurate representations of the companies' textual policies. As the P3P language is significantly more complex than the Nutrition Label, there might also be issues regarding the conversion between the formats.

Based on these findings we suggest the following change to the Nutrition Label:

- Consider removing the "third dimension" and replace its functionality in an alternative way.

4.2.3 Lack of a clear statement of what information is collected

Another issue we encountered with the Nutrition Label was the lack of a clear statement on which types of information is collected. The focus in the label is on how the information is used and with whom it is shared to, and it is up to the user to make the connection between used/shared and collected.

The following section explains this in detail, starting with a user story to outline the problem:

User story: As a prospective user of Acme Inc's services, John would like to see what information they will collect from him.

Consider the Nutrition Label in figure 4.12 for the fictive corporation Acme Inc. as used in the study by Kelley et al. [2010]. The column "information we collect" is defined, but does not provide any information regarding whether this information is collected or not. An experienced user would most likely look for rows with red squares, but without training or knowledge of the label structure or the

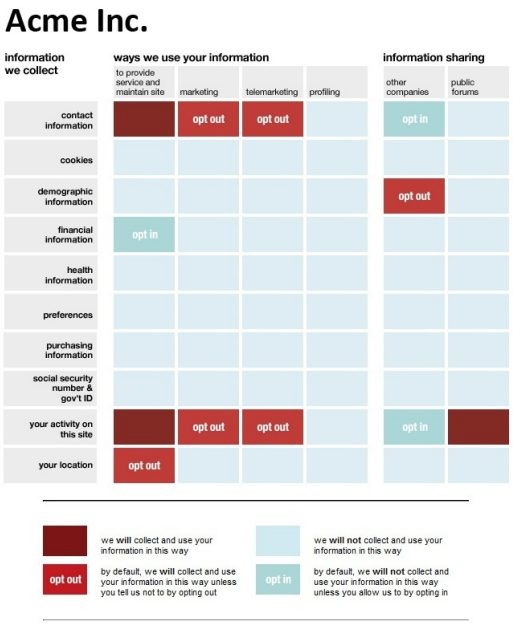


Figure 4.12: The fictive Acme Inc. policy from Kelley et al. [2010]

various symbols, a novice user would need to explore each row in detail to acquire this information.

The first row of the label is given in figure 4.13. The user is presenteted with four different symbols, four categories of how information might be used, and two categories of information sharing. As highlighted by the arrows (which indicate how the user most likely would navigate through the Nutrition Label), two of the four symbols indicates that the information will be collected by default.

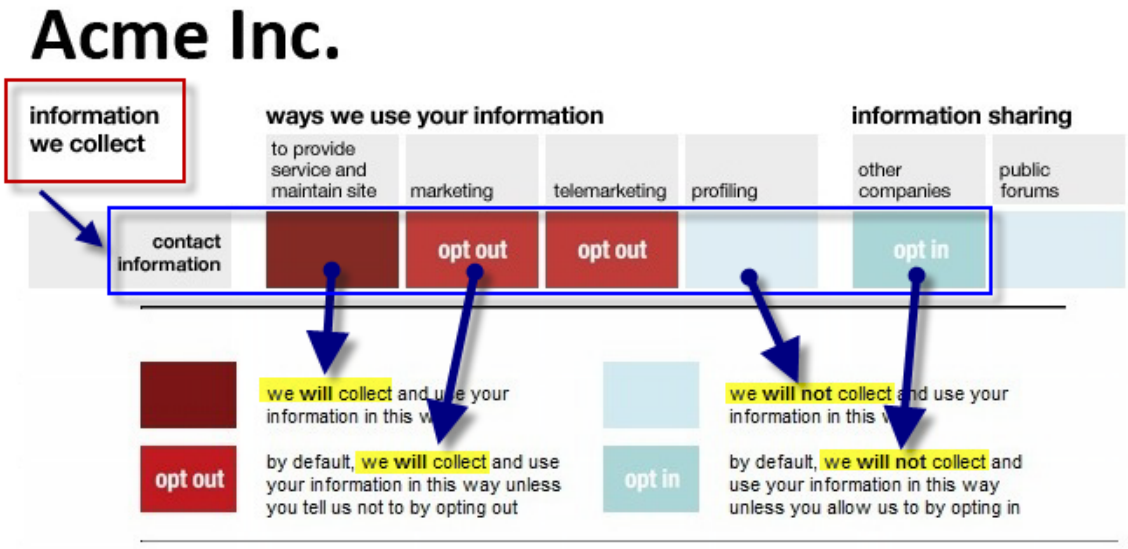


Figure 4.13: Demonstration of four different sharing practices for one data type (contact information)

Given that the first square in this row is red, the user would most likely quickly conclude that this information is collected by Acme Inc. The nature of the Nutrition Label does however not require any of the first elements in a row to be colored red for a information type to be collected. A company might collect information with the sole purpose of sharing this with another company, as seen in figure 4.14.

Now consider figure 4.14. While the first element in the row (arrow 2) indicates that "we will not collect and use your information in this way", the opt-out-symbol under "information sharing" indicates that Acme Inc. in fact does collect demographic information. In order to get to this conclusion, the user would need to familiarize himself with both the concept of "ways we use your information" (arrow 3) and also "information sharing" (arrow 4) while at the same time keeping control of the various symbols. In other words, the label does not answer whether information type A or B is collected, but rather how A and B is used or with whom it is shared. It is up to the user to make the connection between collected and used for/shared with.

We therefore suggest the following change to improve the usability of the label:

- Provide a clear statement of what information types are collected

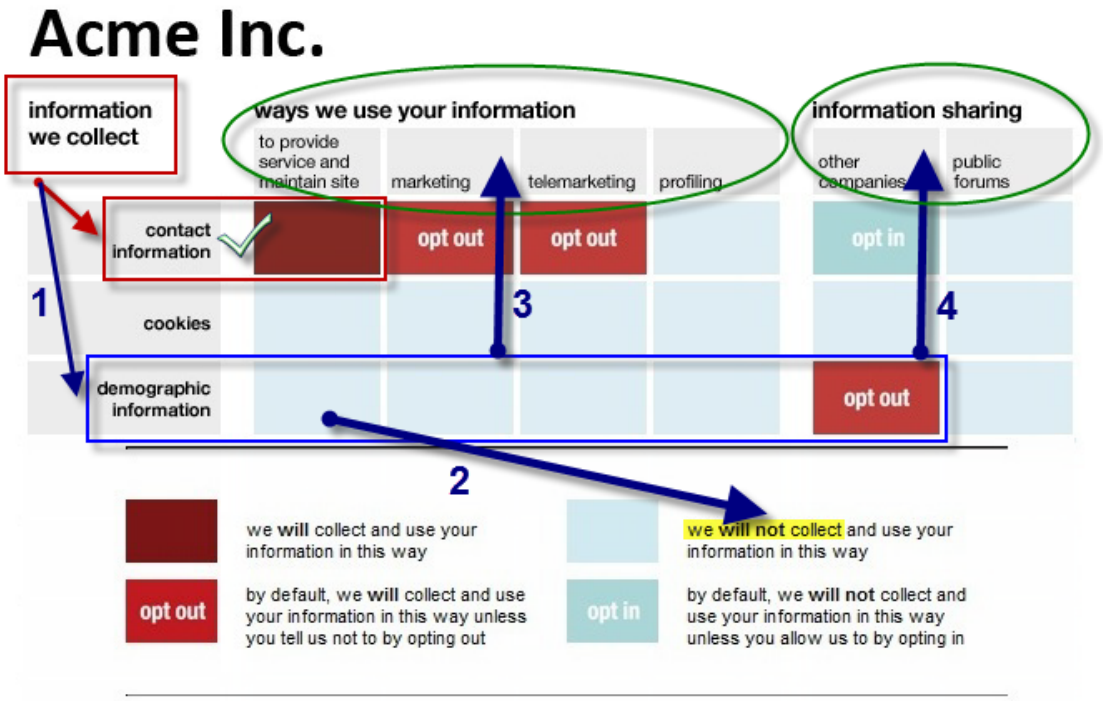


Figure 4.14: Demonstration of a row where the first four symbols indicates that demographic information will not be collected

4.2.4 Other issues

We also encountered several other possible issues regarding the Nutrition Label. First and foremost we are unsure whether some of the terms used in the label are well understood by non-native English speakers. Examples are "opt-in" / "opt-out", demographic information, profiling and cookies. While this issue has been addressed by the latest expansion of the label where an additional page with definitions has been added, a consideration could be to use more universally understood terms, or provide examples in the matrix structure itself.

A second, and often mentioned issue in the preceding sections, is the fact that the Nutrition Label is based on the P3P language. As this is out of the scope for this paper it will not be discussed in details, but we believe the authors behind the Nutrition Label should consider basing it on a different language in future prototypes. According to Wang and Kobsa [2008], the P3P language has gained little popularity, as demonstrated through our search for Nutrition Labels using the Privacy Finder search engine. The complexity of the P3P language is not needed to represent the Nutrition Label, and we believe a simple XML-based format, with one variable per symbol could be a useful approach for any further development.

Also, while we did not investigate each Nutriton Label we retrieved in detail, the content of several policies seemed odd when considered the purpose of the websites they belonged to. For example, the policy belonging to Match.com, which is a dating

site, claimed to not use any information for profiling purposes, which again might be due to a error in converting the P3P policy to the Nutrition Label format.

4.3 Design process

The following section presents the design process of our alternative solution. We first give details regarding our overall design strategy, before summarizing the main findings from the preceding sections in forms of a design criteria list. We end this section and the design chapter by presenting the resulting design of the solution, which was titled the "Privacy Table".

4.3.1 Design strategy

An important decision to make before starting the design process was to choose between keeping or modifying the original content of the Nutrition Label. We chose the latter due to several reasons.

First and foremost, as demonstrated by the currently 5954 word long Facebook.com policy in figure 4.15, compared to the 39 word long policy in figure 4.16, todays privacy policies vary from just a few sentences to many pages. Representing its content in form of a simple label requires knowledge in two areas: which practices is important and crucial to communicate to the user, and which practices the user is interested to see. Defining these two areas would require a thorough analysis which would be out of the scope for this paper. As the Nutrition Label and the P3P language has been through several evaluations, we believe that the current content represents these two areas well.

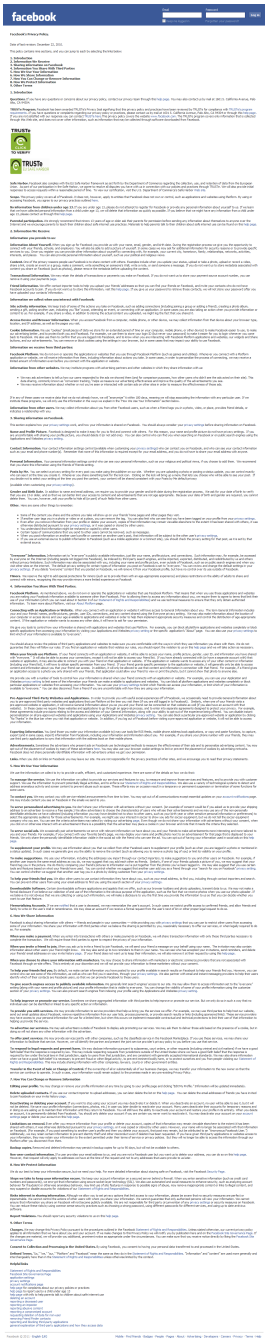


Figure 4.15: The Facebook.com privacy policy as of May 2011

Privacy Policy

We probably have the shortest privacy policy you'll ever see:

We don't share anything. Period. No vague marketing partners or marketing affiliates - just no sharing of any kind.

We value your privacy the way we value our own.

Figure 4.16: The privacy policy for "highpressureshowerheads.com"

Secondly, by modifying the content of the Nutrition Label we would have made the evaluation of the two formats more challenging. In a comparison setting, any questions asked would need to be suitable for both formats. For example, if we had removed the "information sharing to public forum" category from our new design (as neither of the 19 retrieved polices in section 4.1 utilized this column), any question regarding this in the evaluation could only have been answered by using the Nutrition Label.

Based on this, we chose to keep the original structure of the Nutrition Label with the various data information types and the collect, use, and share-categories. By doing this, it also enabled us to directly measure the performance of the proposed design changes as both formats in the evaluation would contain the same information.

We also chose to exclude the latest contextual additions to Nutrition Label from the design process. As previously mentioned, we believe future research on a standardized format should focus on further development of the core label design, rather than additions. The contextual information would also have been the same for both formats, resulting in similar results from the evaluation.

4.3.2 Design criteria

The following list summarizes the findings from the previous sections and atcs as a list of criterias for the design of the Privacy Table:

- Providing a clear starting point for reading the label
- Making the row and column titles and headers more dominant
- Making it clearer which titles belongs to which columns and rows
- Decreasing the size and the amount of symbols in the matrix
- Adding more white space
- Reducing the number of axis in the label
- Reducing the amount of scrolling needed to understand symbols
- Removing the "third dimension" and replace its functionality in an alternative way
- Providing a clear statement of what information types are collected

As most of these suggestions were related to either the surrounding structure (e.g. headers) or the matrix structure (e.g. changes to symbols) of the Nutrition Label, we grouped the findings into the following two categories:

- The surrounding structure: titles, categories, the row and column headers and additional information
- The matrix structure: the core structure with the various symbols and corresponding legends

Table 4.2 shows which finding belongs to which category, where criterion 9 was relevant for both. We will refer to this list through the following sections describing the design process, and through the evaluation planning in chapter 5, where these criteria worked as a basis for developing an evaluation hypothesis.

The surrounding structure	The matrix structure
1 Providing a clear statement of what information types are collected	5 Removing the "third dimension" and replace its functionality in an alternative way
2 Making the row and column titles and headers more dominant	6 Decreasing the size and the amount of symbols in the matrix
3 Providing a clear starting point for reading the label	7 Reduce the amount of scrolling needed to understand symbols
4 Making it clearer which titles belongs to which columns and rows	8 Adding more white space
9 Reducing the number of axis in the policy	

Table 4.2: The findings grouped by label area

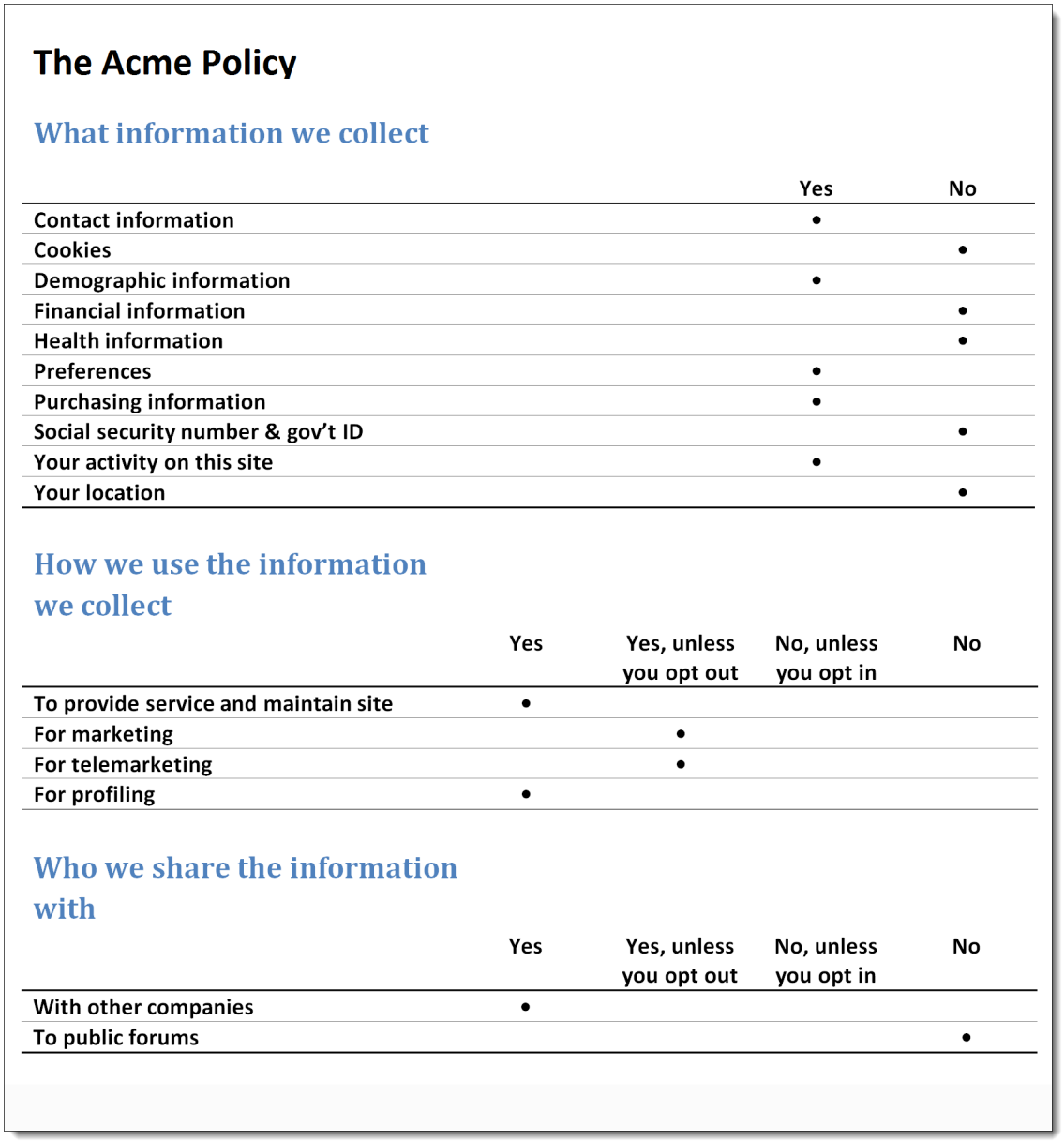
4.3.3 First version

Our first design proposal can be seen in figure 4.17, with the corresponding Nutrition Label in figure 4.18. While our newly proposed design might seem significantly different from the Nutrition Label, it is able to present the exact same content of the Microsoft Corp. policy.

The data information categories (e.g. contact information or financial information), and the three groups for collecting, using and sharing has all been kept the same as in the Nutrition Label. As mentioned in section 4.3.1, we did not include the definition page nor the additional information area from the Nutrition Label.

The following changes were related to the matrix structure:
The biggest change we did was to eliminate the "third dimension" (5) and to reduce the number of axis by splitting the design into three main categories (9). By removing the use of colored symbols (6) we have also removed the need to scroll down for symbol-definitions (7) and at the same time added white space between the dots to provide breathing space in the label (8).

The following changes were related to the surrounding structure:
By reducing the number of axis in the label (9) and replacing the functionality of the "third dimension" by three clearly defined categories, where one specifies what information is collected (1), we have also provided a clear starting point for reading the label (3). We have increased the size of the headers to make them more dominant (2) and to clearly separate the three sections from each other (4).



The Acme Policy

What information we collect

	Yes	No
Contact information	•	
Cookies		•
Demographic information	•	
Financial information		•
Health information		•
Preferences	•	
Purchasing information	•	
Social security number & gov't ID		•
Your activity on this site	•	
Your location		•

How we use the information we collect

	Yes	Yes, unless you opt out	No, unless you opt in	No
To provide service and maintain site	•			
For marketing		•		
For telemarketing		•		
For profiling	•			

Who we share the information with

	Yes	Yes, unless you opt out	No, unless you opt in	No
With other companies	•			
To public forums				•

Figure 4.17: The first version of the Privacy Table

The system with black dots was inspired from the Facebook.com's privacy setting page, which can be seen in figure 4.19. We believe this system should be easier to understand than a symbol and legend structure, and by removing the legend-box we have also freed up space and eliminated the need to scroll up and down in the label.

	Everyone	Friends of Friends	Friends Only	Other
Your status, photos, and posts			•	
Bio and favorite quotations			•	
Family and relationships				•
Photos and videos you're tagged in			•	
Religious and political views			•	
Birthday			•	
Permission to comment on your posts			•	
Places you check in to [?]			•	
Contact information				•

Figure 4.19: Facebook.com privacy settings

The main drawback with this design is that it is unable to represent a Nutrition Label with variying elements along a single column as described in section 4.2.2 on page 43. For example, it is not able to represent the eBay Inc. policy in figure 4.9 on page 44. This is due to the eBay policy specifying that one need to opt-in for contact information to be shared with other companies, while cookies are shared this way unless opted-out.

To cover for this we introduced an asterisk with a corresponding footnote as seen in figure 4.20. This system is however not very scalable, and only usable when there are just one or two differences within a singe column.

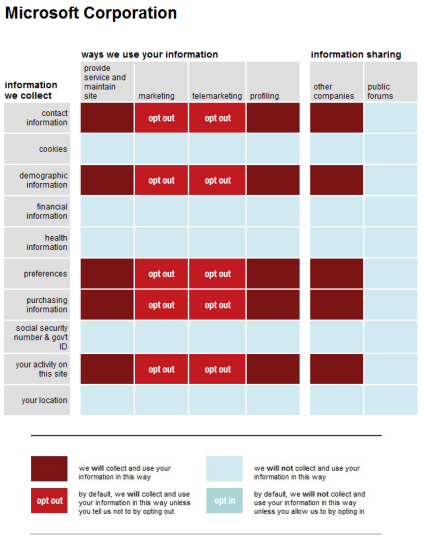


Figure 4.18: The same policy in the Nutrition Label format

How we use the information we collect

	Yes	Yes, unless you opt out	No, unless you opt in	No
To provide service and maintain site	•			
For marketing		•	(*)	
For telemarketing		•		
For profiling	•			

(*) = Financial information

Figure 4.20: Asterisk system in the Privacy Table

4.3.4 Subsequent versions

As described in the introduction, the design of the Privacy Table was an iterative process, based on the theory of design science. Guideline 6, proposed by Hevner et al. [2004] described the design process as a generate/test cycle, where alternative designs are generated and then tested against requirements.

The version presented in this section is therefore the result of the first iteration. Further iterations are presented throughout the subsequent chapters, as indicated by the list below.

- The Second version, following the pre-test (see figure 5.6 on page 76)
- The Third version, following the laboratory experiment (see figure 5.25 on page 98)
- The Fourth and final version, following the Internet experiment and subsequent discussion of results (see figure 7.6 on page 150)(this solutions is regarded as a merged version of the Privacy Table and the Nutrition Label)

CHAPTER 5

Evaluation method

Having proposed an alternative solution based on the findings from the Nutrition Label analysis, we wanted to evaluate whether our changes would have any effect on its performance in an experimental setting. Guideline 3 from the paper on design science by states that the utility, quality and effiacy of the designed artifact must be rigorously demonstrated via well-executed evaluation methods.

Previous research on the Nutrition Label has shown that it outperforms various textual alternatives such as full text, summarized or layered policies ([McDonald et al., 2009], [Kelley et al., 2009], [Kelley et al., 2010]). Through our pre-study we did however notice that while the Nutrition Label has been thoroughly tested against textual policies, it has just been tested once to any similar design (the shortened label version in Kelley et al. [2010]). It is therefore hard to know whether the Nutrition Label has performed well due to it being the best solution for presenting a privacy policy, or whether the good results are due to the lack of any similar competing solutions.

The following chapter first presents the purpose of the evalution, before elaborating on all related aspects regarding the evaluation methodology. The structure of the questionnaire is then given, before the pre-test, laboratory experiment and the additional Internet experiment is presented in details with corresponding questionnaires, policies, as well as procedure and instructions.

5.1 Purpose

The overall purpose of conducting an evaluation was to measure how our Privacy Table design would perform compared to the Nutrition Label under similar conditions as used in previous research experiments. Given the similarities between the Privacy Table and the Nutrition Label, other policy formats were+ excluded from the study in order to enable a stronger focus on the differences between the two designs.

More detailed, we wanted to measure whether the proposed design changes (table 5.1) would have any effect on the evaluation criteria that was derived from the initial problem definition (table 5.2). As the content of Privacy Table was kept the same as in the Nutrition Label, and considering that both solutions has the same relation to Privacy Enhancing Technologies (due to both being a visual representation of any underlying language), only criteria 1-4 in table 5.2 were relevant for the evaluation.

The surrounding structure	The matrix structure
1 Providing a clear statement of what information types are collected	5 Removing the "third dimension" and replace its functionality in an alternative way
2 Making the row and column titles and headers more dominant	6 Decreasing the size and the amount of symbols in the matrix
3 Providing a clear starting point for reading the label	7 Reduce the amount of scrolling needed to understand symbols
4 Making it clearer which titles belongs to which columns and rows	8 Adding more white space
9 Reducing the number of axis in the label	

Table 5.1: Criteria list for the design of the Privacy Table

1.	How easy is it to find the information the user is looking for?
2.	How easy and user-friendly is the solution to use?
3.	Does the solution require experience or can a novice quickly understand the format?
4.	How well does the solution allow for policy comparison?
5.	* How well does the solution manage to cover the details of a textual policy?
6.	* Can the solution be used in current or future Privacy Enhancing Technologies (PETs)?

Table 5.2: Evaluation criteria derived from the initial problem definition

5.2 Method

According to Oates [2006], an experiment is designed to disprove a hypothesis, excluding all factors that might affect the result, and is regarded as the most "scientific" and therefore most acceptable research strategy. If the researchers are confident that all other factors are excluded, the hypothesis that factor A causes the outcome B, is proven. A laboratory experiment also permits a high level in measuring outcomes and analyzing the data, and minimizes the time and cost incurred in visiting field sites.

According to Hevner et al. [2004], a controlled experiment enables the study of an artifact in a controlled environment and is appropriate for measuring usability. A natural choice of research method was therefore to conduct an experiment to test whether the design changes (A) in table 5.1 would have any effect on the evaluation criteria (B) in table 5.2.

5.2.1 Variables and hypothesis

When conducting an experiment it is important to distinguish between independent, dependent and controlled variables. By manipulating the independent variable (cause), we can observe the changes in in the dependent variable(s) (effect), while at the same time keeping the controlled variables the same throughout the experiment [Oates, 2006]. By controlling all variables this way, just one factor remains as the only viable cause for the observed change. Oates [2006] recommends using only one or maximum two independent variables to ease the statistical analysis.

Related variables for our evaluation can be seen in table 5.3. As we wanted to compare two different policy formats and see which performed best, the independent variable was the type of design given to the participant. Half of the participants used design A to answer the questions, while the other half used design B. Controlled variables included asking the same questions to both groups, and not telling them which solution was the Privacy Table and which was the Nutrition Label. For further information regarding variables and threats to validity we refer to section 5.2.7 on page 65.

Independent Variable (What is changed)	**Dependent Variables** (What is observed)	**Controlled Variables** (What is kept the same)
The policy format given to the participants,	- The time spent on answering the questions measured in minutes - The accuracy on answering the questions measured by correct answers - The enjoyability of using the format measured by a series of likert-questions	Asking the same questions to both groups Participants in both groups has similar previous knowledge and experience of the topic Participants unaware of which solution is our design and which is the Nutrition Label

Table 5.3: The research variables

To measure any change in the evaluation criterias in table 5.2, three dependent variables were relevant: time, accuracy and enjoyability. Time measured in minutes it took the participant to complete the various sections, accuracy measured by amount of correct answers, and enjoyability measured by the response to a series of likeability-questions.

Based on this, the nullhypothesis for the experiment was defined as:

- **H0:** The Privacy Table will perform the same way as the Nutrition Label in terms of accuracy, response time and likeability.

With the following alternative hypotheses:

- **H1:** The Privacy Table will perform better than the Nutrition Label in terms of accuracy.
- **H2:** The Privacy Table will perform better than the Nutrition Label in terms of response time.
- **H3:** The Privacy Table will perform better than the Nutrition Label in terms of likeability.

5.2.2 Repeatability

With a strong focus on the differences between the two formats, any other policy format was excluded from the experiment. This allowed us to directly compare the two formats in order to measure whether our proposed changes would have any effect, but disallowed us to compare any findings against other policy formats such as full text or layered notices. As previous research has thoroughly tested the Nutrition Label versus textual alternatives, a main research strategy was therefore to recreate the research conditions from these studies as best as possible. By eliminating the textual solutions from our experiment, but asking the same questions from previous studies, we were able to get an indication for the performance of the Privacy Table compared to full text and layered notices.

We therefore initiated the evaluation planning by analyzing the settings for each related experiment conducted on the Nutrition Label. The findings from this analysis, which can be found in Appendix B, was used thoroughly through the planning stages of the evaluation.

Kitchenham and Pfleeger [2008] list two important advantages for reusing existing research instruments:

- The existing instruments have already been assesed for validity and reliability (see section 5.2.7 on page 65 for further explanation of validity)
- By using common instruments, it is easy to compare new results with the results of other studies

Oates [2006] also highlighs that repeating experiments is important. Even the most carefully designed experiment might have been contaminated by some unrecognized other factor, and conclusions should not be drawn until the experiment have been repeated many times by both themselves, and other researchers. Other researchers should repeat the experiment to ensure that the same results are achieved and that they weren't influenced by faulty equipment or measurements by the original researcher.

As we were not repeating the previous experiments on the Nutrition Label, but rather reusing the instruments, we were unable to provide any verification of the previous conclusions. This is especially true for the findings from the experiments which utilized a design where each participant answered questions about multiple formats (a within-participant design as described in section 5.2.5 on page 62). For the experiments which had separate groups for each policy format, we were however able to provide some evidence for whether the performance of the Nutrition Label group in our experiment was similar to the performance of the Nutrition Label groups in those experiments.

5.2.3 Internet vs. laboratory experiment

An important decision was to choose between conducting an internet or a laboratory experiment. According to Cohen et al. [2007], conducting experiments over the internet is a popular approach for research on IS-systems. It is useful in situations where a large number of participants are required, and due to the wider sampling it offers greater generalizability of the findings. It has also been found that as participating in an Internet experiment is based on a greater degree of voluntariness and takes place in a familiar environment, more authentic behavior from the participants can be observed [Cohen et al., 2007].

While conducting an Internet experiment could provide greater generalizability, it would also require careful planning of the participant recruitment. By randomly distributing an invitation to participate, there is no control of who actually participates, resulting in the findings not being generalizable to any population [Kitchenham and Pfleeger, 2008]. As the Internet is not under control by the researcher, varying network speed, different browsers and platforms and so on might also affect the outcome of the experiment [Cohen et al., 2007].

Due to the large number of unknown factors which could affect the result, [Oates, 2006] claims that Internet experiments should be considered as a quasi-experiment approach. It has also been found that Internet-based experiments have a greater number of dropouts than laboratory experiments, due to motivation or how interesting the experiment was [Reips, 2002]. This could however be improved by offering incentives such as payments or lottery tickets, which have been found to reduce the dropout rate by as much as 31% [Reips, 2002].

Participant recruitment plays a vital role in choosing between an internet and a lab-

oratory experiment. As seen in Appendix B, the laboratory experiments by Kelley et al. [2009] and Reeder et al. [2008] included 24 and 12 participants respectively, while the Internet experiments by Reeder et al. [2008], McDonald et al. [2009] and Kelley et al. [2010] recruited an average of 765 respondents. While the difference in participants between previously conducted laboratory and Internet experiments can be explained by the number policy formats tested in the latter (Kelley et al. [2010] had 152 participants per format while McDonald et al. [2009] had 53 per format) the length of the experiments did however differ significantly. The laboratory study by Kelley et al. [2009] required the participants to spend 45 minutes answering the same questions for both formats (using a within-participant design, see section 5.2.5 on 62), while the Internet study by Kelley et al. [2010] divided the participants between the formats, resulting in a response time of approximately 15 minutes.

Key factors for choosing between online and internet experiments are therefore:

- The number of formats tested
- The number of participants needed, how many can be recruited, for how long they can be expected to participate and which degree of authentic behavior is needed
- To which population the results should be generalizable to
- Possible factors affecting the outcome

At first we considered conducting an Internet experiment based on the online study of the Nutrition Label approach by [Kelley et al., 2010]. This experiment utilized the Amazon Mechanical Turk which is a crowdsourcing Internet marketplace that enables computer programmers (known as Requesters) to co-ordinate the use of human intelligence to perform tasks that computers are unable to do [Amazon, 2011]. Workers can then browse among tasks and complete them for a small monetary payment. The Amazon Mechanical Turk eliminates the challenge of recruiting participants, and also permits setting requirements for who can participate [Amazon, 2011].

Kelley et al. [2010] also developed a custom designed tool called "Surveyors Point" for use with the Amazon Mechanical Turk. This tool enabled for accurate measuring of a number of variables, such as time and number of times the participant looked at a policy. While we acquired permission to use this tool for our experiment, we were disqualified of using the Amazon Mechanical Turk as it was only available for American based companies.

Based on the previous analysis, we therefore decided to conduct a laboratory experiment. The main reason for this was to be able to control who participated in the experiment. As we were unable to use the Amazon Mechanical Turk, the only recruitment method available for an Internet experiment would have been to randomly distribute an invitation. We did not acquire permission to use any e-mail lists

for this purpose at the university where this master thesis was written at, further complicating such a recruitment method.

A laboratory experiment permitted a higher degree of control over the experimental variables, and unknown factors which could have threatened the validity. We also expected the participants to be willing to spend more time in a laboratory setting, and with control over the sample population our findings would be generalizable to a target population.

5.2.4 Participant recruitment

We were offered to conduct the experiment as a part of a guest lecture on privacy online, held by Karin Bernsmed from Sintef ICT. The guest lecture was held for students taking the course TTM4135 Information Security (slides from the guest lecture can be found in Appendix D.3). The course was available for students at study level 3 (third-year students), primarily from the study programs Telematics, Industrial economics and technology management, Communication Technologies and Computer Science. 74 students were enrolled in the course during the spring semester 2011.

Our target population for the experiment would therefore be students from a technological study program. In other words, our findings would say something about how a student handled the Privacy Table and Nutrition Label, rather than the average user as originally intended. This recruitment method was therefore not ideal, but given the potential access to a large number of participants and a pre-reserved location, we decided to accept the guest lecture invitation. An overview of the advantages and disadvantages of this recruitment method is given in table 5.4.

+	Easy access to up to 74 students for 45 minutes
+	Pre-reserved location
+	Participants with similar background and experience - less variations between the groups.
+	As our experiment would replace a normal lecture, the students would have available time to participate and they were already present in the experiment location.
-	We did not know how many students would show up for the guest lecture
-	The students could choose to leave after the guest lecture
-	Using students for the experiment could threaten external validity as the results are not generalizable to a wider population
-	As the experiment would take place in an auditorium we were limited to conducting a paper-based questionnaire

Table 5.4: Advantages and disadvantages of the participant recruitment method

Given that the participants could choose to leave after the guest lecture (our experiment was planned for the 2nd lecture hour), we faced the challenge of pursuing as many students as possible to stay and participate, whilst not forcing or directing anyone to volunteer. Oates [2006] highlights the importance of volunteer participation and states that no one should be forced to participate by wheedling or threats. An important part of the experiment procedure was therefore to make the students aware of the importance of the experiment, while making it clear that participating was voluntary and that they could withdraw at any time.

Further details regarding the procedure of the experiment and the instructions given can be found in section 5.5.3 on page 92.

5.2.5 Between-group vs within-subject design

Experiments can be divided into two different types of design: between-group or within-subjects. In between-group experiments separate groups of particpants are used for each conditions, and each participant is tested once only [Field and Hole, 2003]. In between-subject designs each participant is exposed to all of the conditions of the experiment. Advantages and disadvantages of between-participant and within-subject design can be seen in table 5.5 and table 5.6 respectively [Field and Hole, 2003].

+	**Simple** - All you need to do is make sure participants are allocated randomly to the groups
+	**Less chance of practice and fatigue effect** - Performance in one condition can not affect performance in another
+	**Useful when it is impossible for an individual to participate in all experimental conditions** - for example if two different methods is used to learn a language. The participant cannot un-learn the language to try the alternative method.
-	**Expense in terms of time, effort and participant numbers**
-	**Insensitivity to experimental manipulations** - Differences within the groups could affect the outcome, rather than the experimental manipulations

Table 5.5: Advantages and disadvantages with between-participant design

As previously mentioned, the laboratory studies by Kelley et al. [2009] and Reeder et al. [2008] used a within-participant design, and the web-studies by Reeder et al. [2008], McDonald et al. [2009] and Kelley et al. [2010] used a between-participant design. Both Kelley et al. [2010] and McDonald et al. [2009] found it unrealistic to eliminate learning effects by reordering policies, and to expect participants willing to spend more than 20 minutes on an online questionnaire. Both studies that performed

+	**Economy** - Economical to run in terms of time and effort
+	**Sensitivity** - less "'noise"' from the differences between the participants
-	**Carry-over effect from one condition to another** - Participants might become tired, bored, better practices at doing things and so on.
-	**The need for conditions to be reversible** - Repeated measure design can only be used if being in one condition does not have irreversible effects that prevent the participant being used in another condition.

Table 5.6: Advantages and disadvantages with within-subject design

within-participant studies tested the their newly proposed design (Nutrition Label or P3P expandable grid) against natural language policies, which might partly explain the choice of design as learning effects between the two would be reduced.

Given the similarities between the Privacy Table and the Nutrition Label, it was concluded that the learning effects would have been too big by having the participants answering the same questions for both formats. As we also wanted to see how a novice user handled the format, it was important that the participants were unfamiliar with the policy structure.

The fact that the students were free to leave before the experiment started, also played an important role in choice of design. We did not "have the participants trapped in a dark room where we could test several things on them" [Field and Hole, 2003], but we rather had to keep the experiment as short as possible to attract students to remain in the auditorium and participate. As up to 74 students were expected to the guest lecture, this number of participants would also result in similar participant-per-format rate (up to 37) as in previous experiments (Appendix B).

5.2.6 Between-participant design alternatives

Field and Hole [2003] and Oates [2006] list four different types of between-participant designs: Static group comparison, Post-test only/control group design, Pre-test/post-test control group design and Solomon four-group design. See table 5.7 for an overview.

While *post-test only/control group design* often is mistaken for being a *static group comparison*, there is one significant difference between the two: randomization (indicated by R in table 5.7). In a *post-test only/control group design*, participants are randomly assigned to one of two groups, where one group (O) is given some treatment which is not given to the other group (treatment is indicated by an X in table 5.7). The performance of the groups are then measured: if it differs, one can be reasonably confident that it is attributable to the experimental manipulation [Field and Hole, 2003].

Type	Description	Figure
Static group comparison	A group (O) has experienced X and is compared with one which has not, for purposed of establishing the effect of X	X O O
Post-test only/control group design	Similar to static group comparison, but participants are randomized (R) into the two groups	R X O R O
Pre-test/post-test control group design	Similar to the above, but includes a pre-test before the treatment is applied to one of the groups to ensure there are no pre-treatment differences between the groups	R O_1 X O_2 R O_1 O_2
Solomon four- group design	A robust combination of a Post-test only/control group design and a Pre-test/post-test control group design	R O X O R O O R X O R O
		X = treatment O = observation R = randomization

Table 5.7: Overview of between-participant designs. The figures in the third column indicate the procedure in each design type, with each row corresponding to a single group. For example, the figure for the "*Post-test only/control group*" design indicates that: Group 1 was randomized (R), applied some treatment (X) and then observed (O). Group 2, the control group, was randomized (R) and then observed (O).

The *post-test only/control group design* design is commonly used, but has one weakness: if the randomization process of delegating participants to the groups fails to produce equivalence, there is no way of knowing that it has failed [Field and Hole, 2003]. Oates [2006] highlights the importance of a random assignment of participants to the two groups to ensure internal validity. If anything else than the manipulation of the independent variable produces systematic variation in performance, we won't be able to tell whether this is due to our manipulation or any unknown factor. For example, when using students as participants in the experiment, we might unknowingly assign all the best students to one of the groups.

While using a within-subject design would handle the problem of randomization (as all participants are tested for all conditions), there are also ways of ensuring randomization in between-participant designs. One way is to use a randomized number system to delegate participants as they arrive, but a more sophisticated way is to use a pre-test to make sure the groups are equivalent [Field and Hole, 2003].

In the *pre-test/post-test control group design*, performance of both groups are measured both before and after the treatment is applied (to one of the groups), and

any difference is assumed to be caused by the treatment. The pre-test might however have affected the participant's subsequent performance. *Solomon four-group design* controls the possibility for learning effects, but as four groups is required, it is expensive due to the number of participants needed [Oates, 2006].

For our evaluation, we chose a *post-test only/control group design* for several reasons. Most importantly, we felt a pre-test was unnecessary as we believed there were little differences in knowledge about privacy policy solutions among the students. Such a pre-test would also have been time-consuming and the potential learning effects could have been significant. To ensure equal knowledge regarding both formats, research on the Nutrition Label was not mentioned in the initial guest lecture on privacy online (the slides from the guest lecture can be found in Appendix D.3).

Based on this, we assumed that a random delegation process would be sufficient to ensure equal groups in the experiment. Several pre-cautions were taken during the randomization process. As Field and Hole [2003] highlights, delegating students to one of the groups as they arrive is not a good idea, as the best students are most likely to show up early. As our experiment took place in an auditorium following a regular lecture, we were also aware that their seat placement could be a factor: better students could be more likely to take a seat near the front.

Assuming their horizontal auditorium position were more or less random, our randomization process was therefore to split the auditorium in half by the middle. This method was also time-saving as we did not have to move the students around. While this process is as random as possible under the given conditions, internal validity could still be threatened as we had no control whether all the best students were seated to the left or right in the auditorium. An alternative to ensure complete randomization could have been to assign a random number to each student, and then draw two groups from these numbers.

5.2.7 Validity

When conducting an experiment it is important to control all the variables, to ensure that no unknown factors affect the outcome. Oates [2006] highlights the importance of this, but states that it can be difficulty or even impossible to achieve. Participants in group-experiments can have different previous experience, and their age, gender and work status might differ, something which could result in unbalanced groups. To minimize the impact of unforeseen factors, Oates [2006] suggests eliminating participants with previous experience, choosing random participants, using control groups, or making the researchers and subjects blind to avoid expectations that could influence the result.

5.2.7.1 Internal validity

An experiment has good internal validity if the measured outcomes are indeed due to the manipulations of the independent variable. According to Oates [2006], a common threat to internal validity is participants dropping out of the experiment, which can lead to bias in the remaining sample of participants. Our countermeasure to this was to clearly inform the participants about the time required to answer the questions and that participation was voluntary. We also requested that the ones not wishing to participate had to leave the room before the experiment started.

By not telling the participants which design was our Privacy Table and which was the established Nutrition Label, we also avoided what Oates [2006] refer to as "helping the researcher getting good data" which could serve as a threat to internal validity of the research. As experiments are usually performed in an artificial setting such as a laboratory, participants might also change their behavior as this is an unknown setting for them.

Faulty instruments are another threat to internal validity. One related instrument in our setting was the clock needed for the participants to be able to write down the time in the questionnaire. As we were unsure whether the participants would bring their own cellular phones or carry a wrist watch, we provided a digital PowerPoint clock in addition to an analog clock which was present in the experiment location.

The questionnaire itself can also be defined as an instrument in this setting, and it is important to make sure that it is measuring what we want it to measure. Fink and Litwin [1995] highlights that it is essential to evaluate a survey in order to check that the questions are understandable, to address the likely response rate, and to evaluate the reliability and validity of the instrument. According to Kitchenham and Pfleeger [2008], the two most common ways to organize an evaluation are focus groups and pilot studies.

As we were reusing questions and settings from previous experiments on the Nutrition Label, we did not consider the instrument to be of any major threat to internal validity. Our target group did however differ from the target groups of the previous experiments, and as we modified some of the reused questions we decided to conduct an pre-test which is further described in section 5.4 on page 71. This pre-test enabled us to see whether the questions and instructions were well-understood, and also to get an initial opinion of the Privacy Table design, as a part of the iterative design process.

5.2.7.2 External validity

While internal validity refers to whether the observations are indeed due to the manipulations of the independent variable, external validity refers to whether the results obtained are generalizable. With generalizable we mean whether a result that has been demonstrated in one research setting would be obtained in other settings,

with different research participants and different research procedures [Oates, 2006]. In other words, an experiment has a good external validity if the results are not unique for the particular circumstances. According to Oates [2006], the best way of demonstrating generalizability is to repeat the experiment in many situations, but also a carefully designed experiment could achieve a high external validity.

Threats to external validity include [Oates, 2006]: Over-reliance of special types of participants, too few participants, non-representative participants for the purpose of the test and non-representative test cases. Oates [2006] states that students or volunteers have been found to have characteristics that differentiate them from the general population. Given our participants were all students from the same university level, our results might therefore not be generalizable to a general population, but rather give an idea of how students with a technological background would use the formats.

5.3 Questionnaire design

The importance of repeating experiments was described in section 5.2.2 on page 58, and is particularly important when it comes to questionnaire design. Developing a new questionnaire from scratch is not only time consuming in terms of the design process, but it must also be assesed for reliability and validity, and tested through pilot-tests to ensure it measures what we want it to measure [Kitchenham and Pfleeger, 2008].

5.3.1 Questionnaire structure

Based on the strategy of recreating previous experiments on the Nutrition Label in order to compare findings, the starting point for designing the questionnaire was to retrieve as many questions as possible from the related studies on the label. As our dependent variables were similar to those measured in the studies by Kelley et al. [2009] and Kelley et al. [2010] (time, accuracy and enjoyability), these two studies were used as a basis for outlining the questionnaire.

As all questions from these studies were not published, and those made available did not necessarily test all the aspects regarding the differences between our Privacy Table and the Nutrition Label, the final questionnaire was a mix of reused, modified and additional questions. Having amended the research instrument, we therefore had to repeat the pilot testing through the pre-test (described in section 5.4).

Table 5.8 and 5.9 provides an overview of the questionnaires used in the studies by Kelley et al. [2009] and Kelley et al. [2010]. For a detailed overview of retrieved questions from these the two studies and their subsequent use in the pre-test and final questionnaire, we refer to table B.2 and B.3 in Appendix B.

Kelley et al. [2009]		
	Part	**Description**
1.	Information Finding	8 Yes/No information finding questions regarding a single policy
2.	Perceived Privacy Policy Understanding	6 Questions on a 5-point Likert scale, from Strongly Disagree (1) to Strongly Agree (5).
3.	Policy Comparison Questions	6 policy comparison questions- Four True/False statements and two opinion questions
4.	Policy Comparison Enjoyment	4 Questions on a 5-point Likert scale, from Strongly Disagree (1) to Strongly Agree (5).

Table 5.8: Question parts in the survey by Kelley et al. [2009]

Kelley et al. [2010]		
	Part	**Description**
1.	Demographic questions	Standard information about the participants: gender, age, and current occupation
2.	Internet and privacy	4 Questions regarding Internet usage and prior knowledge of privacy. Answer options: "Yes," "No," or "The policy does not say." (except question 4)
3.	Single policy simple tasks	6 Questions which could be answered by looking at a specific row or column in the table. Answer options: "Yes," "No," "Yes, unless I tell them not to," "Only if I allow them to," or "The policy does not say."
4.	Single policy complex tasks	6 Questions with interaction between some category of data and either data use or data sharing.
5.	Single policy Likeability	6 Likert questions for qualitative feedback on the format
6.	Policy comparison tasks	5 Questions, 3 three information-finding and 2 preference questions. Customized answer options for each question
7.	Policy comparison likeability	3 Likert questions for qualitative feedback on the task of comparing two policies

Table 5.9: Question parts in the survey by Kelley et al. [2010]

As our participants were students from the same university-level, we chose not to include any demographic questions. Given they were all enrolled in various technological study programs, we also assumed that they had similar experience with technology and privacy, and part 2 from Kelley et al. [2010] was therefore also excluded from the experiment. The remaining five (3-7) parts from Kelley et al. [2010] with corresponding questions in part 1-4 in the Kelley et al. [2009] study therefore became the base structure for our experiment.

The overarching structure of the experiment is given in figure 5.1. To clearly separate the tasks of answering singly policy questions and comparing policies, the questionnaire was divided into two parts.

	Part 1
1.	Single policy simple tasks
2.	Single policy complex tasks
3.	Single policy Likeability
	Part2
4.	Policy comparison tasks
5.	Policy comparison likeability

Figure 5.1: Questionnaire structure

5.3.2 Policies used

When it came to selecting privacy policies for the experiment, we had two main design strategies to choose between:

1. Reusing the fictive policies from the studies by Kelley et al. [2010] and Kelley et al. [2009].
2. Using a "real-life" company policy

While reusing the fictive policies from the previous studies would be the most time-saving strategy, as these were already custom designed for the questionnaires, this strategy had some flaws. First of all, the policies used in these studies are not necessarily a good representation of how a realistic Nutrition Label policy looks like, as outlined in the design chapter. As seen in figure 5.2, the Acme policy utilizes all aspects of the Nutrition Label format, while most of the policies we retrieved from the privacy Finder search engine looked more like the three policies in figure 5.3. The Acme policy in figure 5.2 also makes use of the "third dimension", a functionality which only 3 (out of 22) of the policies we retrieved in our search utilized.

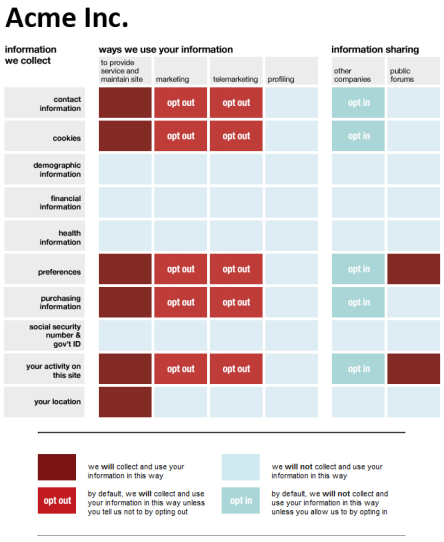


Figure 5.2: The Acme policy used in the study by Kelley et al. [2010]

As the design of the Privacy Table was somewhat based on the more simple "real-life" Nutrition Labels, reusing the Acme policy could therefore favor the Nutrition Label design.

Using a "real-life" policy also had some drawbacks. Primarily, it would require us to modify all the questions in the questionnaire to fit the policy. Secondly, by using a too simple policy, we could favor the Privacy Table and also make the questions

Figure 5.3: The privacy policies of Microsoft Corp., IBM and Yahoo! Inc. as represented in the Nutrition Label format by the Privacy Finder search engine

less challenging. Thirdly, for the comparison part we would have to find a second "real-life" policy which differed enough to allow for interesting comparison questions. These policies would also have to be clearly distinct as one question should only have one answer, and there should be no room for participants misinterpreting any of the questions.

Finally, the major changes required to the questionnaire by using a "real-life" policy would result in our findings not being comparable to the findings from the studies by Kelley et al. [2010] and Kelley et al. [2009], which was one of the main strategies for the evaluation.

Based on this, and aware of the possibility of favoring of the Nutrition Label, we decided to reuse the Acme policy from the Kelley et al. [2009] study for our evaluation. More spesifically, we reused the content of the Kelley et al. [2009] policy and presented it in the updated design from Kelley et al. [2009] (Note that for the pretest we also used the design from the Kelley et al. [2009] study as further explained in section 5.4.4).

5.3.3 Measuring time

To measure time spent on answering the questions, time boxes as seen in figure 5.4 was added before and after the various parts.

Figure 5.4: Time boxes used to measure time

While the accuracy of this method to measure time would not be precise and only provide results in terms of minutes, it was the only alternative given the question-

naire was paper-based. A PowerPoint clock was provided during the experiment in case the participants did not have any clock or the auditorium clock was not visible.

5.4 Pre-test

The following section describes the pre-test in detail, starting with a presentation of the purpose followed by a short description of the procedure and participants recruited. The questionnaires and policies used are then elaborated, before we analyse the findings, and present a re-designed version of the Privacy Table based on these.

5.4.1 Purpose

In order to test the structure and get feedback on both the questionnaire and the Privacy Table, a pre-test was conducted. A "pre-test" is in this setting defined as an evaluation of the questionnaire prior to the final experiment, and not a pre-test to measure existing knowledge or ability of the participants as referred to in section 5.2.6 on page 63. The participants in the pre-test did not take part in the final experiment.

The main goal of the pre-test was to get oral feedback on the questionnaire structure, and on to find out whether the questions from the previous studies were understandable precise. As correct or wrong answers, and how each format performed was not important, we only recruied four participants for the pre-test. Three of the participants had previous experience with the Nutrition Label, and one participant was updated on the development of the Privacy Table.

5.4.2 Procedure & participants

The pre-test was carried out in a meeting room with four employees from Sintef ICT. The participants were instructed to act as they would in a real experiment and briefed of the background and goal of the experiment. They were not told which label was the Nutrition Label and which was the Privacy Table, but as three participants already had previous experience with the Nutrition Label, we assume this might have affected the outcome of the pre-test.

The participants were divided in two groups, where one group received the Nutrition Label and the other group the Privacy Table. As there was no clock available in the meeting room, the participants used their own cellular phones for filling out the start and end time for the various parts.

Having completed the questionnaire, the participants provided feedback through a discussion session.

5.4.3 Questions

The questionnaire used in the pre-test can be found in Appendix C.1, and mainly consisted of reused questions from the Kelley et al. [2009] and Kelley et al. [2010] studies. A complete overview of retrieved questions from these studies and their subsequent use in the pre-test can be found in table B.2 and B.3 in Appendix B.

	Part2
1.1	**Single policy simple tasks**
	Question 1,2,3,5 and 6 from Part 3 in [Kelley et al., 2010] were reused (see table B.3 in Appendix B). Question 4 was excluded as we did not include any privacy seal icons in the policy design. The following question was added: "*Does the policy allow Acme to collect information regarding your household income?*"
1.2	**Single policy complex tasks**
	All questions from Part 4 in [Kelley et al., 2010] were reused (see table B.3 in Appendix B). Question 8 was modified from "home phone number" to "mobile number".
1.3	**Single policy Likeability**
	All questions from Part 2 in [Kelley et al., 2009] were reused (see table B.2 in Appendix B).
	Part2
2.1	**Policy comparison tasks**
	All questions from Part 6 in [Kelley et al., 2010] were reused (see appendix table B.3 in Appendix B), in addition to question 10 from Part 3 in [Kelley et al., 2009]) (see table B.2 in Appendix B). Question 11 and 12 from the latter were excluded as they were preference questions. The following question was added: "*By default, both companies collects your current location and can use this information to improve their services*".
2.2	**Policy comparison likeability**
	Question 15 and 16 from Part 4 in [Kelley et al., 2009]) were reused, while question 13 and 14 from the same study were excluded as they were single policy likeability questions. The following question was added: "*If all policies looked like this I would compare privacy practices across websites more often*"

Table 5.10: Overview of the questions in the pre-test questionnaire

Time blocks (as described in section 5.3) was added before and after part 1.1, 1.2 and 2.1. To capture opinions regarding the questionnaire format and the policy design, two open ended questions were added at the end of the questionnaire. The first was "Do you have any comments or suggestions regarding the privacy policy

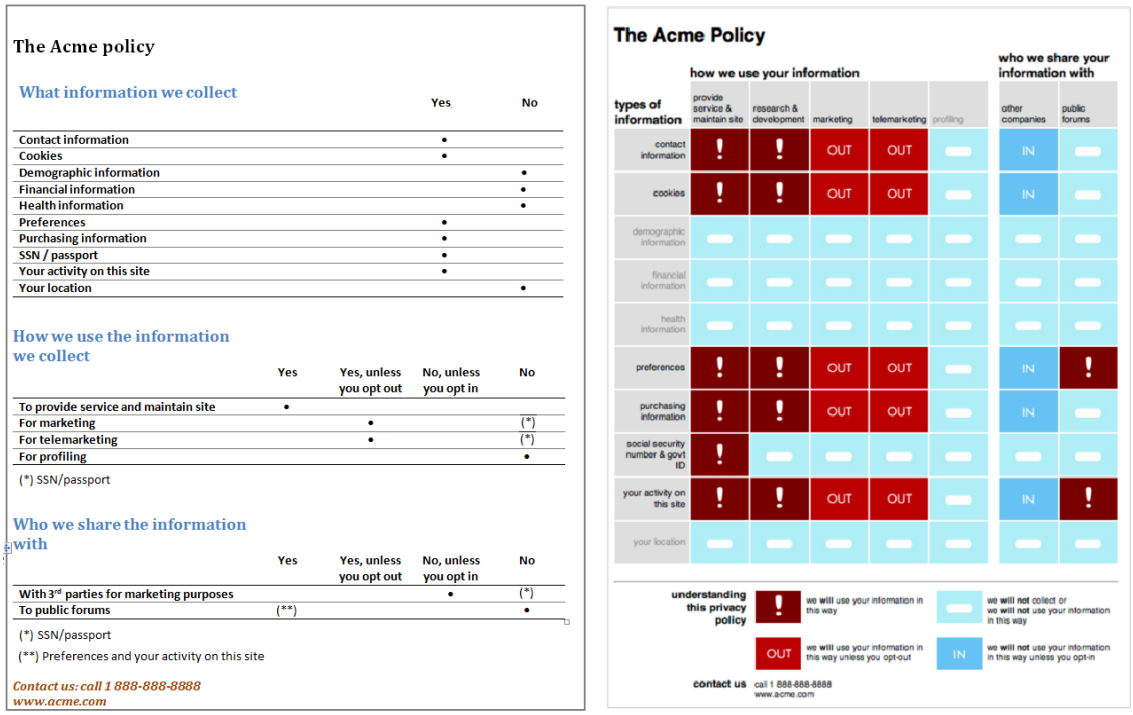
format?" and the second was "Do you have any feedback regarding the format of this survey?".

The participants in the pre-test answered these two questions orally through a discussion session.

5.4.4 Policies used

The Privacy Table policy used in the pre-test, as seen in figure 5.5(left), was the initial design (described in section 4.3.3 on page 51) with some modifcations to match the Nutrition Label. Its content was based on the reused Acme policy (figure 5.5 (right)) from the Kelley et al. [2009] study. As seen in the figure, our asterisk system did not scale well to the complexity of the Nutrition Label.

As most information finding questions was reused from this study, no changes had to be done to the policy content. The additional questions in part 1.1 and 2.1 were customized to fit the Acme policy. Due to a lack of time we did not update the original Nutrition Label ("Acme policy") to the latest design, but as the focus of the pre-test was on the questionnaire structure we did not consider this to be an .



The Acme policy

What information we collect

	Yes	No
Contact information	•	
Cookies	•	
Demographic information		•
Financial information		•
Health information		•
Preferences	•	
Purchasing information	•	
SSN / passport	•	
Your activity on this site	•	
Your location		•

How we use the information we collect

	Yes	Yes, unless you opt out	No, unless you opt in	No
To provide service and maintain site	•			
For marketing		•		(*)
For telemarketing		•		(*)
For profiling				•

(*) SSN/passport

Who we share the information with

	Yes	Yes, unless you opt out	No, unless you opt in	No
With 3rd parties for marketing purposes			•	(*)
To public forums	(**)			•

(*) SSN/passport

(**) Preferences and your activity on this site

Contact us: call 1 888-888-8888
www.acme.com

The Acme Policy

how we use your information

who we share your information with

types of information	provide service & maintain site	research & development	marketing	telemarketing	profiling	other companies	public forums
contact information	!	!	OUT	OUT		IN	
cookies	!	!	OUT	OUT		IN	
demographic information							
financial information							
health information							
preferences	!	!	OUT	OUT		IN	!
purchasing information	!	!	OUT	OUT		IN	
social security number & govt ID	!						
your activity on this site	!	!	OUT	OUT		IN	!
your location							

understanding this privacy policy

! we will use your information in this way

we will not collect or we will not use your information in this way

OUT we will use your information in this way unless you opt-out

IN we will not use your information in this way unless you opt-in

contact us call 1 888-888-8888 www.acme.com

Figure 5.5: Policies used in the pre-test, Privacy Table v1 (left) and Nutritional Label (Right)

5.4.5 Pre-test findings

The following list summarizes the findings from the pre-study with corresponding suggestions for change:

1. Observed: the participants studied the policy before writing the start-time and answering the questions, which could have influenced the time spent on each task. In other words, they could already be familiar with the content of the policy before answering the questions. Some participants also questioned whether they could switch between the policy and the question sheet.
   Change: Better explanation the survey-structure. Tell the participants that they should answer each question by looking it up in the policy. Also consider moving the policy to a separate sheet.

2. Some participants commented that they would like a better explanation of what to do, and what this was all about.
   Change: A more thorough introductional text, and a note about how the collected information will be used, anonymously, time to perform etc.

3. The time: Participants responded that they felt they had to write down the time too many times. Unnecessary to report it for each part. The reason behind this was to separate the time spent on the information finding question from the preference-questions, as we wanted to measure how long time it took to find answers in the policy.
   Change: Either collect start/end time for the entire questionnaire or for each part.

4. The participants missed an "I don't know" option for the questions. The respondent might not have understood the question, for example he/she could be unfamiliar with the term "cookies". The questions used different terms than the policies. For example, to answer a question about the use of credit card, the participant had to understand that it belonged to the 'financial information' category.
   Change: Add an "I don't know" option.

5. Some participants commented that question part 1.2 was splitted across two sheets of paper, and suggested that it should be on a single page. Some participants also suggested to remove the various parts, and present all questions in a single table. This would be difficult to achieve as the parts had variying answer options.
   Change: Make sure no parts are splitted across two sheets of paper.

6. In part 2.1, some participants responded that question 18-20 and 21-22 were of different format. 18-19 were yes/no-questions, while 20-22 was true/false-statements. It was suggested to modify question 18-19 to be true/false statements and change the answer options.
   Change: Modify question 18-19 to true/false-statements.

7. The use of the word'enjoyable' in part 2.2 was also questioned. Participants commented that reading a policy can never be an enjoyable experience, and suggested using a neutral term like "ok". It was also suggested to ask a negatively posed question, such as 'it was hard to read' or "hard to find information".
   Change: Use a different word than enjoyable, and modify some questions to a negatively proned form.

8. Observed: Include a guide on how to correct an answer. For example, by filling the entire box black.

9. It was suggested to improve the language of question 13 and 18

10. In the comparison part, the Bell policy was placed to the left of the Acme policy, which some participants found confusing.
    Change: Exchange the placement of the Bell and Acme policy.

11. One participant commented that the "who we share information with" on the Privacy Table was a bit unclear. More specifically it was unclear what sharing to "3rd parties for marketing" implified.

5.4.6 Proposed design changes

The time between the pre-test and the laboratory experiment was spent on improving the questionnaire design and the Privacy Table design. As described in the introduction, the design of the Privacy Table was an iterative process, based on the theory of design science (with guideline 6 proposed by Hevner et al. [2004]describing design as a search process). While the main purpose of the pre-test was to improve the questionnaire, we also received feedback regarding the format of the Privacy Table. These findings resulted in a re-design process of the Privacy Table.

The main issue we encountered with the Privacy Table was that the two participants using it to answer the questionnaire spent longer time on completing the tasks than the ones using the Nutrition Label. While the Privacy Table performed better than the Nutrition Label on the simple information finding questions, the Nutrition Label had a higher rate of correct answers on the complex information finding questions. The good results on the simple tasks indicated that the design change with including a clear statement of what information was collected had an effect, but we had expected better results on the complex tasks.

More specific, while our label worked well for questions on the following format:

- Does the policy allow Acme to collect information about **<example/scenario>**,

it did not perform well on questions formed as:

- Does the policy allow Acme to **<share/use>** your **<information type> <on place/ for reason>**.

Given we only had four participants in the pre-test, these findings are however just weak indications of the performance of the Privacy Table. Some of the changes presented below were already under consideration, and not direct results of the pre-test findings.

The four main changes to the Privacy Table as a result of the re-design-process were:

- We returned to the matrix system of the Nutrition Label
- We introduced a second light-gray icon in the matrix to indicate opt-in or opt-out
- We added an extra row below the table to indicate what choices the user has regarding the use of the information. This can be seen on as a implementation of the "column summary" as indicated in figure figure 5.8.
- We shaded the rows for collected data types in a light grey color

The decision of returning to the matrix system of the Nutrition Label was mainly due to that we did not believe the current version of the Privacy Table handled advanced policies very well. The asterisk solution presented in section 4.3.3 on page 51 did not scale very well, and we also found the separation of the policy into three distinct sections to be wasting unnecessary space. The resulting second version of the Privacy Table can be seen in figure 5.6.

The Acme Policy

What information we collect			How we use this information				Who we share this information with	
	Yes	No	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	•		•	◦	◦		◦	
Cookies	•		•	◦	◦		◦	
Demographic information		•						
Financial information		•						
Health information		•						
Preferences	•		•	◦	◦		◦	•
Purchasing information	•		•	◦	◦		◦	
Social security no. / Govt. ID		•						
Your activity on this site	•		•	◦	◦		◦	•
Your location	•		•					
Do you have any choices regarding the use of this information?	:		No	◦= You can opt out from this use	◦= You can opt out from this use		◦= Only if you opt in	No

Figure 5.6: The second version of the Privacy Table

The surrounding structure	The matrix structure
1 Providing a clear statement of what information types are collected	5 Removing the "third dimension" and replace its functionality in an alternative way
2 Making the row and column titles and headers more dominant	6 Decreasing the size and the amount of symbols in the matrix
3 Providing a clear starting point for reading the label	7 Reduce the amount of scrolling needed to understand symbols
4 Making it clearer which titles belongs to which columns and rows	8 Adding more white space
9 Reducing the number of axis in the policy	

Table 5.11: The design criteria list from the design chapter

While this version had more in common with the nutriton label-design than the first version, it still followed the original design criteria list from the design chapter, repeated in table 5.11.

The following changes were related to the surrounding structure:
As indicated by the yellow area in figure 5.7, we still provide a clear statement of what information types are collected or not (1), which also acts as a clear starting point for reading the label (3). The row and column headers are still dominant (2) (indicated by the red circle in figure 5.7), and the thick black lines separating the sections makes it clearer which titles belongs to which columns (4).

The following changes were related to the matrix structure:
By replacing the third dimension functionality by the summarized column (5) (blue circle in figure 5.7, further described on the next page), there is still a reduced number of axis in the matrix (9), and the minimalistic use of symbols (6) still eliminates the need to scroll down for any symbol description (7) while providing more white space in the structure (9).

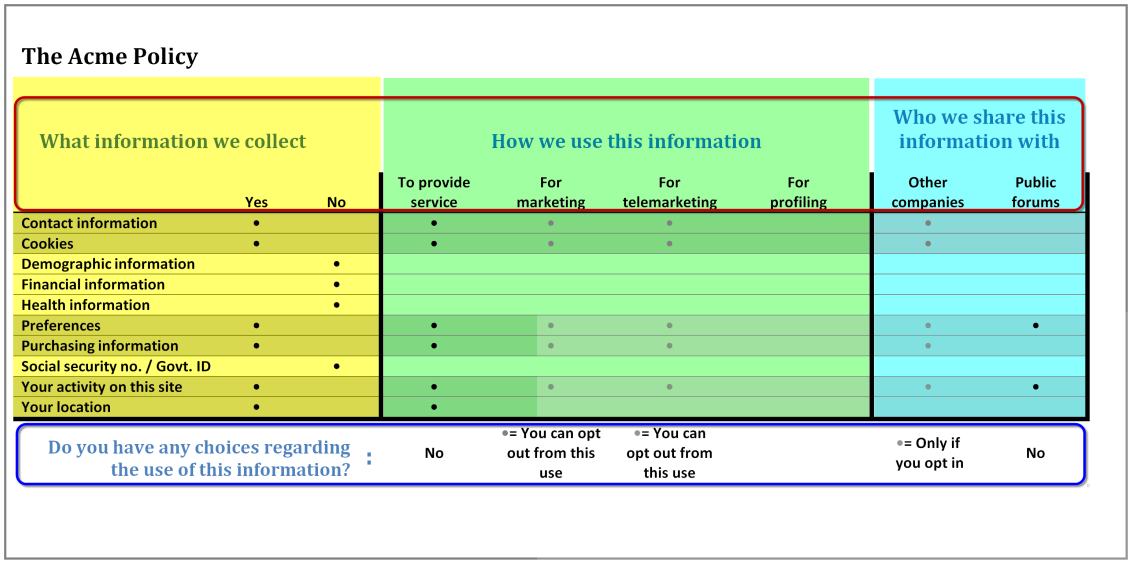


Figure 5.7

By reintroducing a modified version of the Nutrition Label "third dimension" we did however add an extra axis to the design. While this added extra complexion to the label, we felt it was necessary to provide a more detailed description for the privacy practices of each information type. The extra row as indicated by the blue circle in figure 5.7, can be seen on as a direct implementation of the summarized column as explained in section 4.2.2 (page 43) and seen in figure 5.8.

To indicate choices regarding the use of the collected information, a light grey dot-symbol was introduced. In order to keep the overall design as simple as possible, this symbol was used for both opt-in and opt-out cases.

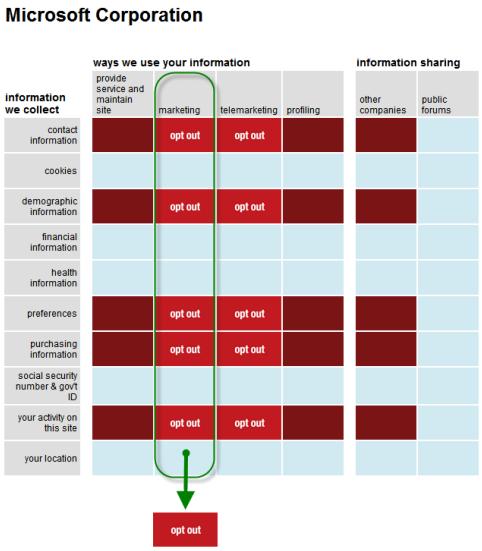


Figure 5.8

5.5 Laboratory experiment

The following section describes the laboratory experiment in detail, starting with a introduction of the polices used, followed by a description of each section in the questionnaire. The experiment procedure and instructions given to the participants are then explained in detail, before ending the section with a short evaluation of the experiment outcome.

The final questionnaire was based on the version used in the pre-test (section 5.4 on page 5.4) with implemented changes from the findings outlined in section 5.4.5 (page 74). The overall questionnaire structure was kept the same, only the questions were modified from the pre-test as explained through the following subsections.

5.5.1 Policies used

The latest iteration of the design progress of the Privacy Table was a result of the findings from the pre-test, as explained in the preceding section. The version used in the laboratory experiment can be seen in figure 5.10 with the corresponding Nutrition Label version in figure 5.9. Full size versions of the policies can be found in Appendix D.2.

The content of these policies were the same as for the policies used in the pre-test, with one change: In this policy Acme collects the current location by default (for providing service and maintaining site) rather than the social security number and gov't ID. This was changed due to uncertainty regarding the knowledge of the terms "SSN" and "gov't ID" which are not commonly used in the country where the experiment was taking place. The Nutrition Label was also upgraded to the latest version, where the column "research development" had been removed.

For part 1 we presented the participants with the policy as seen in figureres 5.9 and 5.10. As in the pre-test, both policies were identical in terms of content, only the format differed.

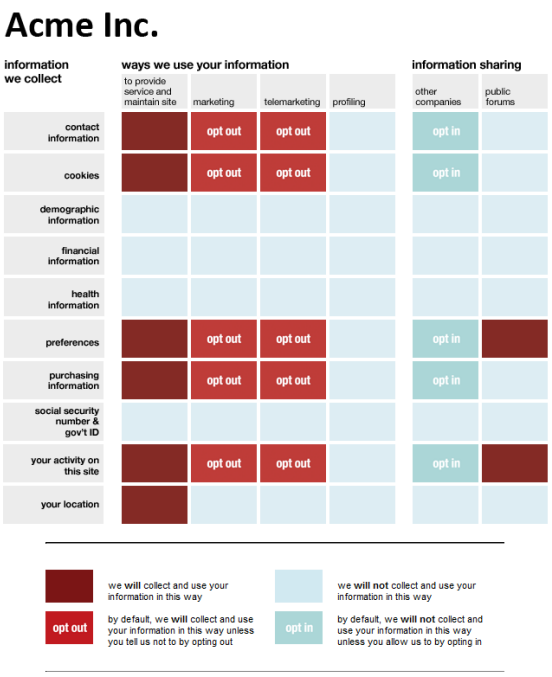


Figure 5.9: The Nutrition Label for part 1 in the laboratory experiment

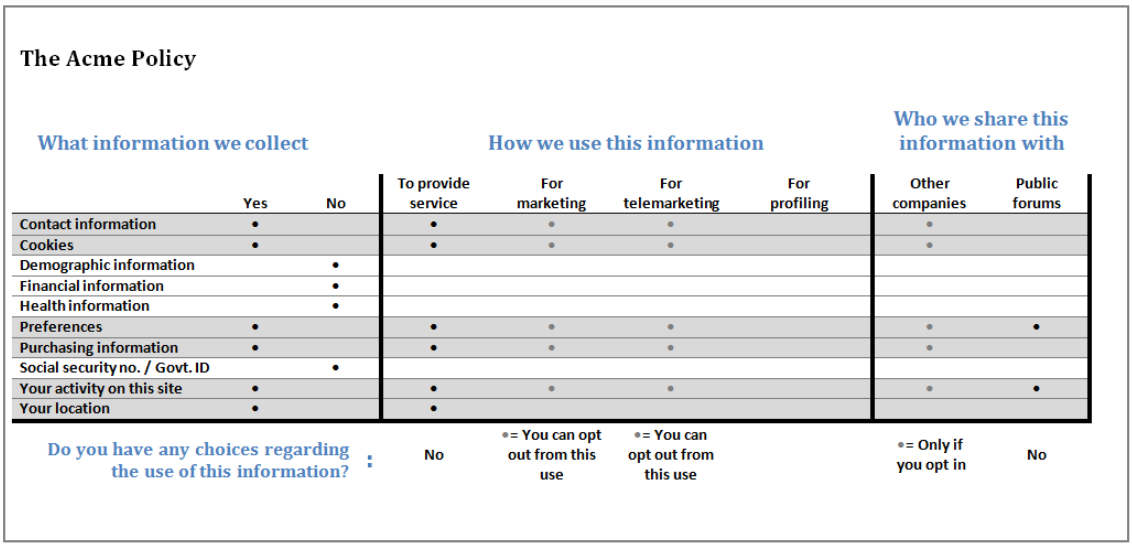
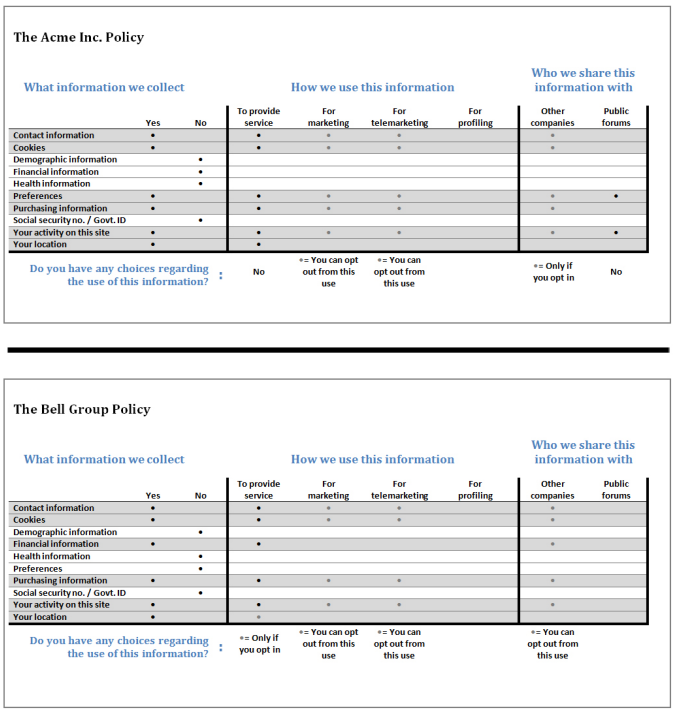


Figure 5.10: The Privacy Table for part 1 in the laboratory experiment

Part 2 of the experiment required the participant to answer questions by comparing two policies of the same format. We reused the 'Acme Inc.' policy from part 1, and introduced a second policy from the fictive company 'Bell group' which was based on the Acme policy and modified to match the comparison questions in part 2.1.

To ease the comparison task we printed both policies on a single sheet of paper, with a thick black line separating the two as seen in figures 5.11 and 5.12. Due to the different format of the two labels, the Nutrition Labels were aligned vertically and the Privacy Tables aligned horizontally.

Full size versions of the policies can be found in appendix D.2.



The Acme Inc. Policy

What information we collect			How we use this information				Who we share this information with	
	Yes	No	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	•		•	*	*		*	
Cookies	•		•	*	*		*	
Demographic information		•						
Financial information		•						
Health information		•						
Preferences	•		•	*	*		*	•
Purchasing information	•		•	*	*		*	
Social security no. / Govt. ID		•						
Your activity on this site	•		•	*	*		*	•
Your location	•		•					
Do you have any choices regarding the use of this information?	:		No	*= You can opt out from this use	*= You can opt out from this use		*= Only if you opt in	No

The Bell Group Policy

What information we collect			How we use this information				Who we share this information with	
	Yes	No	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	•		•	*	*		*	
Cookies	•		•	*	*		*	
Demographic information		•						
Financial information	•		•				*	
Health information		•						
Preferences		•						
Purchasing information	•		•	*	*		*	
Social security no. / Govt. ID		•						
Your activity on this site	•		•	*	*		*	
Your location	•		*					
Do you have any choices regarding the use of this information?	:		*= Only if you opt in	*= You can opt out from this use	*= You can opt out from this use		*= You can opt out from this use	

Figure 5.11: The Privacy Tables for part 2 in the laboratory experiment

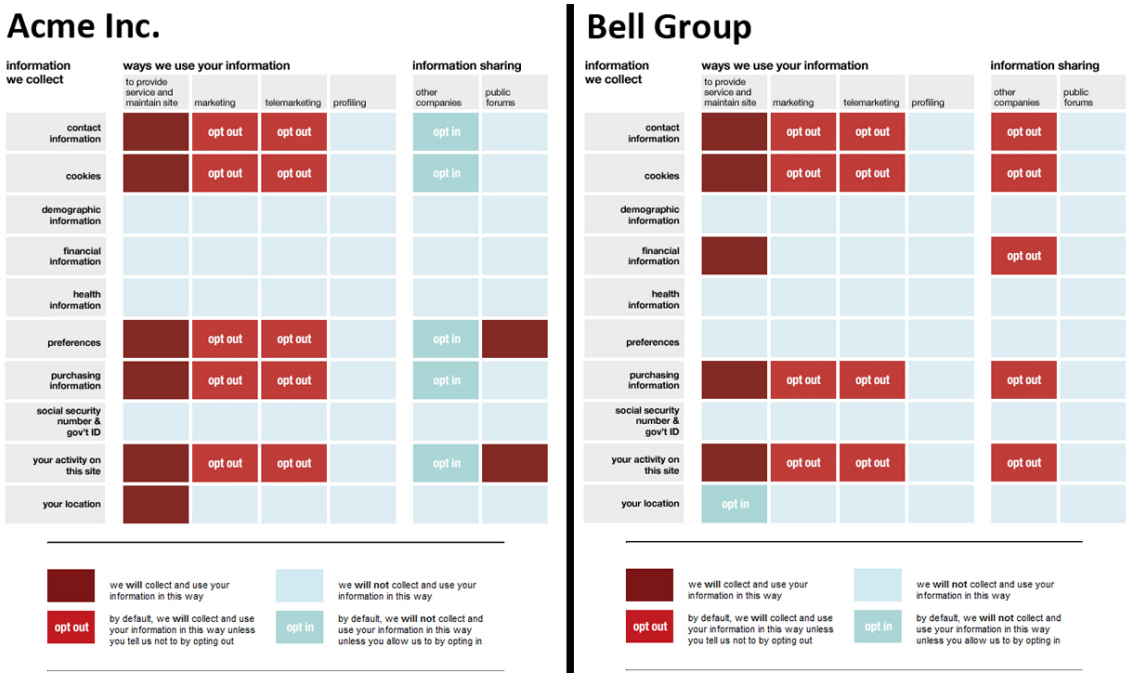


Figure 5.12: The Nutrition Labels for part 2 in the laboratory experiment

5.5.2 Questionnaire

The following subsections describe each questionnaire section in detail. Each sections gives an overview of which questions were reused from previous studies, which changes were done from the pre-test and which additional questions were included. For a full overview of the retrieved questions from previous studies and the subsequent use in this report, we again refer to table B.2 and B.3 in Appendix B.

As the participants in the pre-test felt they had to write down the time too many times, the amount of time-boxes was reduced. In the laboratory experiment questionnaire, combined time was therefore measured for part 1.1 and 1.2, while part 2.1 was kept separate. Time was not measured for part 1.3 and 2.2 as these parts consisted of likeability-questions.

Figure 5.13 gives an overview of the symbols used to indicate the use of a question throughout the preceding sections. A green icon indicates that the question was reused, an orange icon that a question was modified and reused while a red icon signifies that the question was excluded from our study.

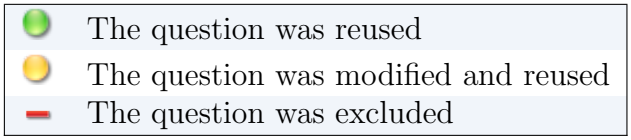


Figure 5.13: Icon legend

The questions in the information finding parts could be answered by "yes", "no", "Does not say" or "I don't know". We included the answer options 'Does not say', in case the respondents were unable to find the answer, and 'I don't know', in case the respondent did not understand the question.

According to Krosnick [1999], by introducing an "I don't know" option we might have provided a safe answer for participants not willing to spend any additional effort necessary to consider a different option stand. We still felt it was necessary to include such an alternative in order to avoid participants guessing or answering incorrectly in case they did not understand the terminology used in the question.

5.5.2.1 Part 1.1 - Single policy simple questions

Part 1.1 consisted of six simple information finding questions, designed to act as warm-up tasks to make the respondent familiar with the policy format.

In this part, we reused question 1, 2 (modified), 3 and 6, and excluded question 4 and 5 from the Kelley et al. [2010] study (see table 5.12). Two changes were done from the pre-test: the format of question 2 was modified as we wanted all questions in this part on a similar format, and question 5 regarding cookies was excluded. This was done as some participants in our pre-test were unsure whether non-native English speakers would understand the meaning of "storing cookies" and relate this

to any sharing practice in the policies.

In replacement for the excluded questions from the Kelley et al. [2010] study, two additional simple information finding tasks were added. The first asked whether "the policy allows Acme to collect information regarding household income?" and the second whether "the policy allows Acme to use information about gender for marketing purposes".

Kelley et al. [2010] - Part 3 - Single policy simple tasks			
	Question	**Pretest/ Final**	**Note**
1.	Does the policy allow Acme to collect information about which pages you visited on this web site?	🟢/🟢	
2.	Acme might want to use your information to improve their website. Does this policy allow them to use your information to do so?	🟢/🟡	Modified to: "Does the policy allow Acme to use your information to improve their website?" as we wanted all questions in this part on a similar format
3.	Does the policy allow Acme to collect information about your current location?	🟢/🟢	
4.	Based on the policy will Acme register their secure certificate with VeriSign or some other company?	➖/➖	Excluded before the pre-test as we did not include any privacy seal icons in the policy design
5.	Based on the policy may Acme store cookies on your computer? eeeeeeeee	🟢/➖	Excluded as some participants in our pre-test were unsure whether non-native English speakers would understand the meaning of "storing cookies".
6.	Does the policy allow Acme to collect information about your medical conditions, drug prescriptions, or family health history?	🟢/🟢	This question changed position with question 2, as we wanted the first three questions to be the simplest to answer
Kelley et al. [2009] - Part 1 - Information finding questions			
	Question	**Pretest/ Final**	**Note**
1.	Does the policy allow the Acme website to use cookies?	➖/➖	See question 5 above.
🟢 = Reused		🟡 = Modified & reused	➖ = Excluded

Table 5.12: Related single policy simple information finding questions

The final version of part 1.1 with correct answers marked in green is given in figure 5.14. Please note that the question numbers in figure 5.14 does not correspond to those in table 5.12.

Question 1, 2, 3 and 5 (in figure 5.14) were formed as single-element questions, asking whether Acme collected a type of information or not. These questions did not take into account how the information would be used, with whom it will be shared or whether one could opt out or in of this use. Question 4 and 6 were formed different as they asked whether or not Acme could use certain information in a specific way, but they did not specifically mention that this information was collected by Acme. These questions can therefore be regarded as more advanced than question 1, 2, 3 and 5.

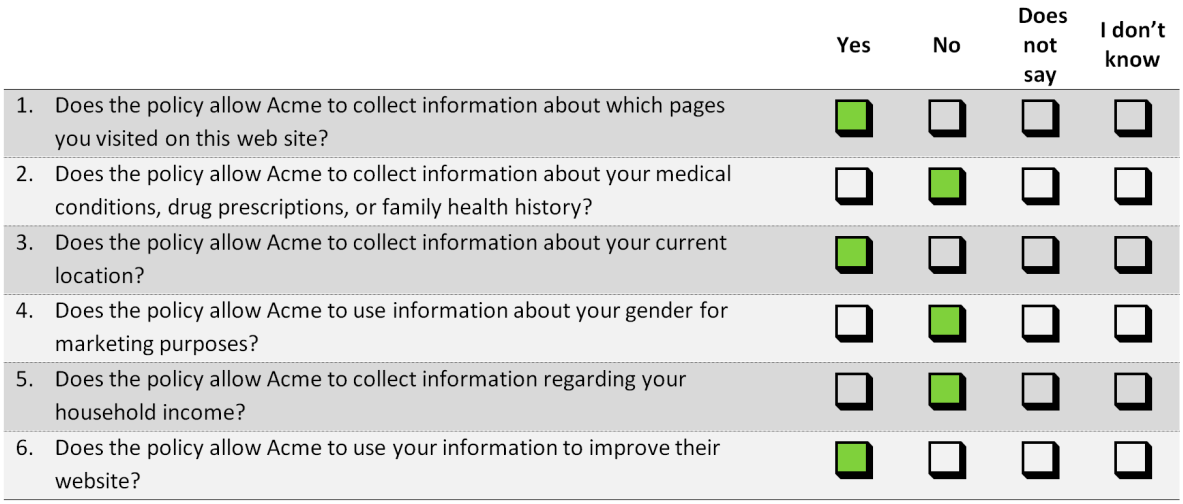


Figure 5.14: Part 1.1 of our questionnaire

As the overall goal of the design changes regarding the surrounding structure of the policy format was to improve the first impression of the label, and to aid in understanding what information was collected, we expected the Privacy Table to perform slightly better than the Nutrition Label on these questions in terms of accuracy and time.

5.5.2.2 Part 1.2 - Single policy complex tasks

In the single policy complex task part, all questions from the study by Kelley et al. [2010] were reused. One change was done from the pre-test: As we were unsure how well known the term "Bulletin boards" was, question 7 was modified to "public forums". In addition, "home phone number" was changed to "mobile number" in question 8 before the pre-test.

Kelley et al. [2010] - Part 4 - Single policy complex tasks			
	Question	**Pretest/ Final**	**Note**
7.	Does the policy allow Acme to share some of your information on public bulletin boards?	🟢/🟡	We were unsure how well known the term "Bulletin boards" is in 2011, and therefore changed it to "public forums".
8.	Does the policy allow Acme to share your home phone number with other companies?	🟡/🟡	"Home phone number" was changed to "mobile number"
9.	Does the policy allow Acme to use your buying history to design custom functionality targeted at you?	🟢/🟢	
10.	Does the policy allow Acme to share your cookie information with other companies?	🟢/🟢	
11.	Will Acme contact you with advertisements? eee eeee eeeee eeeee eeeee	🟢/🟢	
12.	Does Acme give you control regarding their sharing of your personal data?	🟢/🟢	
Kelley et al. [2009] - Part 1 - Information finding questions			
	Question	**Pretest/ Final**	**Note**
2.	Does the policy allow the Acme website to share your information on public bulletin boards?	🟢/🟡	See question 7 above.
3.	By default, does the policy allow the Acme website to collect your email address and use it for marketing?	➖/➖	Excluded before the pre-test as we only wanted six questions in each part.
🟢 = Reused	🟡 = Modified & reused	➖ = Excluded	

Table 5.13: Related single policy complex information finding questions

Question 10 from the Kelley et al. [2010] study, regarding cookies, was included as it directly asked whether cookies were collected and shared with other companies, and not whether Acme might "store cookies on your computer" as asked in question 5 in part 1.1 (which was excluded).

The questions in part 1.2 were of similar form as question 4 and 6 in part 1.1, but also included the answer options "Yes, unless I tell them not to" and "Only if I allow them to". Question 7, 8, 10, and 12 covered Acme's sharing of collected data, while

9 and 11 questioned Acme's own use of this data. The final version of part 1.2 with correct answers marked in green can be seen in figure 5.15.

	Yes	Yes, unless I tell them not to	Only if I allow them to	No	The policy does not say	I don't know
7. Does the policy allow Acme to share your personal information on public forums?	☐	☐	☐	■	☐	☐
8. Does the policy allow Acme to share your mobile number with other companies?	☐	☐	■	☐	☐	☐
9. Does the policy allow Acme to use your buying history to design custom functionality targeted at you?	☐	☐	☐	■	☐	☐
10. Does the policy allow Acme to share your cookie information with other companies?	☐	☐	■	☐	☐	☐
11. Will Acme contact you with advertisements?	☐	■	☐	☐	☐	☐
12. Does Acme give you control regarding their sharing of your personal data?	☐	☐	■	☐	☐	☐

Figure 5.15: Part 1.2 of our questionnaire

As we expected the respondent to be more familiar with the policy structure after having answered the simple questions, the questions in part 1.2 were more related to the design of the matrix itself than the surrounding structure. Given the reduced complexity in the Privacy Table, we expected the Nutrition Label to perform slightly better in terms of accuracy and time on the complex information finding questions.

5.5.2.3 Part 1.3 - Single policy likeability

Following the simple and complex information finding questions was a likeability part with the intention of measuring the enjoyability of using the format to answer single policy questions. As the paper by Kelley et al. [2010] did not publish any of its likeability-questions, we reused questions from the Kelley et al. [2009] study in this part.

As seen in table 5.14, three changes were made from the pre-test. One finding from the pre-test (section 5.4.5 on page 74) was that some participants did not like the word "enjoyable", and argued that reading a privacy policy can never be an enjoyable experience. As question 8 and 9 also covered the enjoyability of reading a policy, question 6 was excluded.

According to Kelley et al. [2009], question 7 in table 5.14 "investigates participants' perceived accuracy in the earlier questions". We did not find this relevant for the evaluation objectives, and therefore excluded this question as well. These two questions were replaced by two custom designed statements to better test the differences between the two formats.

Kelley et al. [2009] - Part 2 - Perceived Privacy Policy Understanding			
	Question	**Pretest/ Final**	**Note**
4	I feel secure about sharing my personal information with Acme after viewing their privacy practices	🟢/🟢	
5	I feel that Acmes privacy practices are explained thoroughly in the privacy policy I read	🟢/🟢	
6	Finding information in Acmes privacy policy was a pleasurable experience	🟢/➖	Excluded based on findings from the pre-study and as policy enjoyment was covered by statement 8 and 9
7	I feel confident in my understanding of what I read of Acmes privacy policy	🟢/➖	This statement "investigates participants perceived accuracy in the earlier questions" [Kelley et al., 2009]. We did not find this relevant for our evaluation objectives.
8	It was hard to find information in Acmes policy	🟢/🟡	Changed to "I had problems finding the information I was looking for in Acmes policy"
9	If all privacy policies looked just like this I would be more likely to read them	🟢/🟢	
🟢 = Reused		🟡 = Modified & reused	➖ = Excluded

Table 5.14: Related single policy likeability questions

The first additional question (question 16 in figure 5.16) was "When I first looked at Acme's policy, it was easy to understand what information they will collect from me", which attempted to capture the respondents first impression of the policy label. As mentioned previously, the first impression of the label was related to the design changes regarding the surrounding structure of the policy, and we therefore expected the Privacy Table to be more liked than the Nutrition Label on this question. We expected the same result for question 14 (figure 5.16), which asked whether the respondent had any problems finding information in the policy.

The second additional question (question 18 in figure 5.16) was "The use of symbols (e.g. icons, characters or colors) in Acmes policy made it confusing to understand how the information they collect will be used", which tried to capture whether the decreased size and amount of symbols in the matrix on our label had any effect. We therefore expected a higher likeability score for the Nutrition Label group on this question, which would indicate that the respondent would prefer fewer icons in

the format. In order to make the question valid for both formats, we specifically mentioned various examples of different types of symbols (icons, characters and colors).

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree
13. I feel that Acme's privacy practices are explained thoroughly in the privacy policy I read	❑	❑	❑	❑	❑
14. I had problems finding the information I was looking for in Acmes policy	❑	❑	❑	❑	❑
15. I feel secure about sharing my personal information with Acme after viewing their privacy practices	❑	❑	❑	❑	❑
16. When I first looked at Acme's policy, it was easy to understand what information they will collect from me	❑	❑	❑	❑	❑
17. If all privacy policies looked just like this I would be more likely to read them	❑	❑	❑	❑	❑
18. The use of symbols (e.g. icons, characters or colors) in Acmes policy made it confusing to understand how the information they collect will be used	❑	❑	❑	❑	❑

Figure 5.16: Part 1.3 of our questionnaire

The final version of part 1.3 can be seen in figure 5.16, and table 5.15 summarizes the purpose of each statement in this part. The description of the reused statements 13, 14, 15 and 17 were taken from Kelley et al. [2009], while the purpose of statement 16 and 18 was described above.

13.	Questions whether participants believe their practices are well displayed
14.	Rates the difficulty the participant had in finding information
15.	Captures participants reaction to the actual content of the privacy policy they read
16.	Captures the participants first impression of the label
17.	Captures whether the proposed label would encourage more people to read policies
18.	Captures whether the use of symbols aids or confuses the participants in finding information

Table 5.15: Summary of the purpose of each statement in part 1.3

With the exception of question 18, we expected the Privacy Table to be averagely more liked than the Nutrition Label in this part, due to the simplified design, which had the purpose of improving the first impression of the policy and making it easier to use for novice users.

5.5.2.4 Part 2.1 - Policy comparison tasks

In part 2.2 we gave the participants five true/false-statements regarding the privacy practices for Acme and Bell. Only four relevant statements were published from previous studies as indicated in table 5.16, where two were modified and one was excluded after the pre-test.

Kelley et al. [2010] - Part 6 - Policy Comparison tasks			
	Question	**Pretest/ Final**	**Note**
14.	Does either company give you options with regards to cookies?	🟢/ ➖	Excluded as we found the statement"Does either company give you options" diffuse.
15.	Does either company collect sensitive information (such as banking or medical records)?	🟢/ 🟡	Modified to a true/false statement and added "passport number" as an example: "Neither Acme nor Bell collects any sensitive information (such as banking or medical records, or passport numbers)"
16.	By default, Acme can collect information about your age and gender in order to market to you by email, but the Bell Group cannot.	🟢/ 🟡	"Age and gender" was changed to "your use of their website"
Kelley et al. [2009] - Part 3 - Policy Comparison Questions			
	Question	**Use**	**Note**
10	By default, Button Co. can share information about your purchases with other companies, but Acme cannot.	🟢/ 🟢	
11	Which company will better protect your information online? eee eeee eeee eee	➖/ ➖	Excluded before the pre-test as this was not a true/false statement, and has no correct answer
12	Youre looking to buy a gift online. At which company would you prefer to shop?	➖/ ➖	Another preference question which was excluded before the pre-test for the same reasons as question 11
🟢 = Reused	🟡 = Modified & reused	➖ = Excluded	

Table 5.16: Related policy comparison tasks

This part originally consisted of a mix of true/false-statements and information finding questions, but based on the feedback from the pre-test all questions in this part

was modified to true/false-statements. Question 15 (in table 5.16) from the Kelley et al. [2010] study was therefore changed to "Neither Acme nor Bell collects any sensitive information (such as banking or medical records, or passport numbers)". "Passport numbers" was also added as an example of sensitive information in this statement.

Question 14 (in table 5.16) from the Kelley et al. [2010] study was excluded as we found it too unspecific. The "Use of cookies" could signify both internal company use and external sharing, and the statement would have been both true and false for the Acme and Bell policies (as both companies could share cookie information to other companies if opted in, but not on public forums). The remaining questions (11 and 12) in table 5.16 did not fit the true/false-statement format and was thus excluded.

Two additional questions were added to this part. The first (question 19 in figure 5.17) was "Both Acme and Bell lets you control whether non-sensitive information (such as name and address) is shared with other companies" and the second (question 23 in figure 5.17) was "By default, both companies collects your default location and can use this information to improve their services". An overview of all the questions with the corresponding correct answer marked in green can be seen in figure 5.17.

	True	False	Does not say	I don't know
19. Both Acme and Bell lets you control whether non-sensitive information (such as name and address) is shared with other companies	■	☐	☐	☐
20. Neither Acme nor Bell collects any sensitive information (such as banking or medical records, or passport numbers)	☐	■	☐	☐
21. By default, Acme can collect information about your use of their website in order to market to you by email, but Bell cannot.	☐	■	☐	☐
22. By default, Bell can share your history of purchased items with other companies, but Acme cannot	■	☐	☐	☐
23. By default, both companies collects your default location and can use this information to improve their services	☐	■	☐	☐

Figure 5.17: Part 2.1 of our questionnaire

As for part 1.3, we expected the Nutrition Label to perform slightly better in terms of accuracy and time on the comparison questions in part 2.1. As for the complex information finding questions, the reduction in complexity in the Privacy Table compared to the Nutrition Label, could favor the latter when asked complex questions.

5.5.2.5 Part 2.2 - Policy comparison likeability

As for part 1, we also wanted to measure the respondent's enjoyment of the comparison experience in part 2. Few related policy comparison likeability questions were published in the previous studies, and two of the ones that were published (question

13 and 14 in table 5.17) from the Kelley et al. [2009] study were excluded as they were single policy likeability questions.

Question 15 and 16 in table 5.17 from the Kelley et al. [2009] study were both similar and positively toned ("enjoyable experience" vs. "easy to do"). As one of the participants in the pre-test commented that reading a policy could never be an enjoyable experience, and suggested using a word like "ok" instead, we merged question 15 and 16 to a single question asking "I found comparing two policies an easy task to do".

Kelley et al. [2009] - Part 4 - Policy Comparison Enjoyment & Ease			
	Question	**Pretest/ Final**	**Note**
13	Looking at policies to find information was an enjoyable experience	– / –	Excluded before the pre-test as it was a single policy likeability question
14	Looking at policies to find information was easy to do	– / –	Excluded before the pre-test as it was a single policy likeability question
15	Comparing two policies was an enjoyable experience	● / –	Merged with question 16 due to its similarities.
16	Comparing two policies was easy to do	● / ●	Merged with question 15 and modified to "I found comparing two policies an easy task to do"
● = Reused		● = Modified & reused	– = Excluded

Table 5.17: Related policy comparison likeability questions

Four additional likeability-questions (question 25-28) was therefore added to complement this part, as seen in figure 5.18. Question 25, "*If all policies looked like this I would compare privacy practices across websites more often*" was added before the pre-test. This question was similar to question 17 in part 1.3 (see figure 5.16 on page 87), and tested whether the policy format would motivate for a more frequent comparison of privacy practices across sites.

Question 26 and 28 in figure 5.18 were specifically designed to test what effect the reduced use of symbols and colors would have on comparison enjoyment (suggested changes to the matrix structure). Question 26, "*It was easy to keep control of what the different symbols (e.g. icons, characters or colors) in the policy stood for*", tested whether it was easy to keep control of the symbols in the policies during the comparison experience. Question 28, "I would feel more confident in understanding the privacy practices of Acme or Bell if the policies had more text and less symbols (e.g. icons, characters or colors)", tested whether the respondent would prefer more text and less symbols in the policy.

Question 27, "*I feel that these policies would cover most of the privacy concerns I*

would have if I were to sign up for Acme or Bell in real life", was a general question that tested whether the respondent felt that the content of either format covered his or hers privacy concerns if he/she was to sign up for Acme or Bell in real life.

The final version of part 2.2 can be seen in figure 5.18.

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
24. I found comparing two policies an easy task to do	❑	❑	❑	❑	❑
25. If all policies looked like this I would compare privacy practices across websites more often	❑	❑	❑	❑	❑
26. It was easy to keep control of what the different symbols (e.g. icons, characters or colors) in the policy stood for	❑	❑	❑	❑	❑
27. I feel that these policies would cover most of the privacy concerns I would have if I were to sign up for Acme or Bell in real life	❑	❑	❑	❑	❑
28. I would feel more confident in understanding the privacy practices of Acme or Bell if the policies had more text and less symbols (e.g. icons, characters or colors).	❑	❑	❑	❑	❑

Figure 5.18: Part 2.2 of our questionnaire

Given we expected the Nutrition Label to perform slightly better in terms of accuracy and time on the comparison questions in part 2.1, due to the Label being better suited for complex questions, we also expected it to be slightly more liked than the Privacy Table.

5.5.2.6 Open-ended questions

Ending the questionnaire was the two open-ended questions designed to capture any additional opinions regarding the policy format and the questionnaire in general. The first was "Do you have any comments or suggestions regarding the privacy policy format?" and the second was "Do you have any feedback regarding the format of this survey?".

5.5.3 Procedure and instructions

As the experiment planning has been given in details through the preceeding parts, this section will give an brief overview of the experiment day procedure as well as the instructions given to the participants.

Figure 5.19: The experiment location

As outlined in section 5.2.4 on page 61, the laboratory experiment was conducted as a part of a guest lecture held by Karin Bernsmed from Sintef ICT for students taking the course TTM4135 Information Security at the Norwegian University of Science and Technology (NTNU), on April 1st 2011. The topic for the lecture was to give a brief introduction to privacy online, personal data management and privacy agents (as previously mentioned, the Nutrition Label and the Privacy Table was not mentioned). The guest lecture lasted for approximately 45 minutes, and was (together with the experiment) held in auditorium EL3 as seen in figure 5.19. PowerPoint slides from the guest lecture can be found in Appendix D.3, and the slides from the experiment presentation are given throughout this section.

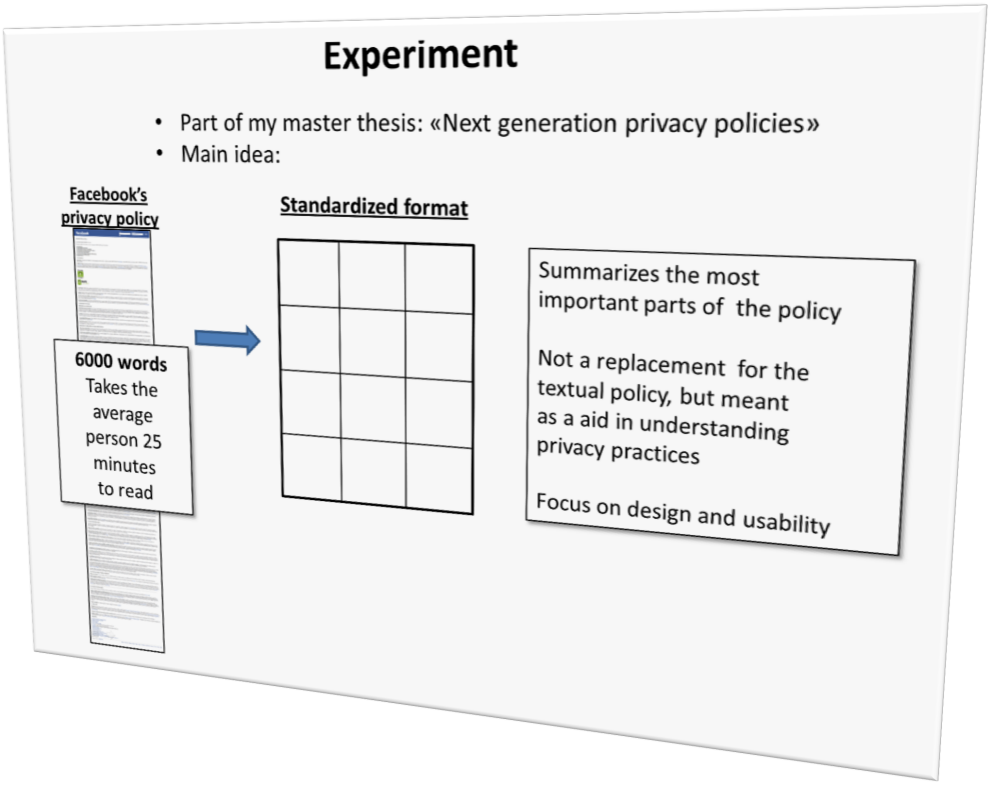


Figure 5.20: The first PowerPoint slide for the experiment presentation

The experiment was announced at the beginning of the lecture, and as explained in section 5.2.4 on page 61, the students were made aware of the importance of participating in the experiment and told that participating was volunteer. To minimize dropouts (section 5.2.7.1, on page 66) they were also told that the experiment would last for 15-20 minutes and that they afterwards were free to leave.

The 15 minute break between the guest lecture and the experiment was spent to prepare the questionnaires and the PowerPoint presentation. As the students arrived back to the auditorium they were first briefed about the thesis and experiment background, as seen in the PowerPoint slide in figure 5.20. This was only a brief explanation of the thesis topic and did not include any presentation of the Nutrition Label nor the Privacy Table.

The students were then instructed of the experiment procedure, as seen in the PowerPoint slide in figure 5.21. They were told that the goal of the experiment was to test two types of standardized policy design, and that they would see either format A or B. They were not told which design was the Nutrition Label and which was our design, to avoid anyone "helping the researcher" as explained in section 5.2.7.1 on page 66. They were also instructed to use the policies actively while they were answering the questions, to avoid a similar situation to what happened in the pretest, where one participant was unsure whether he had to memorize the policy (section 5.4.5, page 74.)

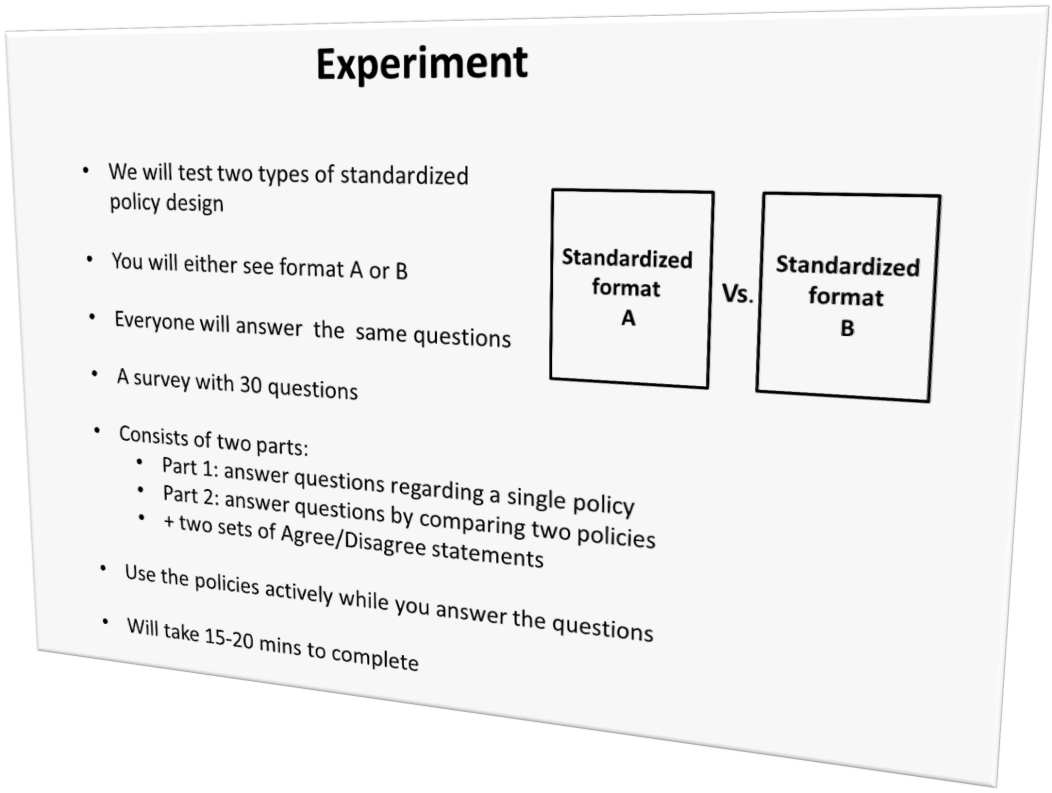


Figure 5.21: The second PowerPoint slide for the experiment presentation

As seen in figure 5.22, they were also told that they could answer "I don't know" in case they did not understand the question. We also explained that no personal information would be collected, and finally we repeated that participating was voluntary and that they could abandon the experiment at any time.

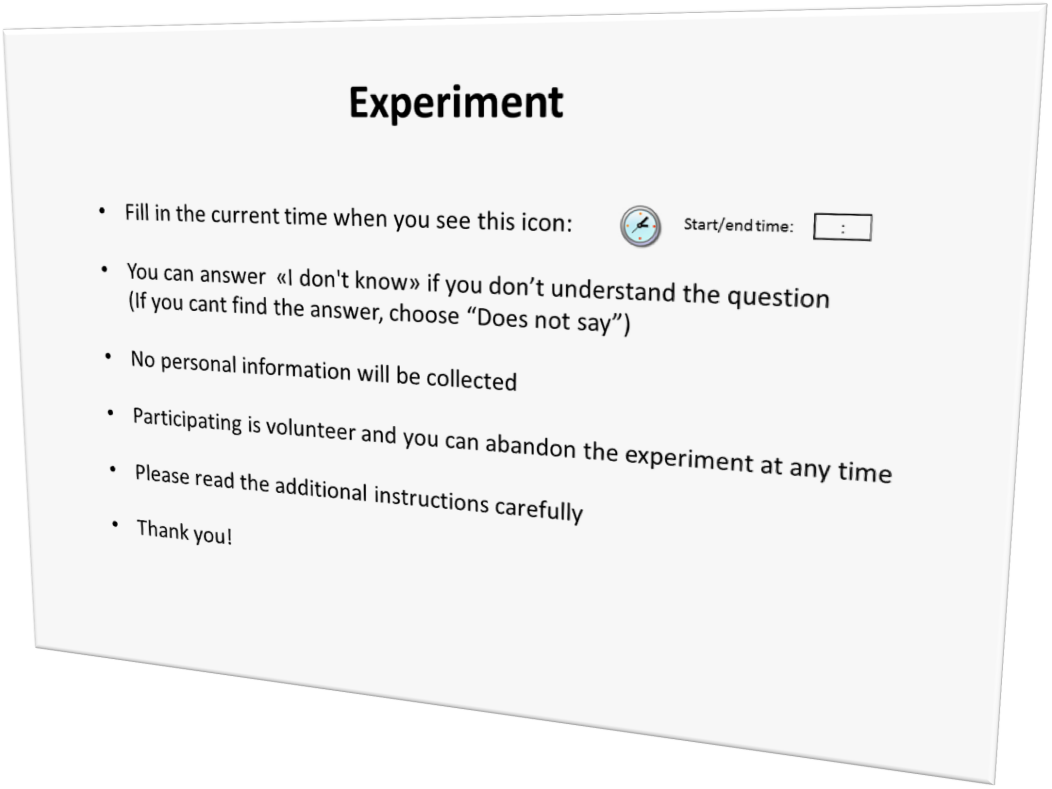


Figure 5.22: The third PowerPoint slide for the experiment presentation

The questionnaires were then handed out to the students. We used the randomization method as described in section 5.2.6 on page 63, and divided the auditorium by half: one side received the Nutrition Label version and the other half received the Privacy Table version.

5.5.4 Summary

While the execution of the laboratory experiment went as planned, its outcome suffered from a disappointingly low number of students showing up for the guest lecture. While 74 students were registered to the course, only 15 students turned up and completed the experiment.

There can be several reasons for this low attendance. First of all, it is common that some courses has a low attendance, and combined with this being a guest lecture not included in the course curriculum, several students could have regarded it as a "day off". The lecture time (08:15 on a Friday) could also have affected the attendance.

Given the low attendance we decided to conduct the experiment a second time: either as another laboratory experiment or as an internet survey. Common for both were that we had to offer some sort of incentive such as a prize or a payment in order to attract participants. Based on a quick evaluation of time required and possible recruitment strategies, we decided to go for an Internet experiment. This way, we could make the questionnaire available online, spread the participation link and wait for responses rather than organizing another time-consuming lab-experiment. We also believed the online recruitment and experiment method would result in a higher number of participants than recruiting through offline channels (such as posters or announcements).

5.6 Internet experiment

As explained on the previous page, we decided to conduct an additional Internet experiment due to the low number of participants in the laboratory experiment. The following section describes the Internet experiment in detail, starting with the participant recruitment and followed by the experiment procedure. The changes made to the policies and the questionnaire is then described, before ending the section with a short summary of the experiment outcome.

5.6.1 Participant recruitment and instructions

As previously mentioned, the recruitment method for the Internet experiment was to randomly distribute a link on a social networking site. While this would result in our results not being generalizable to any population (as discussed in section 5.2.3 on page 59), it could instead give us some idea of how a more varied user group would handle the policies. The findings from the Internet experiment must therefore be regarded as indications rather than statistically valid results.

A facebook event as seen in figure 5.24 was created to recruit participants to the experiment. The event description consisted of a brief description of the purpose and content of the experiment.

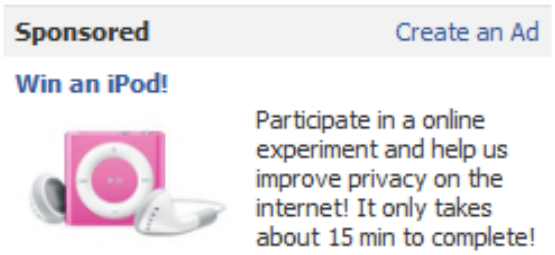


Figure 5.23: The facebook ad

As an incentive to participate, we offered anyone that completed the questionnaire to enter a prize draw for winning a iPod shuffle (2GB), worth approxiamitly USD70 . The participant entered the draw by entering his/hers email address upon completing the survey, but the address was not

connected to the questionnaire answers. A random result generator was used to draw the winner [1]. A Facebook ad as seen in figure 5.23 (funded by a free gift card) was also created, and was displayed 50,981 times and generated 54 visits to the event page.

Figure 5.24: The facebook event page

By clicking the link in figure 5.24, the participant was forwarded to either the questionnaire containing the Nutrition Label or the Privacy Table, where the first page provided the following instructions:

This experiment is a part of a Master Thesis study being conducted at the Norwegian University of Science and Technology, in cooperation with SINTEF ICT. The goal of the study is to explore alternative ways of presenting privacy policies, and in this experiment two new approaches will be tested. You will be assigned to one of these two, and your task is to answer four sets of questions regarding policies in this format.

It should take approximately 15-20 minutes to complete the experiment and no per-

[1] http://www.randomresult.com/

sonal information will be collected. All your answers will be treated confidentially and anonymously, and by completing this survey you accept that your answers will be used for the purpose of answering the research questions in the mentioned thesis.

By completing the experiment you will be eligible to enter a draw for an iPod shuffle (2 GB). In order to participate in the draw, you will need to leave your email address on the final page. Your email address will not be connected to your experiment answers and will only be used for contacting the prize winner.

Please read the detailed instructions for each page carefully.

Thank you for your contribution!

5.6.2 Survey-tool

To convert the paper-based survey to a online format, we utilized the service provided by Surveygizmo.com. Surveygizmo.com was chosen based on good previous experience and the fact that it offered the most customizable look, which enabled us to directly convert the original questionnaire to an online format. The online questionnaire was the same as for the laboratory and for full sized screenshots we refer to Appendix E.1.

Two similar questionnaires were created in Surveygizmo, with one showing the Nutrition Label and the other showing the Privacy Table. To randomly direct participants to one of the formats, the participation link on the Facebook event directed to a hosted script which randomly forwarded the browser to one of the Surveygizmo questionnaires.

5.6.3 The policies used

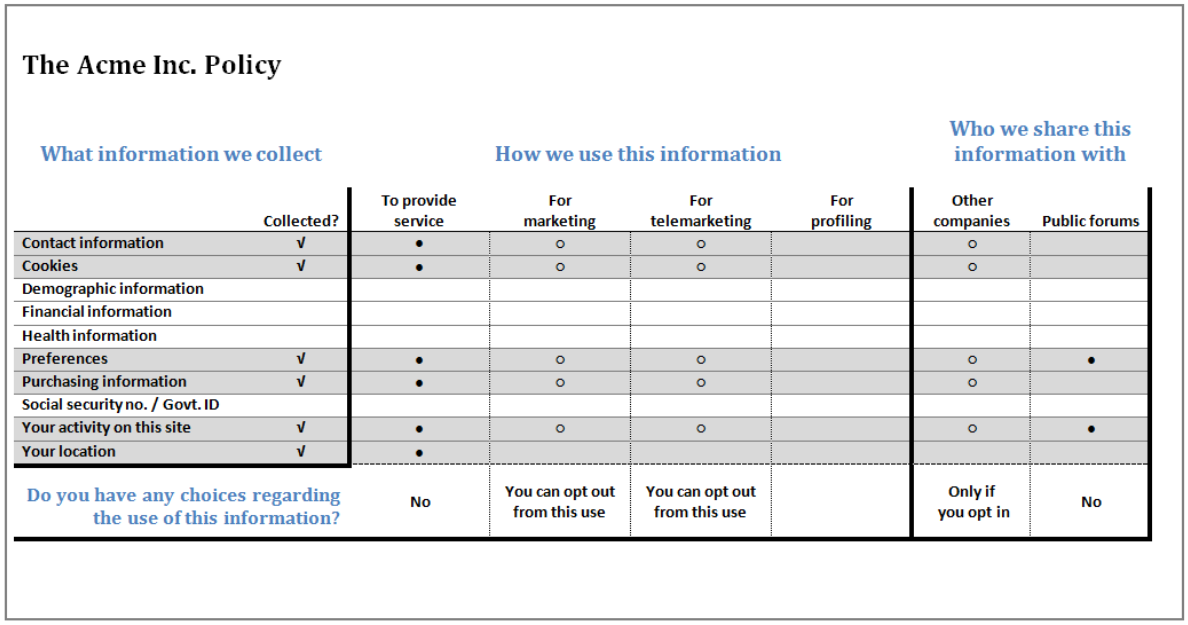
Based on the feedback we recieved in the open-end questions in the laboratory experiment, we modified the Privacy Table design. As seen in the list below, several participants had problems understanding the opt-out/in system and the different symbols.

- "It should be different symbols/colors on "only if opt-in" and "you can opt-out", and not just in the text below"
- "Better coloring of the dots, to distinguish at a glance between opt in/out, yes/no"
- "Very hard to understand the opt in/out choises. Hard to understand what was default settings in the two different firms. 'Didnt feel like the "choices" in the bottom had anything to do with each column. Hard to understand. This is because of the thick line where the table ends. Use multiple colors to explain default settings for "opt in / and opt out""

While some suggested using several symbols, or an opt-in/out elements in each square, we remained true to the original design criteria list (see table 5.1 on page 5.1). The following changes were done:

- The light grey symbol was replaced by a circle for a better distinction from the black dot.
- To better connect the opt-in/out choices with each column, the thick black line was moved to the bottom
- The yes/no for collected information was replaced by a single symbol to save space

The resulting design which was used in the Internet experiment can be seen in figure 5.25.



The Acme Inc. Policy

What information we collect		How we use this information				Who we share this information with	
	Collected?	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	√	•	○	○		○	
Cookies	√	•	○	○		○	
Demographic information							
Financial information							
Health information							
Preferences	√	•	○	○		○	•
Purchasing information	√	•	○	○		○	
Social security no. / Govt. ID							
Your activity on this site	√	•	○	○		○	•
Your location	√	•					
Do you have any choices regarding the use of this information?		No	You can opt out from this use	You can opt out from this use		Only if you opt in	No

Figure 5.25: The third version of the Privacy Table

5.6.3.1 The questionnaire

While no changes were done to the questionnaire itself, the internet format required us to do some modifications. Firstly, as we had little or no control over who answered the questionnaire, a demographic part asking for age, gender and education was added. Full screenshots of the internet questionnaire can be found in Appendix E.1.

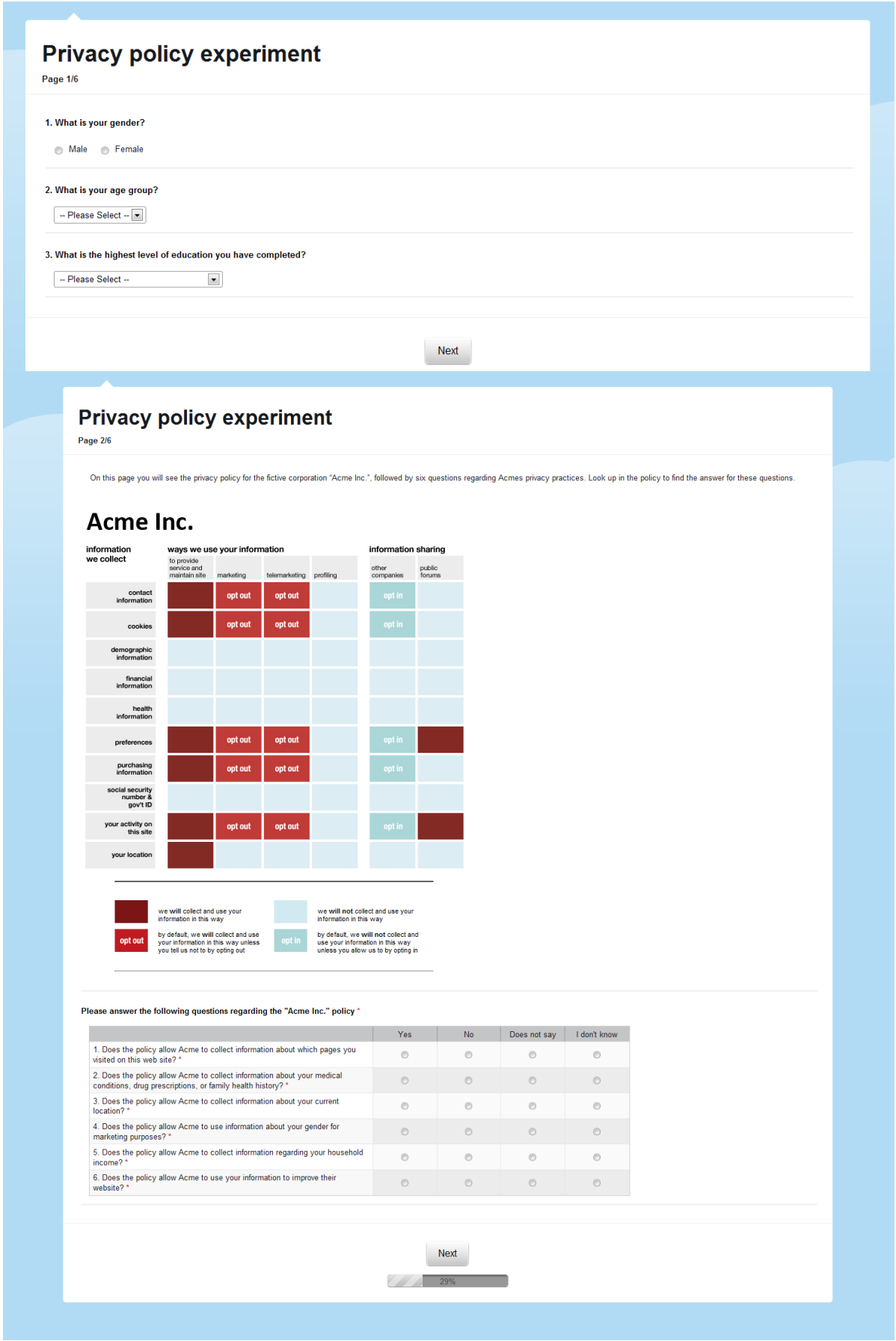


Figure 5.26: The demographic questions (top) and the single policy questions

Secondly, given that we had to present the questions and the policies together, part 1.1 and part 1.2 was split across two pages with the corresponding Acme policy above each section as seen in figure 5.26. In addition, the instruction text "*On this page you will see the privacy policy for the fictive corporation "Acme Inc.", followed by six questions regarding Acmes privacy practices. Look up in the policy to find the answer for these questions.*" was added to aid the participants.

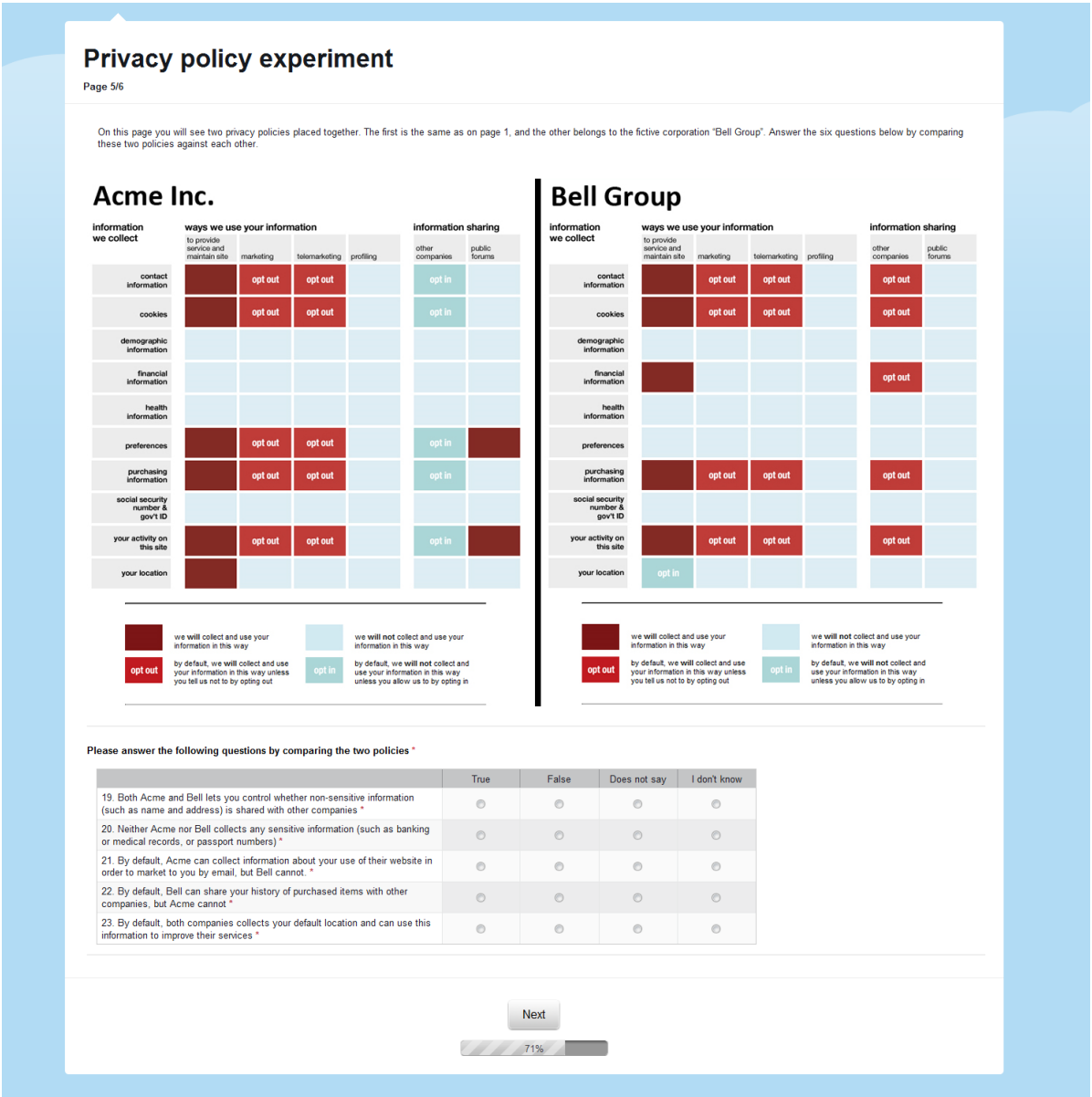


Figure 5.27: The comparison policy questions

The comparison part (figure 5.27) were of similar form as the single policy questions, with the instruction "On this page you will see two privacy policies placed together. The first is the same as on page 1, and the other belongs to the fictive corporation Bell Group. Answer the six questions below by comparing these two policies against each other." As seen in Appendix E.1, the likeability sections were placed on separate pages.

CHAPTER 6

Evaluation results

The following chapter presents the results from the evaluations. We begin the chapter by presenting the demographic information, and given the low number of participants in both experiments, it is followed by a discussion of the validity of the results. We then outline the results from the information finding questions, the likeability questions, the timing data and the open-end questions from both experiments. Detailed statistics regarding the results can be found in Appendix F. We end this chapter by a presentation of the key findings.

6.1 Participants and demographics

As mentioned in the summary of the laboratory experiment in section 5.5.4 (on page 94) , 15 students participated in the Laboratory experiment. The students were all taking the course TTM4135 Information Security, which was available for students at study level 3 (Third-year students); primarily from the study programs Telematics, Industrial economics and technology management, Communication Technologies and Computer Science.

For the Internet experiment, we received a total of 24 responses, with an equal distribution of 12 responses per each policy format. There was however a high number of dropouts (45) from the internet experiment, which was expected due to the challenging tasks (some drop-outs reported that they had expected to participate in an opinion-based survey, and not in a experiment which required them to answer information finding tasks). Neither of the drop-outs completed more than part 1.1 of the questionnaire and was therefore excluded from further analysis.

Demographic information for the Internet participants can be seen in figure 6.1. While there were a majority of male participants in the Nutrition Label group, there was a female dominance in the Privacy Table group. There were no major differences in age between the groups, with a majority of the participants belonging to the 18-29 age groups. 83.3% of the participants in both groups had completed either a bachelor's or a master's degree.

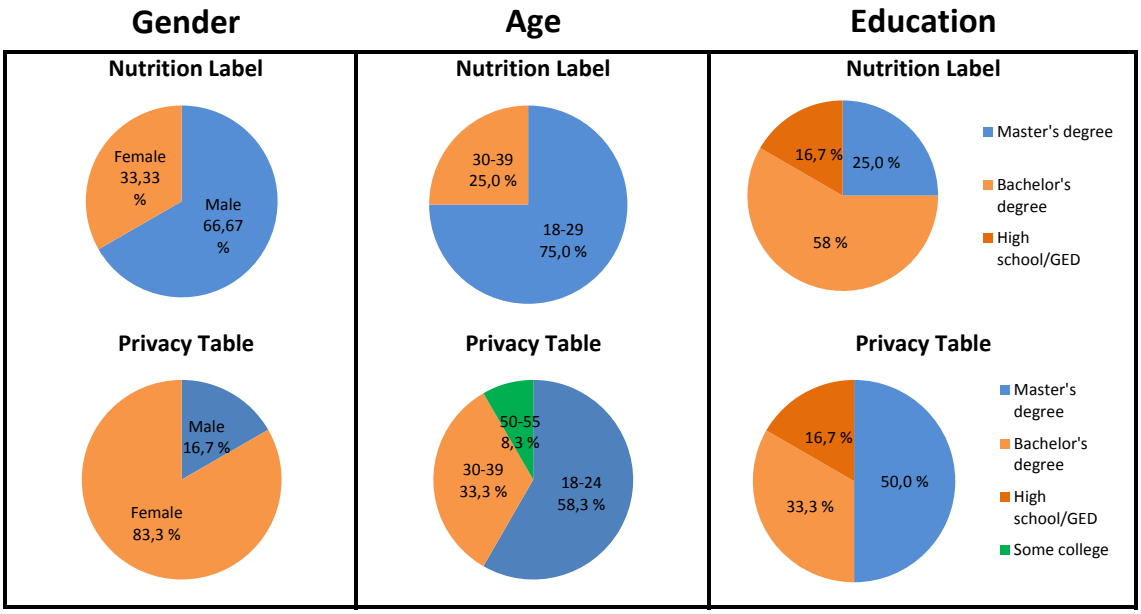


Figure 6.1: Demographic information for the Internet experiment

6.2 Validity of the results

As discussed in section 5.2.7 on page 65, a common threat to the external validity of an experiment is having too few participants. According to Oates [2006] a research project should have a sample size of at least 30 in order to conduct a statistical analysis and get reliable results. Given our final sample was 15 participants for the laboratory experiment, and 24 participants for the internet experiment, the validity of the subsequent statistical analysis is therefore threatened by the small sample sizes. Our results might therefore not be regarded as reliable or generalizable to any population.

The study by Kelley et al. [2009], from which we have reused several questions for the questionnaire, also had a small sample set consisting of 24 participants. While the experiment by Kelley et al. [2009] compared two rather different formats and also used a within-participant design, which could require less participants to get significant results, it indicates that even an analysis of a smaller data sample than 30 participants might be useful for such an experiment of this type.

Based on this we therefore decided to conduct a statistical analysis on the experiment results, assuming our results were generalizable to some degree. As we had two groups in each experiment, a natural choice of method for analyzing the results was to conduct two-sample t-tests. A t-test determines whether the mean from one group differs from the mean of another group, by comparing the mean difference between the two groups [CSPA, 2011]. According to Trochim [2006], a t-test is especially appropriate as the analysis for the post-test-only two-group randomized experimental design.

The t-test does however make two assumptions. Firstly it assumes normally distributed data and secondly it assumes equality of variance [CSPA, 2011]. Equality of variance requires that the two populations must have equal variances, and a normal distribution is defined as a probability distribution that plots all of its values in a symmetrical fashion and most of the results are situated around the probability's mean. Values are equally likely to plot either above or below the mean [Solutions, 2011].

According to Solutions [2011], indications for normally distributed data are a skewness (measure of asymmetry) within a +-1 range, and a kurtosis (measure of "peakness") within a +-3 range. It also exists several tools for testing whether data is normally distributed, and one of the most popular among researchers is the Shapiro-Wilk's W test [Analyseit, 2008]. The Shapiro-Wilk test tests the null hypothesis that a sample came from a normally distributed population [Analyseit, 2008]. If the p-value is significant (less than 0.10 or less than 0.05) the null hypothesis is rejected and one can conclude that the sample is not normally distributed [Analyseit, 2008]. With small sample sizes of 10 or fewer observations, it is however unlikely that the normality test will detect non-normality, and one should rely on non-parametric alternative.

	Information finding				**Likeability**			
	LAB		**INT**		**LAB**		**INT**	
	NL	**PT**	**NL**	**PT**	**NL**	**PT**	**NL**	**PT**
Mean	12,5	11,3	10,3	10,8	3,41	3,23	2,87	3,16
Median	12	13	10,5	12,5	3,41	3,09	2,91	3,091
SD	3,3	2,3	2,4	3,9	0,303	0,355	0,539	0,659
Variance	10,571	5,238	5,66	14,9	0,092	0,126	0,291	0,434
Skewness	-2,19	-0,37	-0,45	-0,86	0	0,69	-0,51	1,07
Kurtosis	5	-1,69	-0,85	-0,48	0,4	-0,65	-0,39	1,07
Shaipro-Wilk W	0,7	0,92	0,93	0,86	1	0,93	0,96	0,91
p	0,002	0,461	0,357	0,044	0,999	0,553	0,847	0,187

Table 6.1: Experiment data analysis. PT = Privacy Table group, NL = Nutrition Label grouo, LAB = Laboratory experiment, INET = Internet experiment

To get an indication for whether our data was normally distributed, we first performed an analysis of the experiment data and based on this and based on this created three types of diagrams: histograms, normality plots and box-and-whisker diagrams. The histograms and normality plots[1] can be found in Appendix F.1 and the box-and-whisker diagrams can be seen in figure 6.2 (Information finding parts) and figure 6.3 (Likeability parts).

The box-and-whisker diagram consists of the following variables, bottom to top: the smallest observation, the lower quartile (Q1), the median (Q2), the upper quartile

[1]We used an 30-day evaluation copy of Analyse-It (http://www.analyse-it.com) for creating the diagrams

(Q3), and the largest observation. The lower quartile is the number that is one quarter from the start of the list and the upper quartile is the number that is three quarters from the start of the list.

For example, consider the laboratory experiment data for the Privacy Table group. The lowest participant score was 8 (out of 17), the lower quartile was 9, the median was 12, the upper quartile 13 and the best score was 14.

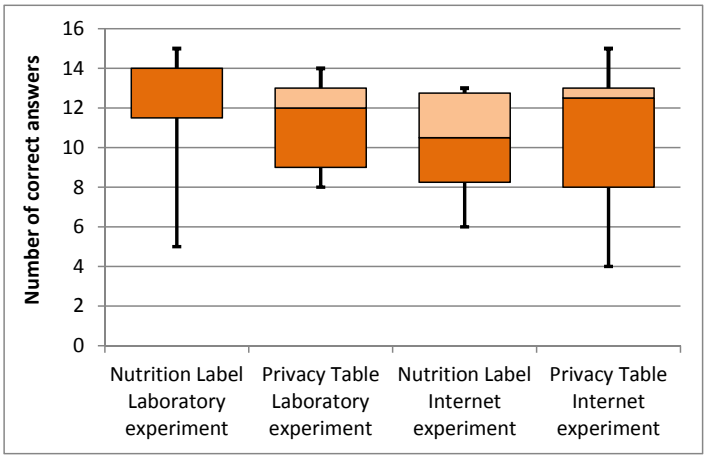


Figure 6.2: Box and whisker plot for the information finding parts

For the information finding parts, we scored each participant 0-17 points depending on correct answers, and performed an analysis for each group (Nutrition Label and Privacy Table) in both experiments. The first four leftmost columns in table 6.1 represent these data. For the likeability parts we scored the participants on a scale of 1-5 depending on level of agreement (with 1 equaling strongly disagree) and averaged each participant's score across all likeability questions. The four likeability parts can be found in the rightmost columns in table 6.1.

According to Flanagan-Hyde [2011], normal distributed data are represented by a symmetric box shape in the box-and-whisker diagram. An initial assumption based on figure 6.2 was therefore that the Nutrition Label group data from the laboratory experiment, and the Privacy Table group data from the Internet experiment were not normally distributed.

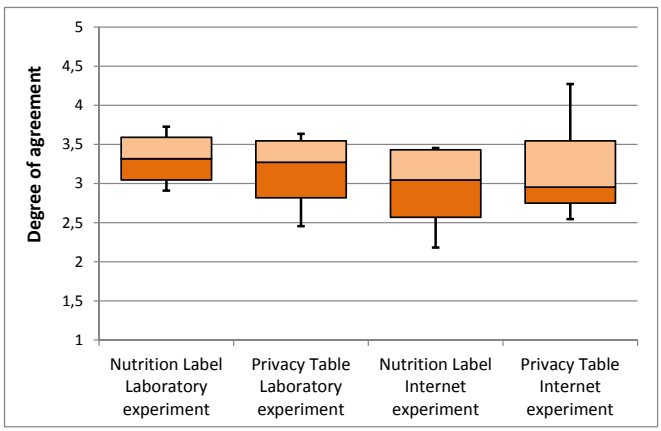


Figure 6.3: Box and whisker plot for the likeability parts

The assumption regarding the Nutrition Label group data from the laboratory experiment was backed by the findings from the Shaphiro-W test (P=0.002) and the values for skewness (-2.19) and kurtosis (5) which both were out of range. The assumption of non-normality in the data for the Privacy Table group in the Internet experiment was also backed by the Shaphiro-W test p=0.044, but the skew-

ness (-0.86) and kurtosis (-0.48) was within range. The corresponding Privacy Table group in the laboratory experiment (p=0.461, skewness=-0,37, kurtosis=-1.69) and the Nutrition Label group in the Internet experiment(p=0.357, skewness=-0.45, kurtosis=-0.85) were both assumed to be normally distributed.

Given the low number of samples in these experiments the Shaphiro-W test might not have detected non-normality, but the histograms and normality plots (see Appendix F.1) provided further evidence for non-normality in both the data for the NL group in the lab experiment, and the PT group in the Internet experiment (Signs for normal distribution is a bell-shaped form in the histogram, and points following the line in the normality plot [Analyseit, 2011].

As seen in table 6.1, there were stronger indications for the Likeability questions being normally distributed. All skewness and kurtosis values were within range, and none of the Shaphiro-W p-values was significant. While there was an indication for non-normality in the Privacy Table group in the Internet experiment (p=0.187), most of the groups had symmetrical box shapes in the box-and-whisker diagrams as seen in figure 6.14. It was therefore assumed that the samples were normally distributed.

Again, the Shaphiro-W tests might not have detected non-normality due to the small sample sizes, and as some of the histograms and normality plots (see Appendix F.1) indicated non-normality, we can only use these findings as indications.

Based on this discussion it is therefore important to highlight that while we have conducted a statistical analysis of the experiment results, the outcome of the analysis should be regarded as indications rather than statistical evidence. While we have assumed normal distribution and equal variances for the t-tests, this assumption might be wrong and the validity of the results could therefore be threatened. In order to explain the outcome of the t-tests, we will use terms such as “significant” and “near-significant” throughout the analysis, but only refer to these findings as indications in the remainder of the thesis.

6.3 The information finding questions

The following section presents the analysis of the information finding question parts. We performed descriptive statistics on each individual question and on combined parts, followed by t-tests to determine whether there was any difference in mean score between the groups. Full results of the descriptive statistics and t-tests can be found in Appendix F.2.

A t-test can either assume equal or unequal variances, and we therefore performed preliminary tests for the equality of variances (F-test) on all sample data before conducting the t-tests. If the measured P-value is less than the significant value (α = 0.05) the samples are assumed to be unequal.

We will present the result from each t-test on the following format:

- (There were no significant / There were significant) differences between the mean scores for Group A (Mean (**M**), Standard Deviation (**SD**, Subjects (**N**) and group B ((**M**), (**SD**), (**N**)) in the (Internet/laboratory experiment), with $t_{(degreesoffreedom)}$ = t-stat, p = p-value (α = x).

The group names will some places be shortened to NL (Nutrition Label) and PT (Privacy Table), and used as: **(NL: M, SD, N / PT: M, SD, N)**
which equals:
(Nutrition Label: Mean, Standard Deviation, Subjects / Privacy Table: Mean, Standard Deviation, Subjects)

α indicates the critical value, which was set to 0.05 for all t-tests. We will categorize the findings in the following three categories based on the p-value:

- P<0.05 - Significant finding
- 0,05<P<0.20 - Near-significant finding
- 0.20<P<0.30 - An indication that there might be a difference in means

We emphasize again that we use these terms with the purpose to describe the result of the t-test, and that all findings should be regarded as indications due to the uncertainty regarding the t-test assumptions.

6.3.1 Overall accuracy results

To get an overview of the performance in the information finding questions, we measured the average score for each format. Each participant answered a total of 17 information finding questions (part 1.1, 1.2 and 2.1) and was scored on a scale from 0 to 17 based on the number of correct answers.

The resulting data is presented in the box-and-whisker diagram in figure 6.4 and the mean scores for both groups in each experiment are also given in the diagram in figure 6.5.

As seen in the figure, the Nutrition Label performed slightly better than the Privacy Table in the laboratory experiment, while the Privacy Table performed slight better in the internet experiment.

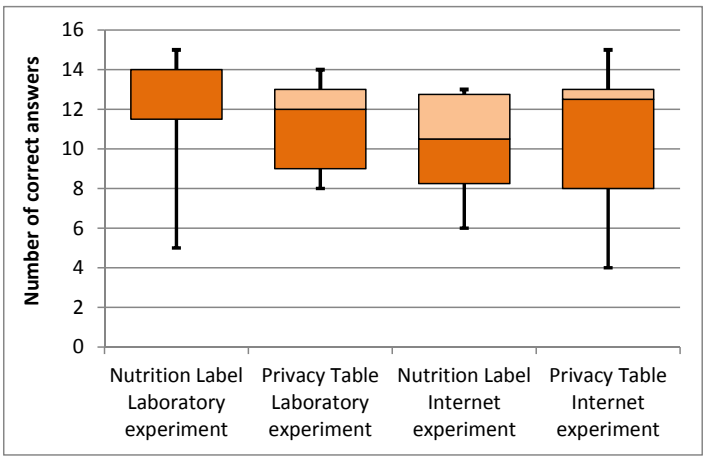


Figure 6.4: Box and whisker plot for the information finding parts

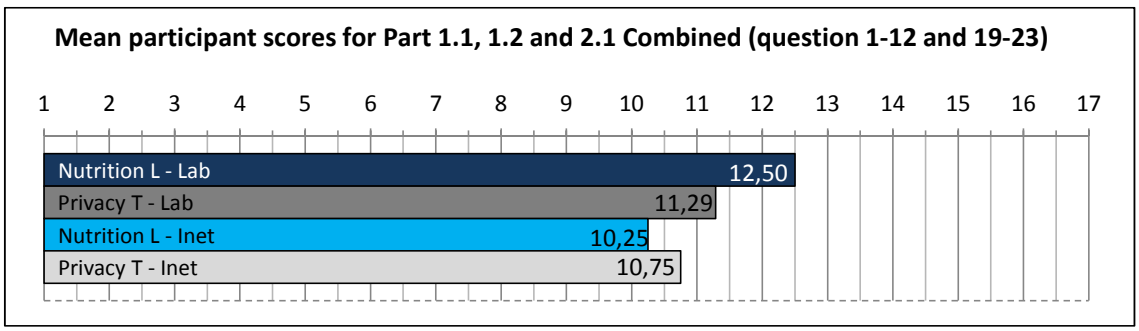


Figure 6.5: Mean scores for all parts

The overall result from the laboratory experiment was that we found no significant difference between the mean score for the Nutrition Label group ($M = 12.5, SD = 3.251, N = 8$) and the Privacy Table group ($M = 11.29, SD = 2.289, N = 7$) using the two-sample t-test for equal variances, $t_{(13)} = 0.824, p = 0.425$ ($\alpha = 0.05$). Based on the result from the preliminary test for the equality of variances ($F = 2,018, P = 0,205$) we assumed equal variance in the t-test.

For the Internet experiment, the result were similar with no significant difference between the mean score for the Nutrition Label group ($M = 10.25, SD = 2.379, N = 12$) and the Privacy Table group ($M = 10.75, SD = 3.864, N = 12$) using the two-sample t-test for equal variances, $t_{(22)} = -0.382, p = 0.706$ ($\alpha = 0.05$). We again assumed equal variances based on the initial F-test ($F = 0.379, P = 0,061$).

We therefore reject hypothesis H1.

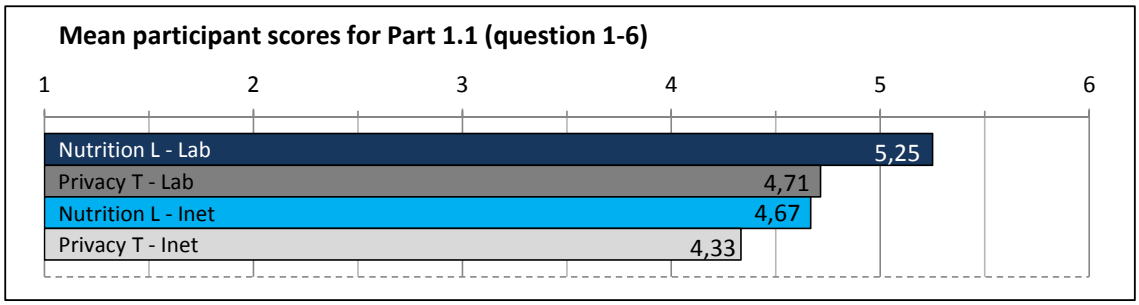


Figure 6.6: Mean scores for - Part 1.1

6.3.2 Part 1.1

6.3.2.1 Overall mean scores

As seen in figure 6.6, the Nutrition Label group scored averagely better in both the laboratory and the Internet experiment on the simple information finding questions. To see whether there was any significant difference between the mean scores, we performed a two-sample t-test on each experiment data set for this question. The preliminary F-test indicated that the variances of the two groups in the Laboratory were not significantly different ($F = 0.875, P = 0.427$), nor were the variances of the two groups in the Internet experiment ($F = 0,419, p = 0,083$). We therefore conducted two-sample t-tests that assumed equal variances.

For the laboratory experiment, we found a near-significant difference between the mean score for the Nutrition Label group ($M = 5.25, SD : 0.707, N = 8$) and the Privacy Table group ($M = 4.71, SD = 0.756, N = 7$) using the two-sample t-test for equal variances, $t_{(13)} = 1,418, p = 0.180$ ($\alpha = 0.05$).

For the Internet experiment, we found no significant difference between the Nutrition Label group ($M = 4.67, SD : 0.888, N = 12$) and the Privacy Table group ($M = 4.33, SD = 1.371, N = 12$) with $t_{(22)}) = 0.707, p = 0.487$ ($\alpha = 0.05$).

6.3.2.2 Question analysis

As seen in figure 6.7, all participants in the laboratory experiment answered correctly on the three first simple information questions. These questions, in addition to question 5 were on the form "Does the policy Acme to collect <information type>" and were regarded as the least challenging tasks. While there were a overall difference between the results from the laboratory and the Internet experiments (The laboratory participants scored an average 97% on question 1,2,3 and, 5, compared to 82% correct answers in the Internet experiment), the two formats scored similarly on these questions.

Question 4 and 6 in part 1.1 was more advanced, and also asked whether the col-

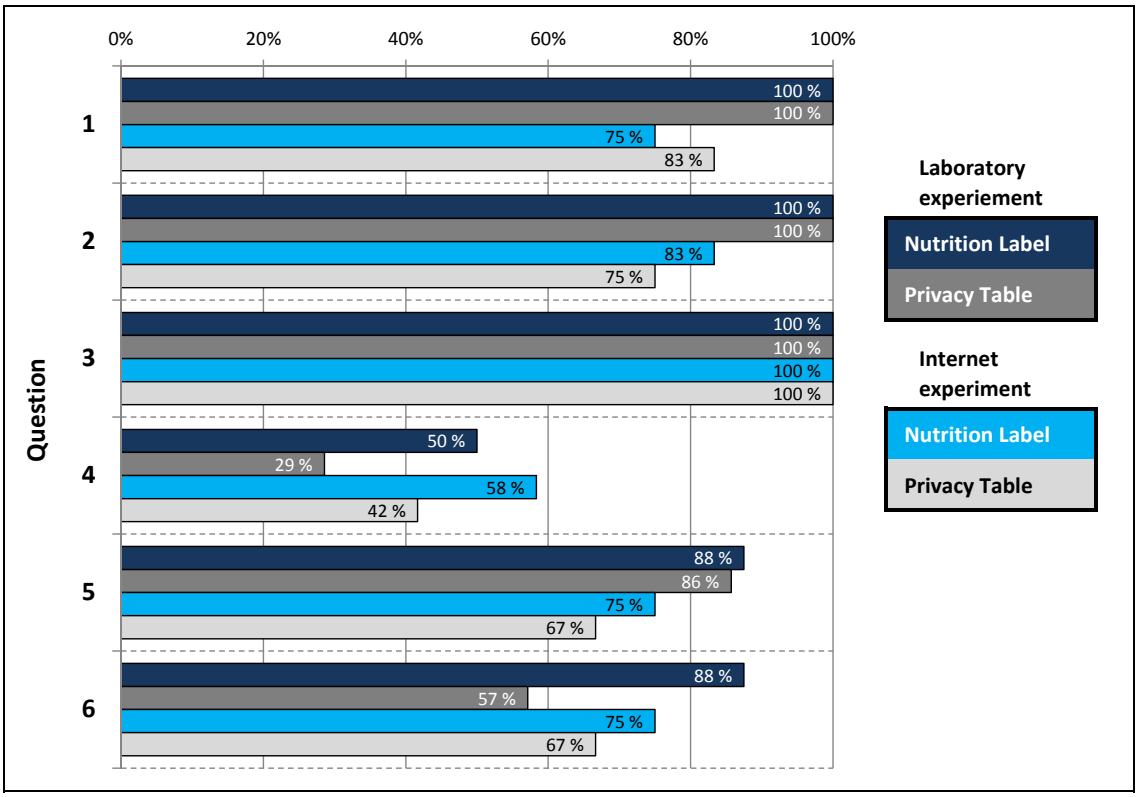


Figure 6.7: Results from information finding questions 1-6

lected information would be used for any purpose. Question 4 asked whether "the policy allow Acme to use information about your gender for marketing purposes", which only 29% (Laboratory experiment) and 42% (Internet experiment) of the participants using the Privacy Table answered correctly. The Nutrition Label scored higher on this, with 50% correct answers in the laboratory experiment and 58% in the Internet experiment.

On question 6, "Does the policy allow Acme to use your information to improve their website?" the laboratory experiment group using the Nutrition Label also scored averagely highest (88%). The Internet group using the same Nutrition Label did however score lower (75%), with both Privacy Table groups below that score (57% correct answers in the laboratory, and 67% correct in the Internet experiment).

6.3.2.3 Internet vs. laboratory results

As mentioned above, we noticed a clear difference between the participants in the Internet group compared to the participants from the Laboratory group on the simple information finding questions 1,2,3 and 5. We therefore scored the participants 0-4 points depending on correct answers in these questions, and combined the samples for both groups in each experiment. We then performed a t-test to see whether the mean score for the Internet participants were significantly different from the laboratory participant score.

The initial test for the equality of variances indicated that the samples were significantly different ($F = 0,168, P = 0,0006$) and we therefore assumed unequal variance in the t-test. We found a significant difference between the mean score for the participants in the Laboratory experiment ($M = 3.87, SD = 0.352, N = 15$) compared to the participants in the Internet experiment ($M = 3.29, SD = 0.859, N = 24$) on question 1,2,3 and 5 using a two-sample t-test for unequal variances, $t_{(33)} = 2,9125, p = 0.0064$ ($\alpha = 0.05$).

By conducting a similar t-test based on the scores in question 1-6, we found a near-significant difference between the mean scores (Laboratory: $M = 5.00, SD = 0.756, N = 15$ / Internet: $M = 4.50, SD = 1.142, N = 24$) on part 1.2 using a two-sample t-test for equal variances, $t_{(37)} = 1.644, p = 0.142$ ($\alpha = 0.05$). (The initial test for the equality of variances assumed equality, $F = 0,438, P = 0,056$).

6.3.3 Part 1.2

6.3.3.1 Overall mean scores

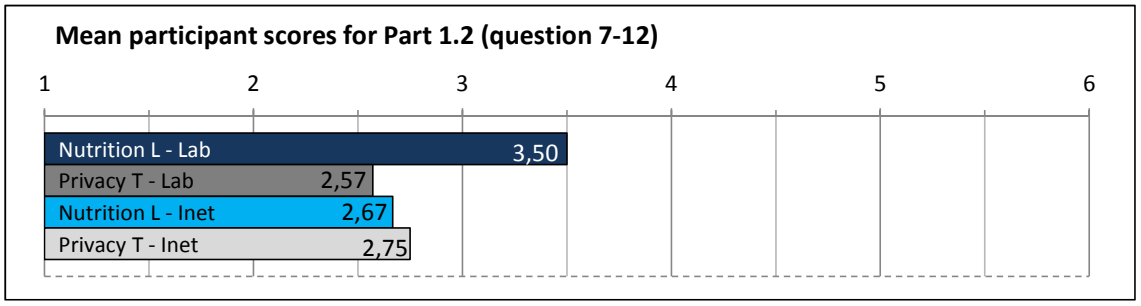


Figure 6.8: Mean scores for - Part 1.2

As for part 1.1, the Nutrition Label group in the laboratory experiment had the highest mean score (mean = 3.5 = 58% correct answers) for the hard information finding questions. The Privacy Table group had a mean score of 2.57 (43% correct answers) in the same experiment. Slightly lower scores was observed in the Internet experiment with 38% correct answers for the Nutrition Label group (mean = 2.67) and 39% for the Privacy Table group (mean = 2.75).

As for part 1.1, we performed two two-sampled t-tests to see whether there were any significant differences between the mean scores for the groups. The initial F-test (equality of variances) indicated that the variances of the two groups in the Laboratory experiment were not significantly different ($F = 1.111, P = 0.457$), while the variances of the two groups in the Internet experiment were significantly different ($F = 0.177, P = 0.004$). We therefore assumed equal variance in the laboratory data, and unequal variance on the Internet data.

For the laboratory experiment, we found a near-significant difference between the mean score for the Nutrition Label group ($M = 3.5, SD = 1.195, N = 8$) and the

Privacy Table group ($M = 2.57, SD = 1.134, N = 7$) using the two-sample t-test for equal variances, $t_{(13)} = 1.537, p = 0.148$ ($\alpha = 0.05$).

For the Internet experiment, we found no significant difference between the mean score for the Nutrition Label group ($M = 2.67, SD = 0.651, N = 12$) and the Privacy Table group ($M = 2.75, SD = 1.545, N = 12$) with $t_{(15)} = -0.172, p = 0.866$ ($\alpha = 0.05$).

6.3.3.2 Question analysis

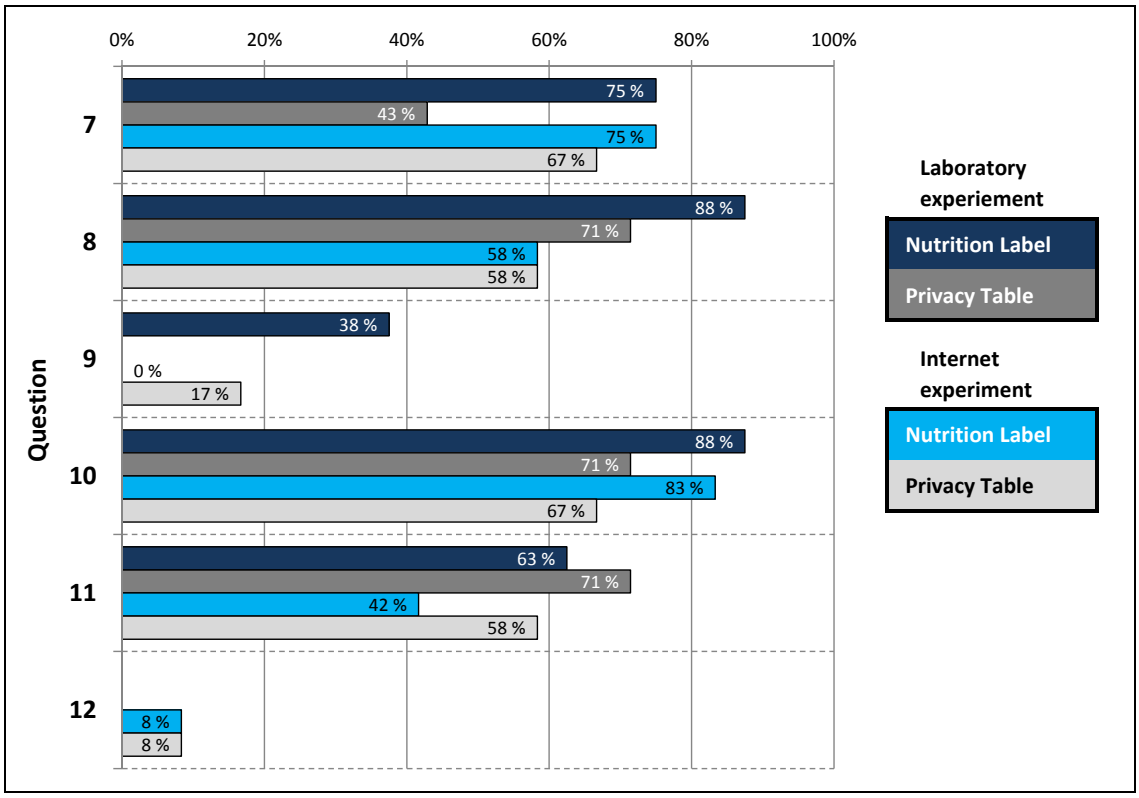


Figure 6.9: Results from information finding questions 7-13

The Nutrition Label group had 32% more correct answers than the Privacy Table group on question 7 ("Does the policy allow Acme to share your personal information on public forums?") in the Laboratory experiment. The formulation of this question was however a bit vague, as it was up to the respondent to interpret what "personal information was". If the respondent interpreted it as "preferences" or "activity on this site" the natural answer to the question would be yes, while the correct answer was no.

While the difference was less in question 8 ("Does the policy allow Acme to share your mobile number with other companies?") and 10 ("Does the policy allow Acme to share your cookie information with other companies?"), the Nutrition Label group in the laboratory experiment also had the highest average score on these question.

The two groups in the Laboratory experiment scored similarly on question 7, 8 and 10.

Common for these questions was that they all asked whether the information would be shared with any 3rd party. We performed a t-test (We assumed equal variance in the t-test($(F = 0.5, P = 0.193)$) where we scored the participants 0-3 points based on correct answers for these questions (7, 8 and 10), and found that the Nutrition Label group ($M = 2.5, SD = 0.756, N = 8$) had a near-significant higher mean score than the Privacy Table group ($M = 1.86, SD = 1.069, N = 7$) on question 7, 8 and 10 in the laboratory experiment, ($t_{(13)} = 1.36, p = 0.197$ ($\alpha = 0.05$)).

The Privacy Table groups had the highest average score on question 11 ("Will Acme contact you with advertisements?") in both experiments.

Two questions in this part, 9 and 12, had a particularly low score in both experiments. Question 9 asked whether "the policy allows Acme to use your buying history to design custom functionality targeted at you?", with "No" as correct answer. Of a combined 39 participants in both experiments, just 5 answered this question correctly. As seen by the detailed overview in figure 6.10, a majority of the participants answered "Yes" or "Yes, unless I tell them not to"(Opt out).

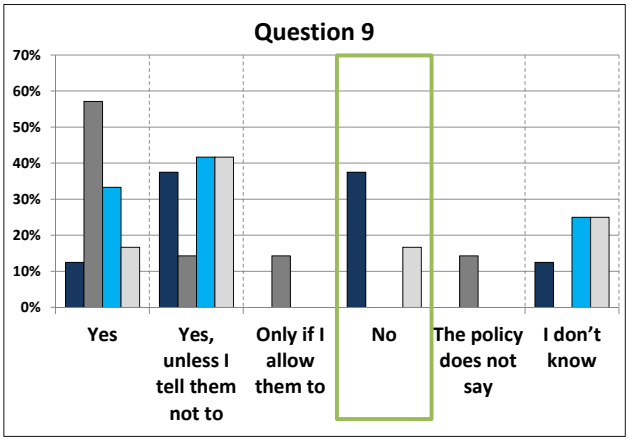


Figure 6.10: The response distribution for question 9. The green box indicates the correct answer.

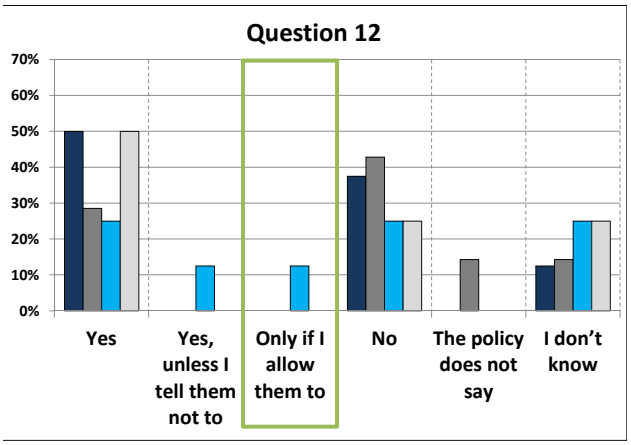


Figure 6.11: The response distribution for question 12. The green box indicates the correct answer.

Question 12 scored even lower, with only 2 out of 39 participants answering correct. This question asked " Does Acme give you control regarding their sharing of your personal data?", with "Only if I allow them to" (Opt in) as the correct answer. As seen figure 6.11, most participants answered either "Yes" or "No" to this question.

6.3.4 Part 2.1

6.3.4.1 Overall mean scores

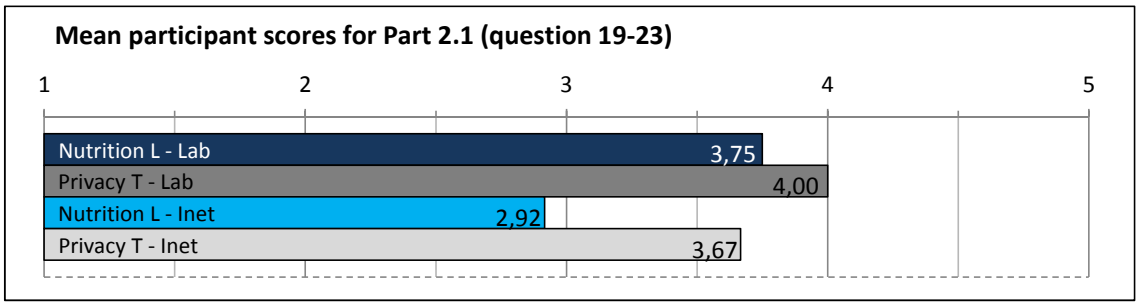


Figure 6.12: Mean scores for - Part 2.1

Moving to the comparison questions in part 2.1, the Privacy Table group performed better than the Nutrition Label in both the laboratory experiment (PT:4.00 (80% correct), NL: 3.75 (75%correct)) and in the internet experiment (PT: 3.67 (73% correct), NL: 2.92 (58%correct)).

As for part 1, we performed two-sampled t-tests to see whether there were any significant differences between the groups in part 2. The initial F-test (equality of variances) indicated that the variances of the two groups was not significantly different in neither the Laboratory experiment ($F = 3.36, P = 0.08$) nor in the Internet experiment ($F = 1.33, P = 0.32$). We therefore assumed equal variance in both t-tests.

For the laboratory experiment, we found no significant difference between the mean score for the Nutrition Label group ($M = 3.75, SD = 1.832, N = 8$) and the Privacy Table group ($M = 4.00, SD = 1.00, N = 7$) using the two-sample t-test for equal variances, $t_{(13)} = -0.321, p = 0.754$ ($\alpha = 0.05$).

While we found no significant differences between the mean scores in the Internet experiment, there were indications that the Privacy Table group ($M = 3.67, SD = 1.497, N = 12$) had a higher mean score than the Nutrition Label group ($M = 2.92, SD = 1.730, N = 12$), with $t_{(22)} = -1.136, p = 0.268 (\alpha = 0.05)$.

6.3.4.2 Question analysis

The Privacy Table and the Nutrition Label group scored more or less similar across all questions in the laboratory experiment. The Privacy Table scored around 10% better than the Nutrition Label on question 19, 20 and 22, while the Nutrition Label was marginally better in question 21 and 23.

The results were more varying in the internet experiment. With the exception of question 19 where the Privacy Table group had 41% more correct answers on average

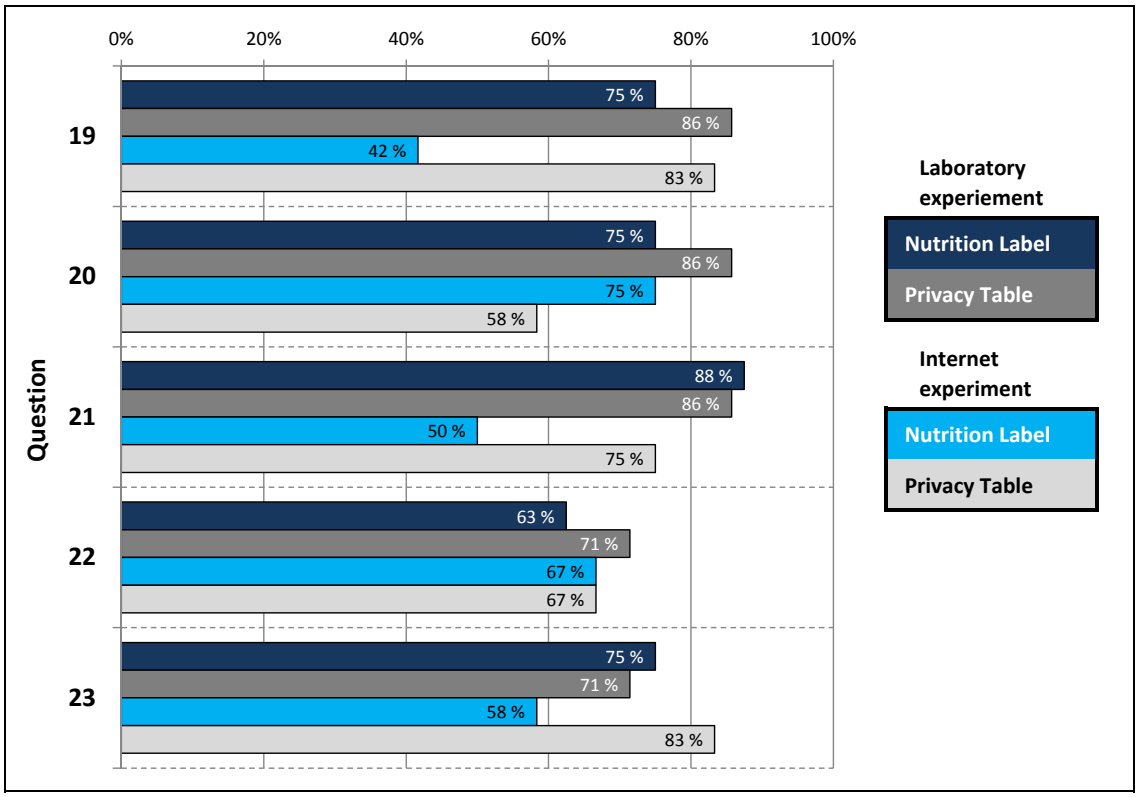


Figure 6.13: Results from information finding questions 19-23

than the Nutrition Label, the findings from the Internet experiment contradicted the results from the laboratory experiment (i.e. on questions where the Nutrition Label group performed best in the laboratory experiment, the Privacy Table group scored best in the Internet experiment and vice versa). The Nutrition Label group performed better than the Privacy Table in question 20; while the Privacy Table group performed best in question 19, 21 and 23.

An interesting finding in part 2.2 was the result from question 19 ("Both Acme and Bell lets you control whether non-sensitive information (such as name and address) is shared with other companies") and question 22 ("By default, Bell can share your history of purchased items with other companies, but Acme cannot"). As mentioned in section 6.3.3, the Nutrition Label group scored higher on all questions regarding how information was shared with 3rd parties. This finding is contradicted by the results from question 19 and 22, where the Privacy Table groups scored better in both experiments.

6.4 The likeability questions

To get an overview of how enjoyable the participants found each format, we summarized the mean score for the 5-point likeability questions. The likert-scale ranged from "Strongly disagree", which was weighted 1, to "Strongly agree" which was weighted with 5. In other words, for these questions a high score would indicate satisfaction with the format and a low score indicate dis-satisfaction.

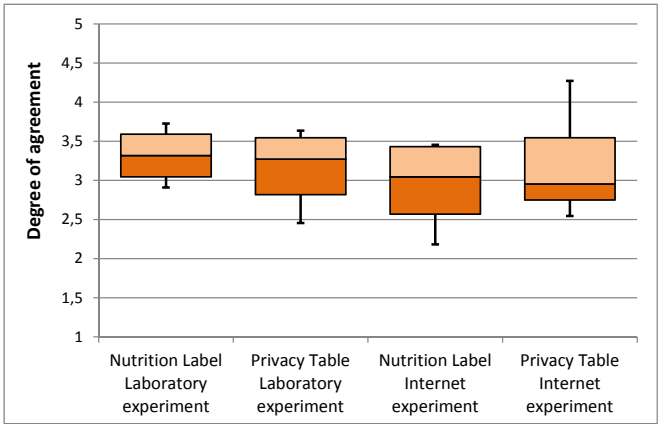


Figure 6.14: Box and whisker plot for the likeability parts

As question 14 was negatively posed ("I had problems finding the information I was looking for in Acmes policy"), we inverted the results for this question (e.g. a "Strongly agree" (5) was converted to a "Strongly disagree" (1).) The box-and-whisker diagram in figure 6.14 presents the overall statistics for the mean scores of all questions in part 1.3 and 2.2.

Given question 18 and 28 was custom designed to measure the effect symbols and colors had in the design, and a high or low agreement would not indicate a general likeability to the format, these questions were excluded from further calculations on the overall statistics. Table 6.2 summarizes the means for each format in the single policy questions (1.3) and in the comparison questions (2.2), as well as aggregate results for both parts, excluding question 18 and 28.

Extended results for the descriptive statistics and the t-tests can be found in appendix F.3.

	Laboratory experiment		Internet experiment	
Part	Nutrition Label	Privacy Table	Nutrition Label	Privacy Table
1.3	3.45	3.49	2.48	3.17
2.2	3.72	3.32	3.10	3.31
Aggr.	3.57	3.41	2.76	3.23

Table 6.2: Mean enjoyability scores on a 5-point likert scale (ranged from "Strongly disagree"=1 to "Strongly agree"=5) for single policy questions (Part1.3) and comparison questions (part 2.2), as well as aggregate results for both parts.

We averaged the likeability scores for each participant and performed two-sample

t-test with unequal sample sizes (equal for the Internet experiment) for the same data categories as shown in table 6.2. I.e. we performed a total of six t-tests: two t-tests per question part (1.3, 2.2 and combined), one for each experiment.

6.4.1 Overall results

Preliminary tests for the equality of variances (F-tests) indicated that the variances between all groups were assumed to be equal [2] (P-values above 0,05 indicates equal variances). Therefore, two-sample t-tests that assumed equal variances were performed on the sample data.

The overall score for both question parts combined (1.3 and 2.2) in the laboratory experiment was that there were no significant difference between the mean scores for the Nutrition Label group ($M = 3.57, SD = 0.947, N = 72$) and the Privacy Table group ($M = 3.41, SD = 1.116, N = 63$), with $t_{(13)} = 0.763, p = 0.475$ ($\alpha = 0.05$).

For the same combined parts in the Internet experiment, the Privacy Table ($M = 3.23, SD = 1.001, N = 108$) mean was near-significantly higher than the Nutrition Label ($M = 2.76, SD = 1.135, N = 108$) with $t_{(22)} = -1.469, p = 0.156$ ($\alpha = 0.05$).

We therefore reject hypothesis H3.

As seen in table 6.2, the two formats had similar mean-values on the likert-questions in part 1.3 in the laboratory experiment. The t-test showed that there were no significant difference between the mean for the two groups (NL: $M = 3.45, SD = 0.959, N = 40$ / PT: $M = 3.49, SD : 1.121, N = 35$), with $t_{(13)} = -0.110, p = 0.914$ ($\alpha = 0.05$).

For the Internet experiment, the Privacy Table scored almost a three quarter point higher on the likert-scale than the Nutrition Label in part 1.3. The result from the t-test showed that the Privacy Table ($M = 3.17, SD = 1.044, N = 60$) was near-significantly more liked than the Nutrition Label ($M = 2.48, SD = 1.157, N = 60$) with $t_{(22)} = -1.879, p = 0.074$ ($\alpha = 0.05$).

There were no overall significant differences in part 2.2, with the Nutrition Label group mean higher in the laboratory experiment (NL: 3.72, PT: 3.32), and the Privacy Table group mean higher in the internet experiment (NL: 3.10, PT: 3.31). Details of these two t-tests, in addition to full results for the other t-tests and descriptive statistics for each question can be found in section F.3 in appendix F.

[2] (Part 1.3: P = 0,328 (Laboratory) and P = 0,473 (Internet), Part 2.2: P = 0,061 (Laboratory) and P = 0,382 (Internet), Combined: P = 0,393 (Laboratory) and P = 0,374 (Internet).

6.4.2 Part 1.3

The following section describes the findings from the analysis of part 1.3 in detail. Only the main findings are presented here, and the detailed results from the t-tests can be found in table F.6 in appendix F.3. Descriptive statistics for this part can be found in table F.3 (laboratory experiment) and table F.4 (Internet experiment) in appendix F.3. The scale values used in the figures corresponds to the following likert scale:

1. - Strongly Disagree
2. - Disagree
3. - Neither agree nor disagree
4. - Agree
5. - Strongly agree

6.4.2.1 Question 13

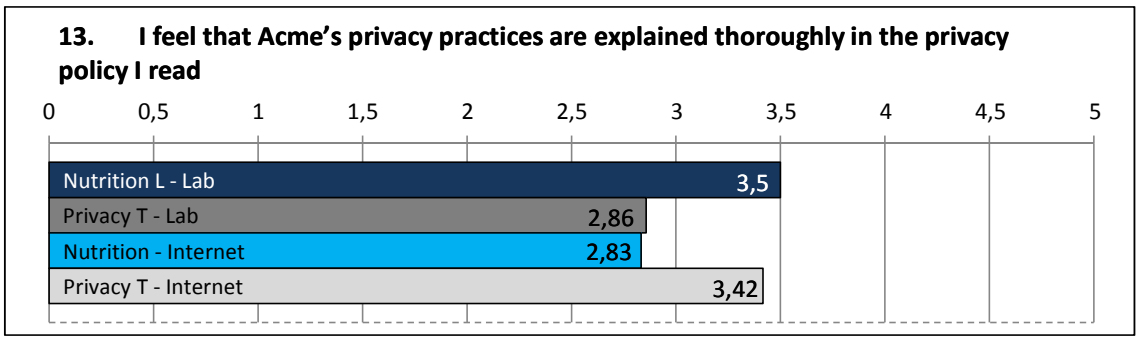


Figure 6.15: Mean scores for question 13 - Part 1.3

Question 13 was reused from the Kelley et al. [2009] study and designed to test whether the participants believed their privacy practices were well displayed in the policy. While the group using the Nutrition Label in the laboratory experiment had a higher degree of agreement to this statement, the results were contradictionary in the internet experiment where the Privacy Table group agreed more.

A interesting finding from the laboratory experiment was that all participants in both groups either disagreed or agreed to this question. Six participants agreed and two disagreed from the Nutrition Label group (median = 4, mode = 4), compared to three participants agreeing and four disagreeing in the Privacy Table group (median = 2, mode = 2). The results were more varied in the internet experiment with both formats having a value of 3 (neutral) as median and mode.

We found an indication but no statistical evidence for that the mean likeability for the Nutrition Label group ($M = 3.5, SD = 0.926 N = 8$) was higher than the Privacy Table group mean ($M = 2.86, SD = 1.069 N = 7$) in the laboratory experiment for this question, with $t_{(13=}1.249, p = 0.234$ ($\alpha = 0.05$).

The findings from the Internet experiment were contradictionary to those above, as it showed that the Privacy Table group ($M = 3.42, SD = 0.996, N = 12$) had a near significantly higher mean than the Nutrition Label group ($M = 2.83, SD = 0.937, N = 12$ with $t_{(22)} = -1.477, p = 0.154$ ($\alpha = 0.05$).

6.4.2.2 Question 14

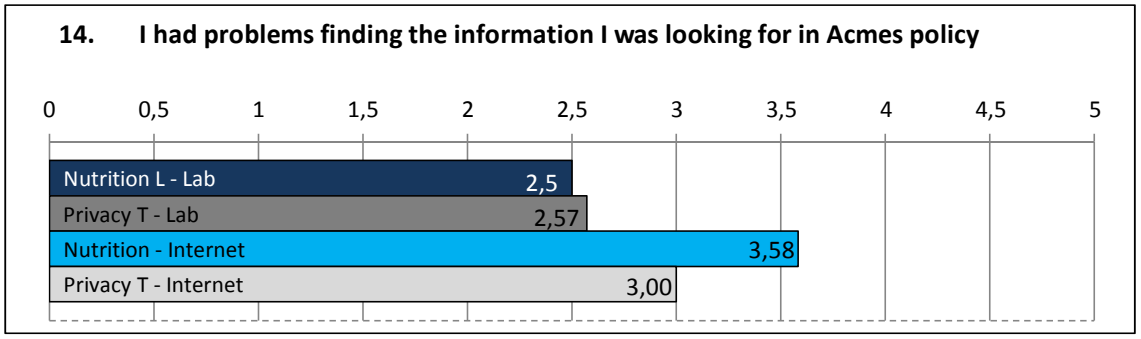


Figure 6.16: Mean scores for question 14 - Part 1.3

Question 14 was a modified version of a question from the Kelley et al. [2009] study (the original was "It was hard to find information in Acme's policy"), and was designed to rate the difficulty the participants had in finding information in the policy. As this was the only question which was negatively posed, the results were inverted in the overall statistics as described in section 6.4. Figure 6.16 presents the non-inverted results for this question.

There were no significant difference between the mean scores for the Nutrition Label group ($M = 2.5, SD = 0.926, N = 8$) and the Privacy Table group ($M = 2.57, SD = 1.134N = 7$) in the laboratory experiment for this question, with $t_{(13)} = -0.134, p = 0.895$ ($\alpha = 0.05$).

The Nutrition Label group ($M = 3.58, SD = 1.045N = 12$) had a higher degree of agreement to question 14 in the Internet experiment compared to the Nutrition Label group ($M = 3, SD = 0.9962N = 8$) and we found that the NL group had a near-significantly higher mean score than the PT group, with $t_{(22)} = 1.402, p = 0.175$ ($\alpha = 0.05$). 58% of the participants using the Nutrition Label in the Internet experiment agreed or strongly agreed to this statement (compared to 42% agreeing in the Privacy Table group), while 33% of the participants using the Privacy Table disagreed or strongly disagreed (compared to 16% of the Nutrition Label group) in the same experiment.

6.4.2.3 Question 15

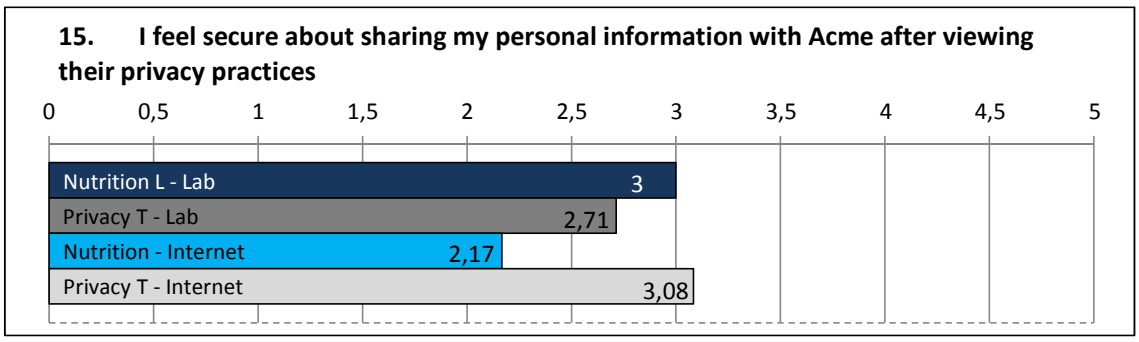


Figure 6.17: Mean scores for question 15 - Part 1.3

Question 15 was reused from the Kelley et al. [2009] study and designed to capture the participant's reaction to the actual content of the privacy policy they read. As for question 13, the results in question 14 were also contradictionary. While the group using the Nutrition Label in the laboratory experiment ($M = 3, SD = 0.926, N = 8$) agreed more to this statement than the group using the Privacy Table ($M = 2.71, SD = 0.951, N = 7$), there were no significant differences with $t_{(13)} = 0.589, p = 0.566$ ($\alpha = 0.05$).

For the Internet experiment we found near-significant differences between the mean scores for the Nutrition Label group ($M = 2.17, SD = 1.267N = 12$) and the Privacy Table group ($M = 3.08, SD = 1.083, N = 12$) for this question, with $t_{(22)} = -1.904, p = 0.070$ ($\alpha = 0.05$). 42% of the participants using the Nutrition Label in the Internet experiment disagreed strongly to this statement (median = 2, mode = 1), compared to none in the Privacy Table group (median = 3, mode = 3).

6.4.2.4 Question 16

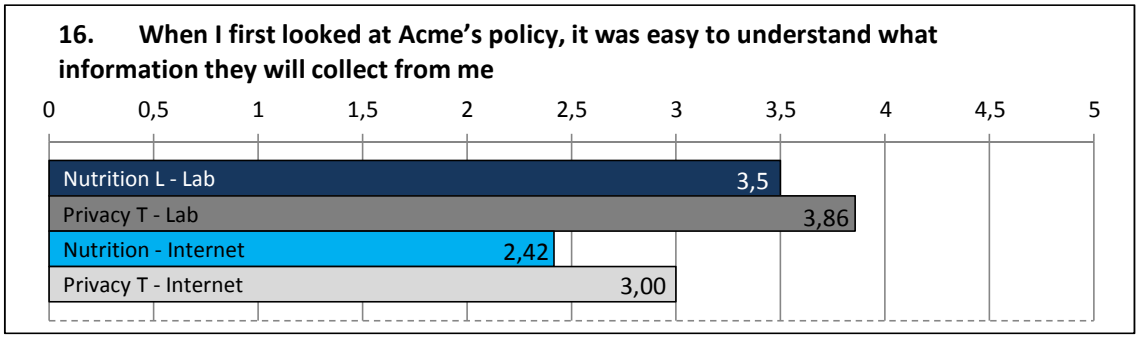


Figure 6.18: Mean scores for question 16 - Part 1.3

Question 16 was designed to capture the participant's first impression of the label, to see how a first-time user would handle the format. The Privacy Table scored highest in both experiments for this question. In the laboratory experiment, a majority of the participants using the Privacy Table agreed to this statement (median = 4,

mode = 4) while the Nutrition Label scored lower (median = 3,5, mode = 2). There were however no significant differences between the mean scores for the groups (NL: $M = 3.5, SD = 1.195, N = 8$ / PT: $M = 3.86, SD = 0.900, N = 7$) in the laboratory experiment for this question, with $t_{(13)} = -0.645, p = 0.530$ ($\alpha = 0.05$).

For the Internet experiment, the Privacy Table group (median = 3, mode = 3) had a higher mean than the Nutrition Label group (median = 2, mode = 1), where 60% of the participants from the latter group disagreed or disagreed strongly to this question. We found an indication, but no statistical evidence for that the mean score for the Privacy Table group ($M = 3.00, SD = 1.044, N = 12$)was higher than for the Nutrition Label group ($M = 2.42, SD = 1.379, N = 12$), with $t_{(22)} = -1.168, p = 0.255$ ($\alpha = 0.05$).

6.4.2.5 Question 17

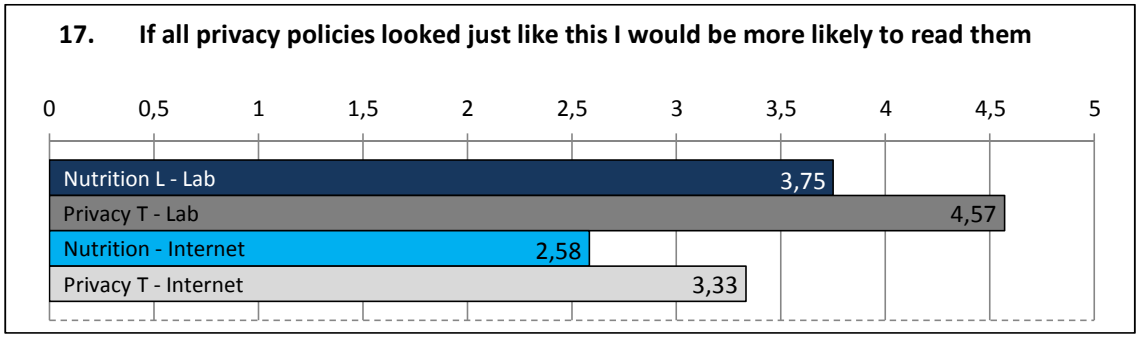


Figure 6.19: Mean scores for question 17 - Part 1.3

Question 17 was reused from the Kelley et al. [2009] study and designed to capture whether the proposed label would encourage more people to read policies. We found significant differences between the mean scores for the Nutrition Label group ($M = 3.75, SD = 0.886, N = 8$) and the Privacy Table group ($M = 4.57, SD = 0.535, N = 7$) in the laboratory experiment with $t_{(13)} = -2.203, p = 0.048$ ($\alpha = 0.05$). A mean of 4.57 for the Privacy Table group in the laboratory experiment was the highest measured value for the likeability-questions, with all participants either agreeing or strongly agreeing (median = 5, mode = 5). The Nutrition Label group had more varied results, with the majority of participants agreeing to question 17 (median = 4, mode = 4).

We also found near significant differences between the mean scores for the two groups (NL: $M = 2.58, SD = 1.240, N = 12$ / PT: $M = 3.33, SD = 1.155, N = 12$) in the Internet experiment, with $t_{(22)} = -1.533, p = 0.139$ ($\alpha = 0.05$). Both formats had lower mean-values in the Internet experiment, with the Nutrition Label group half a point below the neutral value (3), and the Privacy Table half a point above.

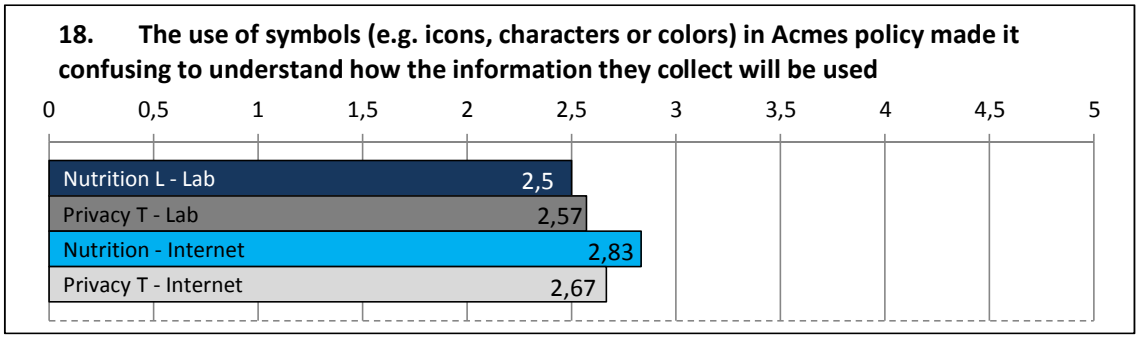


Figure 6.20: Mean scores for question 18 - Part 1.3

6.4.2.6 Question 18

Question 18 was custom designed measure whether the use of symbols in the policy was helping or confusing the participants in the information finding tasks. The question was negatively posed and the mean value for both formats in both experiments came between 2.5 and 3 on the likert-scale.

There were no significant differences between the mean scores for the two groups (NL: $M = 2.5, SD = 0.756, N = 8$ / PT: $M = 2.57, SD = 1.397, N = 7$) in the laboratory experiment for this question, with $t_{(13)} = -0.126, p = 0.902$ ($\alpha = 0.05$), nor in the internet experiment (NL: $M = 2.83, SD = 0.1.030, N = 12$ / PT: $M = 2.66, SD = 0.9847, N = 12$), with $t_{(22)} = 0.405, p = 0.689$ ($\alpha = 0.05$).

6.4.3 Part 2.2

The following section describes the findings from the analysis of part 2.2 in detail. As with part 1.3, only the main findings are presented here, and the detailed results can be found in in Appendix F.3. (t-tests: table F.7, descriptive statistics: table F.3(Lab) and (table F.4 (Internet)) The likert scale for part 2.2 was the same as in part 1.3 and its values corresponds to the same degrees of agreement.

6.4.3.1 Question 24

Question 24 was reused from the Kelley et al. [2009] study and measured whether the participant found the comparison task easy to do.

The Nutrition Label group ($M = 4.13, SD = 0.834, N = 8$) had a mean value half a point higher than the Privacy Table group ($M = 3.57, SD = 0.976, N = 7$) in the laboratory experiment on this question, and we found indications for, but no statistical evidence, that the means could be different with $t_{(13)} = 1.185, p = 0.257$ ($\alpha = 0.05$).

The two groups (NL: M=3.33, SD=1.155, N=12 / PT: M=3.25, SD=0.965, N=12)

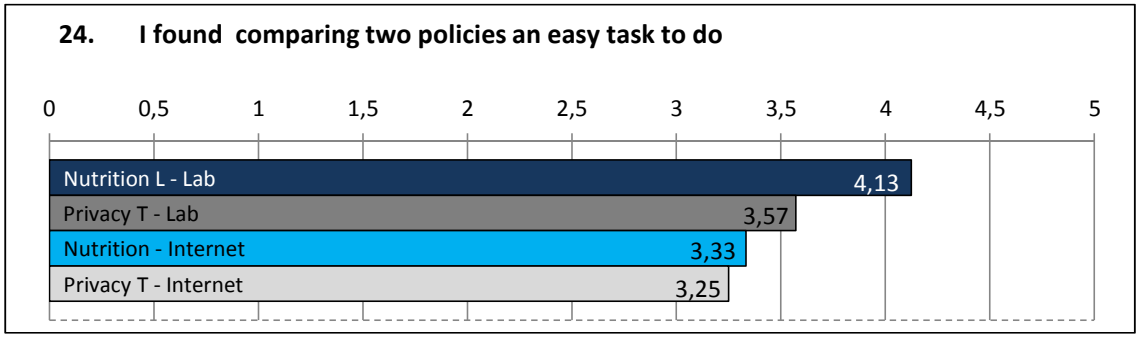


Figure 6.21: Mean scores for question 24 - Part 2.2

had similar means on the internet experiment, with $t_{(22)} = 0.192, p = 0.850$ ($\alpha = 0.05$). The mode value was 4 for both groups in both experiments, showing that most participants agreed to finding the comparison task easy.

6.4.3.2 Question 25

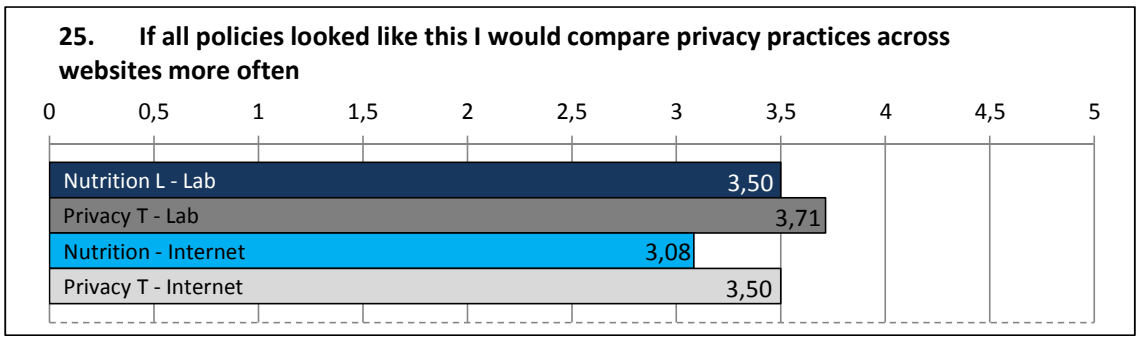


Figure 6.22: Mean scores for question 25 - Part 2.2

Question 25 was designed to test whether the policy format would motivate for a more frequent comparison of privacy practices across sites. This question closely resembled question 17 from the single policy likeability part, which asked whether the format would motivate the participants or read policies more often. While the Privacy Table group scored higher than the Nutrition Label in both experiments on question 25, the results were not significant as for question 17.

While the Privacy Table group ($M = 3.71, SD = 0.951, N = 7$) had a higher mean score than the Nutrition Label group ($M = 3.5, SD = 0.926, N = 8$) in the Laboratory experiment, we found no significant difference between the means, with $t_{(13)} = -0.442, p = 0.666$ ($\alpha = 0.05$).

The Privacy Table group ($M = 3.5, SD = 1.000, N = 12$) also had a higher mean than the Nutrition Label group ($M = 3.08, SD = 0.793, N = 12$) in the Internet experiment, and while we found indications for that the difference in means could be different, there was no statistical evidence: $t_{(22)} = -1.131, p = 0.270$ ($\alpha = 0.05$).

6.4.3.3 Question 26

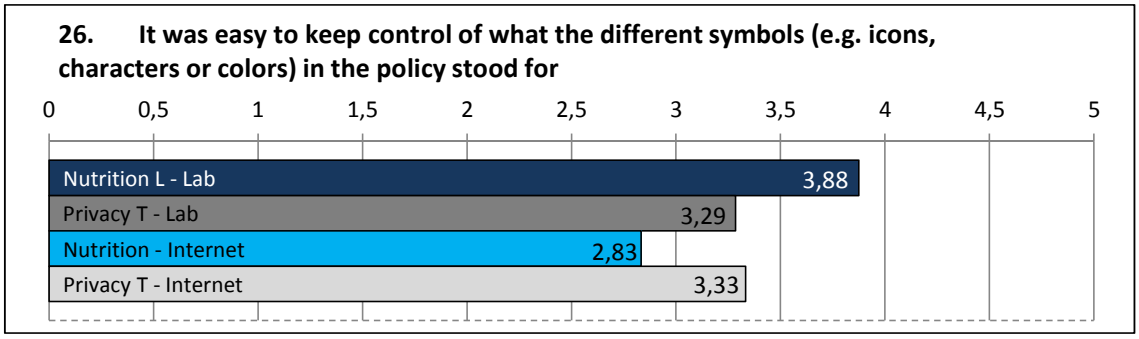


Figure 6.23: Mean scores for question 26 - Part 2.2

Question 26 was designed to test whether the participants found it easy to keep control of the symbols in the policies during the comparison experience. The purpose of the question was to see whether the reduced use of symbols in the Privacy Table would have any effect. As seen in figure 6.23, the mean score of the Privacy Table groups were similar in both experiments, while the Nutrition Label groups differed by over a point on the likert-scale across the experiments. All participants using the Nutrition Label in the laboratory experiment were either neutral or agreed to some degree to this statement (median =4, mode = 3), while just 16% of the Nutrition Label group agreed or strongly agreed in Internet experiment (median = 3, mode = 3). The corresponding median and mode values for the Privacy Table group were similar, with the exception of a mode value of 4 in the laboratory experiment (71% agreed while 29% disagreed or strongly disagreed).

While the Nutrition Label group ($M = 3.88, SD = 0.835, N = 8$) had a higher mean than the Privacy Table group ($M = 3.29, SD = 1.254, N = 7$) in the laboratory experiment, there was no significant difference between the groups with $t_{(13)} = 1.085, p = 0.297$ ($\alpha = 0.05$). For the Internet experiment, the Privacy Table group ($M = 3.33, SD = 0.888, N = 12$) had a higher mean than the Nutrition Label group ($M = 2.83, SD = 1.030, N = 12$), and while there was indications for a difference in means, the t-test gave no statistical evidence $t_{(22)} = -1.274, p = 0.216$ ($\alpha = 0.05$).

6.4.3.4 Question 27

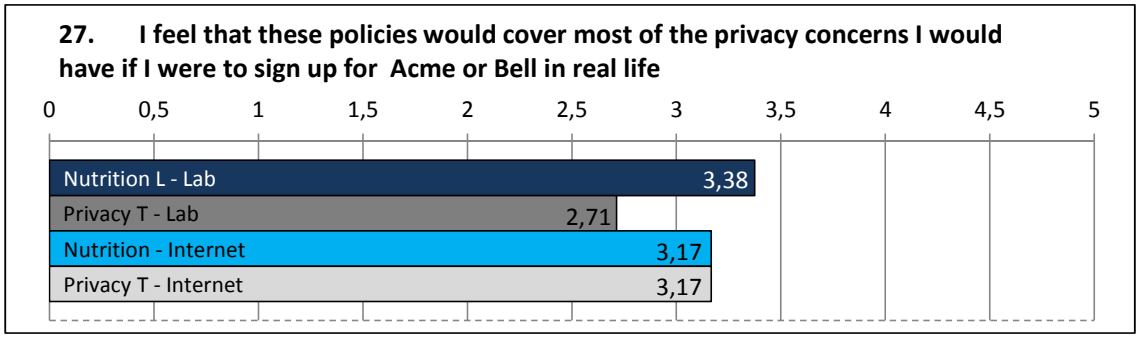


Figure 6.24: Mean scores for question 27 - Part 2.2

Question 27 was designed to test whether the respondent felt that the content of the policy format covered his or hers privacy concerns if he/she was to sign up for Acme or Bell in real life. Given the content of the policies were the same, similar results were expected in this question. The equal mean score of the two groups (NL: $M = 3.17, SD = 1.115, N = 12$ / PT: $M = 3.17, SD = 1.030, N = 12$) in the internet experiment demonstrated this expectation with $t_{(22)} = 0.000, p = 1.000$ ($\alpha = 0.05$). While the mean scores were similar in the Internet experiment, an detailed analysis of the results showed that a majority of the respondents in the Nutrition Label group agreed to this statement (mean = 4, mode = 4), while a majority of the Privacy Table group respondents were neutral (median = 3, mode = 3).

The differences were bigger in in the laboratory experiment, where a majority of the respondents (57%) in the Privacy Table group ($M = 2.71, SD = 1.124, N = 7$) disagreed to some degree to this question, while the majority of respondents in the Nutrition Label group ($M = 3.38, SD = 1.061, N = 8$) agreed. While there was an indication for a difference in mean between the groups, the t-test showed no significant differences with $t_{(13)} = 1.107, p = 0.289$ ($\alpha = 0.05$).

6.4.3.5 Question 28

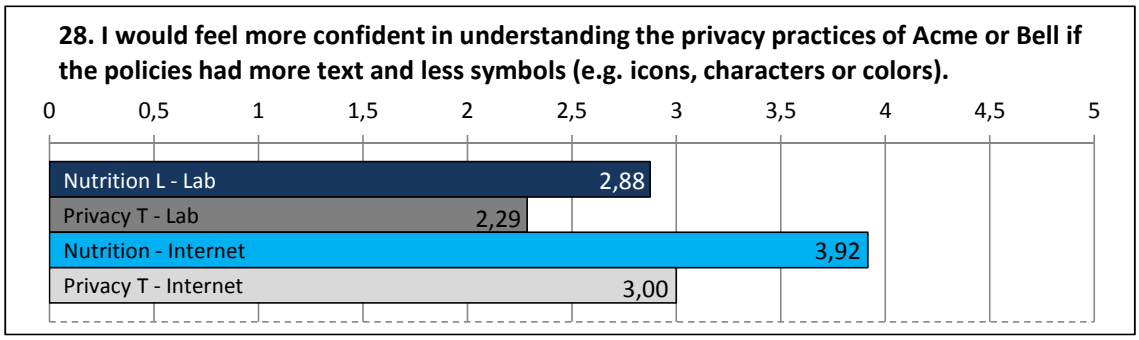


Figure 6.25: Mean scores for question 28 - Part 2.2

The final question, question 28, was designed to test whether the respondent would prefer more text and less use of symbols in the policy. The motivation behind asking this was to see whether the reduced use of symbols and a higher emphasis on the textual content would have any effect. It is important to highlight that the Privacy Table and the Nutrition Label had identical textual content.

No participants using the Privacy Table group ($M = 2.29, SD = 0.756, N = 7$) in the laboratory experiment agreed to this statement, with a majority (57%) disagreeing (43%) or strongly disagreeing (14%). The results from the Nutrition Label group ($M = 2.88, SD = 1.126, N = 8$) in the same experiment were different, with equal distribution between agreeing (38%) or disagreeing (to some degree) (38%). While there was an indication for a difference between the means of the two groups, the t-test found no significant difference with $t_{(13)} = 1.170, p = 0.263$ ($\alpha = 0.05$).

For the Internet experiment, we noticed a increased mean score for both groups. We found near significant difference between the mean score of the Privacy Table group ($M = 3.00, SD = 1.279, N = 12$) and the Nutrition Label group ($M = 3.92, SD = 1.240, N = 12$) with $t_{(22)} = 1.782, p = 0.089$ ($\alpha = 0.05$).

The Nutrition Label group had a consistent mean score in both experiments on this question, with 42% strongly agreeing and 33% agreeing (median = 4, mode = 5) indicating that they would prefer more text and less symbols in the policy. The findings for the group using the Privacy Table were more equally distributed across the scale (median = 3, mode = 2), with a majority disagreeing (33%) to the statement.

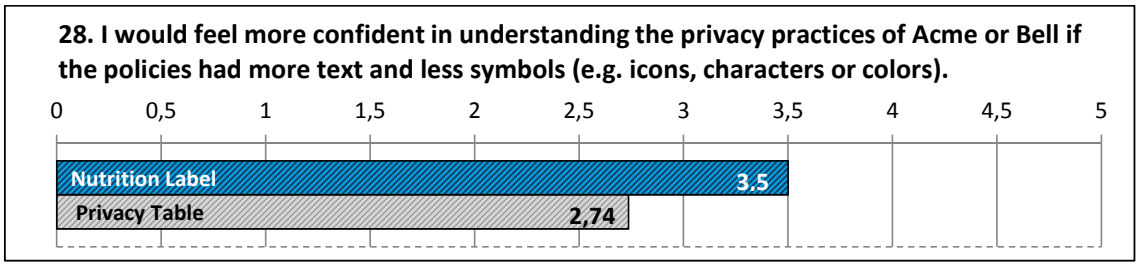


Figure 6.26: Mean scores for combined data samples for question 28 - Part 2.2

We also performed a t-test on a combined sample-set from both experiments, as seen in figure 6.26 (detailed t-test statistics in table F.7, appendix F.3). We found near-significant difference in the mean score for the combined Nutrition Label group ($M = 3.5, SD = 1.277, N = 20$) and the combined Privacy Table group ($M = 2.74, SD = 1.147, N = 19$) with $t_{(37)} = 1.959, p = 0.058$ ($\alpha = 0.05$). The overall statistics for the combined data set showed that a majority of the participants in the Nutrition Label group agreed to the statement (median = 4, mode = 4), while the majority in the Privacy Table group disagreed (median = 3, mode = 2).

6.5 Timing Results

To measure the time spent on finding answers it the policies, we added time-blocks before and after part 1 and 2 in the laboratory experiment. We averaged the mean time across participants per part and performed t-tests to see whether there was any significant difference between the mean scores. Full details regarding the t-tests and descriptive statistics for the mean times can be found in appendix F.4.

Part	Nutrition Label	Privacy Table
Part 1	5.25	5.00
Part 2	2.00	3.00
Aggr.	3.86	4.00

Table 6.3: Mean time values measured in minutes for Part 1, 2, and both combined in the laboratory experiment.

As seen in table 6.3, the Privacy Table group ($M = 5.0, SD = 0.577, N = 7$) spent a quarter of a minute less time on average than the Nutrition Label group ($M = 5.25, SD = 2.188, N = 8$)in part 1. Based on the preliminary F-test we assumed unequal variances ($F = 14.35, P = 0.002$), but did not find any significant difference between the mean times for the groups, with $t_{(8)} = 0.311, p = 0.764$ ($\alpha = 0.05$).

For part 2, the Nutrition Label group ($M = 2.0, SD = 0.894, N = 7$) spent around a minute less than the Privacy Table group ($M = 2.83, SD = 0.408, N = 6$) on average (The count($N$) differ from part 1 as one participant in each group did not write down the time for this part). Based on the preliminary F-test we assumed equal variances ($F = 4.8, P = 0.055$), and we found a near-significant difference between the mean times for the groups, with $t_{(10)} = -2.076, p = 0.065$ ($\alpha = 0.05$).

For both parts combined, the two groups spent similar times on the tasks. Based on the preliminary F-test we assumed unequal variances ($F = 3.78, P = 0.014$), but did not find any significant difference between the mean times for the Nutrition Label group ($M = 3.86, SD = 2.381, N = 15$) and the Privacy Table group ($M = 4.00, SD = 1.225, N = 13$), with $t_{(20)} = -0.198, p = 0.845$ ($\alpha = 0.05$).

We therefore reject hypothesis H2.

6.6 Open end questions

To capture opinions regarding the questionnaire format and general opinions regarding the policy format, two open ended questions were added at the end of the questionnaire. The first was "Do you have any comments or suggestions regarding the privacy policy format?" and the second was "Do you have any feedback regarding the format of this survey?".

Common for both formats and both experiments was comments that the participants did not understand what "demographic information" was. Several participants also commented that they found the experiment useful and interesting to participate in.

6.6.1 Nutrition Label feedback

The main feedback regarding the Nutrition Label was that some participants in the laboratory experiment wanted a better explanation of what the different information categories contained. One participant commented that "*Could use more details on what info is contained in the information types. For example that financial information includes credit card number and such. Possibility of partial collection, ie pharmacy stores prescriptions but not patient journals so some health info is stored. No way of knowing what companies mean by "health info" "financial info" etc.* " while another noted that "*There should be some examples for each type of information which is collected, and for the ways of using/sharing the information. It might not be clear which information goes to which category*"

As explained in Chapter 4, we chose not to include the contextual information page which was added to the latest version of the Nutrition Label in our experiment as our focus was on the design of the label structure itself. The feedback mentioned above does however indicate that the category definition part was a useful extension of the Nutrition Label.

Another participant commented that the "information sharing" part of the policy was not good enough, and that there should be more categories for how information could be shared. He stated confusion regarding that while information in the Acme policy could be shared on public forums without any opt-out options, it was necessary to opt-in for the same information to be shared to other companies. The participant argued that if information is shared to public forums, it would also be accessible by any company interested in acquiring it.

While this was a design flaw for the fictive Acme policy, the observations is important because it indicates that the two sharing categories might be depending on each other. While a policy might state that information is just shared to public forums, it would also automatically be shared to any company visiting these forums.

The participants using the Nutrition Label in the Internet experiment did not provide any feedback on the policy format.

6.6.2 Privacy Table feedback

6.6.2.1 Laboratory experiment

As mentioned in section 5.6.3 on page 97, several participants had problems understanding the opt-out/in system and the different symbols used in the laboratory experiment version of the Privacy Table. One participant commented that "*it should be different symbols/colors on the opt-in an opt-out symbols and not just in the text below the rows*", and another requested "*better coloring of the dots, to distinguish at a glance between opt in/out, yes/no*".

Some participants also reported confusion regarding the "*do you have any choice regarding the use of this information?*" row that indicated opt-in / opt-out options. On commented that it was "*Very hard to understand the opt in/out choises. Hard to understand what was default settings in the two different firms. Didnt feel like the "choises" in the bottom had anything to do with each column. Hard to understand. This is because of the thick line where the table ends.*"

Based on the feedback, we did some design changes to the Privacy Table prior to the Internet experiment. The details regarding the design changes were explained in section 5.6.3 on page 97.

6.6.2.2 Internet experiment

While there was less reports on issues regarding the symbols in the Internet experiment feedback, on participant commented that "*The information along the x axis as to policy (how we will use this information), would be better at the top of the table, in its own row. It seems unnecessary that I need to look at the top and bottom of the table. If the policies are to be truly comparable, and "opt in" and "opt out" are going to vary: I would prefer a separate symbol for "opt in" and another for "opt out"; alternatively I would expect "opt in" and "opt out" to be written in the table.*"

Another respondents requested symbol legends: "*One can never be certain of what a symbol means without a legend. You asked questions about trust and certainty, but without a legend that assures me that my interpretation of a symbol is the same as its intended use, I will not trust it as an agreement. A blank must also be clearly defined, if I am to trust that it means you will not be using the information*".

6.7 Key findings

The following section summarizes the most important findings from the information finding and likeability questions. The results are summarized in tables where each row represents a part or a question as seen in figure 6.27. Each row contains the mean scores for the NL (Nutrition Label) group and the PT (Privacy Table) group from both experiments, in addition to the p-value from the t-test. Green colored cells indicates a significant finding ($p<0.05$), yellow cells indicates a near-significant findning ($p<0.20$) and orange cells indicates a indication for difference between the mean scores ($p<0.30$). Significant and near-significant findings are described textually below each table.

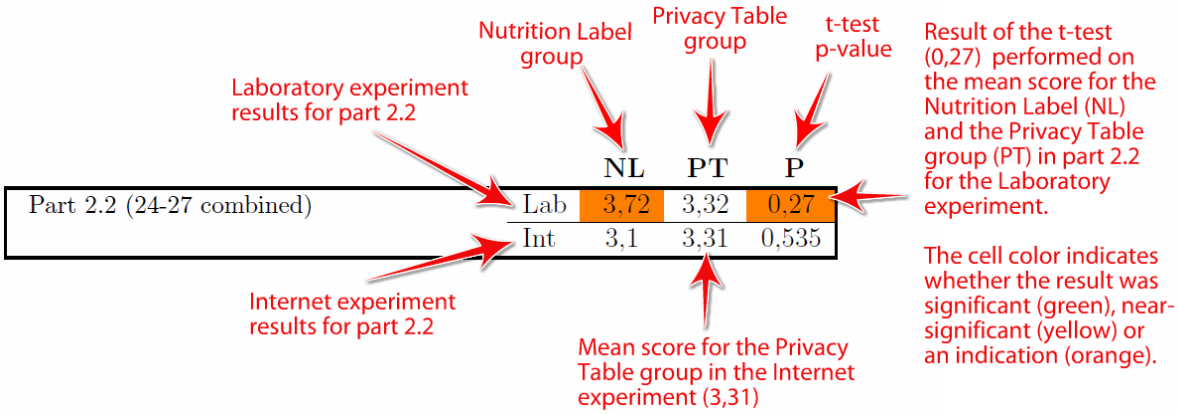


Figure 6.27: Explanation of key findings table

6.7.1 Information finding parts

Overall		NL	PT	P
Part 1.1	Lab	5,25	4,71	0,180
	Int	4,67	4,33	0,487
Part 1.2	Lab	3,5	2,57	0,148
	Int	2,67	2,75	0,866
Part 1.2 -	Lab	2,5	1,86	0,197
Question 7, 8 and 10	Int	2,17	1,92	0,568
Part 2.1	Lab	3,75	4	0,764
	Int	2,92	3,67	0,268
Part 1.1, 1.2 and 2.1	Lab	12,5	11,29	0,425
	Int	10,25	10,75	0,706

Table 6.4: Key findings in the information finding parts

The Nutrition Label group (mean = 5.25) in the laboratory experiment had a near-significant higher mean score than the Privacy Table group (mean = 4.71) in part

1.1.
($t_{(13)} = 1,418, p = 0.180$)

The Nutrition Label group (mean = 3.5) in the laboratory experiment had a near-significant higher mean score than the Privacy Table group (mean = 2.57) in part 1.2.
($t_{(13)} = 1.537, p = 0.148$)

The Nutrition Label group (mean = 2.5) in the laboratory experiment had a near-significant higher mean score than the Privacy Table group (mean = 1.86) for question 7, 8 and 10 in part 1.2. These questions all asked whether the collected information would be shared with any 3rd party.
($t_{(13)} = 1.36p = 0.197$)

We also found that only 5 out of a combined 39 participants for both experiments answered question 9 in part 1.2 correctly, and only 2 out of 39 answered question 12 in the same part correctly. Question 9 asked whether "the policy allow Acme to use your buying history to design custom functionality targeted at you?", while question 12 asked whether "Acme give you control regarding their sharing of your personal data?".

In part 2.1, the Privacy Table group had a higher mean score than the Nutrition Label group in both experiments. While we found no indications for any significant difference between the mean scores in the laboratory experiment (NL mean= 3.75, PT mean = 4.00), there was indications that the Privacy Table group (mean = 3.67) had a higher mean score than the Nutrition Label group (mean = 2.92) in the Internet experiment.

The Nutrition Label group (mean = 12.5) performed better than the Privacy Table group (mean = 11.29) in the laboratory experiment, but there were no indications of any significant difference between the mean scores. The results were opposite in the Internet experiment with the Privacy Table group (mean = 10.75) scoring higher than the Nutrition Label group (mean = 10.25), but again we found no indications of any significant difference between the means.

Part 1.1:	**Laboratory**	**Internet**	**P**
Question 1,2,3,5	3,87	3,29	0,006
Question 1-6	5	4,5	0,142

Table 6.5: Combined results (NL+PT) for each experiment

By combining the means for the two groups in each experiments, we found that the laboratory experiment participants (mean = 5.0) had a near-significant higher mean score ((t(37) = 1.644, p = 0.142) than the internet experiment participants (mean = 4.5) in part 1.1. We also found that the laboratory experiment participants (mean = 3.87) had a significant higher mean score (t(33) = 2.913, p = 0.0064) than the Internet experiment participants (mean = 3,29) on question 1,2,3 and 5, which was the simplest information finding questions.

6.7.2 Likeability question parts

Overall		NL	PT	P
Part 1.3 (13-17 combined)	Lab	3,45	3,49	0,914
	Int	2,48	3,17	0,074
Part 2.2 (24-27 combined)	Lab	3,72	3,32	0,27
	Int	3,1	3,31	0,535
Part1.3 + 2.2 (13-17 and 24-27 combined)	Lab	3,57	3,41	0,475
	Int	2,76	3,32	0,156

Table 6.6: Key findings in the likeability question parts

The Privacy Table group (mean = 3.17) in the Internet experiment had a near-significant higher mean likeability than the Nutrition Label group (mean = 2.48) in part 1.3 (except question 18) .
$(t_{(22)} = -1.879, p = 0.074)$

The Privacy Table group (mean = 3.57) in the Internet experiment had a near-significant higher mean likeability than the Nutrition Label group (mean = 2.76) in part 1.3 and 2.2 combined (except question 18 and 28)
$(t_{(22)} = -1.469, p = 0.156)$

Part 1.3:		NL	PT	P
13. I feel that Acmes privacy practices are explained thoroughly in the privacy policy I read	Lab	3,5	2,86	0,234
	Int	2,83	3,42	0,154
14. I had problems finding the information I was looking for in Acmes policy	Lab	2,5	2,57	0,895
	Int	3,58	3	0,175
15. I feel secure about sharing my personal information with Acme after viewing their privacy practices	Lab	3	2,71	0,566
	Int	2,17	3,08	0,07
16. When I first looked at Acmes policy, it was easy to understand what information they will collect from me	Lab	3,5	3,86	0,53
	Int	2,42	3	0,255
17. If all privacy policies looked just like this I would be more likely to read them	Lab	3,75	4,57	0,048
	Int	2,58	3,33	0,139
18. The use of symbols (e.g. icons, characters or colors) in Acmes policy made it confusing to understand how the information they collect will be used	Lab	2,5	2,57	0,902
	Int	2,83	2,67	0,689

Table 6.7: Key findings in likeability part 1.3

The Privacy Table group (mean = 3.42) in the Internet experiment agreed more to feeling that Acmes privacy practices were explained thoroughly in the policy than the Nutrition Label group (mean = 2.83).
(Question 13, near significant, $t_{(22)} = -1.477, p = 0.154$).

The Nutrition Label group (mean = 3.58) in the Internet experiment agreed more to having problems finding information in the policy than the Privacy Table group

(mean = 3).
(Question 14, near significant, $t_{(22)} = 1.402, p = 0.175$)

The Privacy Table group (mean = 3.08) in the Internet experiment agreed more to feeling secure about sharing their personal information after reading the policy than the Nutrition Label group (mean = 2.17).
(Question 15, near significant, $t_{(22)} = -1.904, p = 0.07$)

The Privacy Table group (mean = 4.57) in the Laboratory experiment agreed more to that they would read policies more often if they looked like this, than the Nutrition Label group (mean = 3.75).
(Question 17, significant, $t_{(13)} = -2.203, p = 0.048$)

The Privacy Table group (mean 3.33) in the Internet experiment also agreed more to that they would read policies more often if they looked like this, than the Nutrition Label group (mean = 2.58).
(Question 17, near significant, $t_{(22)} = -1.533, p = 0.139$)

Part 2.2:		**NL**	**PT**	**P**
24. I found comparing two policies an easy task to do	Lab	4,13	3,57	0,257
	Int	3,33	3,25	0,85
25. If all policies looked like this I would compare privacy practices across websites more often	Lab	3,5	3,71	0,666
	Int	3,08	3,5	0,27
26. It was easy to keep control of what the different symbols (e.g. icons/colors) in the policy stood for	Lab	3,88	3,29	0,297
	Int	2,83	3,33	0,216
27. I feel that these policies would cover most of the privacy concerns I would have if I were to sign up for ..	Lab	3,38	2,71	0,289
	Int	3,17	3,17	1
28. I would feel more confident in understanding the privacy practices of Acme or Bell if the policies had more text and less symbols (e.g. icons, characters or colors).	Lab	2,88	2,29	0,263
	Int	3,92	3	0,089
	I+L	3,5	2,74	0,058

Table 6.8: Key findings in likeability part 2.2

6.7.3 Timing results

	NL	**PT**	**P**
Part 1	5,25	5,00	0,764
Part 2	2,00	2,83	0,065
Part 1 and 2 combined	3,86	4,0	0,845

Table 6.9: Key findings for the timing results

The Nutrition Label group (mean = 2.00) in the laboratory experiment spent near-significant less time answering the comparison questions in part 2.1 than the Privacy Table group (mean = 2.83), measured in minutes.
(Near significant, $t_{(10)} = -2.076, p = 0.065$).

6.7.4 Hypotheses

We rejected hypothesis H0.
We rejected hypothesis H1.
We rejected hypothesis H2.
We rejected hypothesis H3.

CHAPTER 7

Discussion

Based on the initial discussion in the previous chapter regarding how the low number of participants affected the validity of the results, we will begin this chapter by a discussion on whether the data generation method was suitable for the purpose of comparing the two formats, or if an alternative method could have provided better results. Following this section, we will discuss the results in relation to the proposed design criteria list from the design chapter, and argue whether there are any indications for if the proposed design changes had any effect in terms of accuracy, timing data and likeability.

We will then discuss if there were any other important findings from the evaluation, compare our results to the results found in previous research on the Nutrition Label, and then present a final proposal of a standardized design based on our overall findings. We will end the chapter by discussing whether our work through this thesis has answered the problem definition outlined in Chapter 1, and in which ways our proposed artifact has contributed to further work on the area.

7.1 The choice of data generation method

As outlined in the preceding chapter, and further discussed through the subsequent sections, we only received indications for whether or not our proposed design changes had any effect in terms of accuracy, time and likeability. An important question to ask is therefore whether we chose the correct data generation method for the purpose of the experiment. The two formats we compared might have been too similar in order to find any significant differences through a questionnaire.

The combination of similar designs and few participants might have contributed to the indications being vague or contradictionary across the experiments, as further discussed in the subsequent sections. The issue with few participants is connected to the theory that it requires few participants to discover large differences, and many participants to discover small ones [Concort, 2010]. There were several indications for non-normality in the samples from the Laboratory experiment and while we

argued that the results in the likeability-sections seemed to be normally distributed, the groups consisting of 12 participants could also have been too small to discover whether the designs performed significantly different.

Small groups also provide for small margins of error, and the answers from one or two participants could greatly change the mean score for the entire group, resulting in misleading conclusions. It is however important to highlight that few participants was an unforeseen circumstance and that the evaluation method might have provided clearer evidence if the outcome had been based on 74 responses as planned. In groups of 35 students, each participant would have had less impact on the entire group and we would probably have been able to detect clearer evidence for whether the proposed design changes had any effect.

While the low attendance in the guest lecture and experiment was unexpected, we could have utilized strategies such as offering the students incentives to participate, or better informing them of the importance of participating beforehand (e.g. through the lecture the preceding week or through the course email list) in order to ensure a higher attendance. It should be mentioned that the course responsible did inform the students about the importance of showing up for the guest lecture and the experiment, but this seemed to have had little effect.

If the low attendance had been foreseen, alternative recruitment methods would have been limited due to a small budget for offering incentives, and not having permission to utilize university email-lists for this purpose. The number of participants for the Internet experiment also indicated that an initial online recruitment strategy most likely would have resulted in few participants as well.

An alternative if we had expected few participants could have been to consider another data generation method. The reasoning behind choosing a questionnaire as a data generation method was primarily that we wanted to measure the respondents accuracy, timing data and enjoyability of using the format, and secondly that we wanted to compare our results to those found in related research on the Nutrition Label. The questionnaires we reused were also well-designed and tested for the purpose, and by reusing the questions we also saved time which could be spent on the design process itself. It could be argued that the modifications we did to the questionnaire between the pre-test and the experiments diminished its validity in terms of being well-tested for the purpose. To better ensure the final questionnaire was well-designed for its purpose, we could have conducted another pre-test before the laboratory experiment, but due to limited time this was not considered.

Alternative data generation methods have been used for similar experiments previously. Kelley et al. [2009] utilized focus groups in their initial evaluation of the Nutrition Label, and while it was not a focus group per definition, we also had a good experience of a similar discussion session through our pre-test. Focus groups are a form of group interviews that capitalizes on communication between research participants in order to generate data [Kitzinger, 1995]. Participants are encouraged to discuss with each other, and the method is particularly relevant for collecting data

for open ended questions. While we would have been unable to collect statistical data for the comparison of the formats, such a focus group would have provided us with useful contextual knowledge of how the participants perceived the format. By simply presenting the participants with the two formats, and asking which design they preferred (and why) we could possibly have collected stronger indications for whether our design changes had any effect than through single likeability-questions.

Focus groups does however have some limitations, such as some participants not raising their voice due to the group situation, or participants affecting each other's opinions [Kitzinger, 1995]. By utilizing focus groups for our evaluation we would also not have been able to compare our results to previous findings from related research, which was a main argument for choosing to conduct an experiment using a questionnaire.

Another relevant data generation method could have been to conduct an observation. According to Oates [2006] observations can be used within a range of research strategies, including experiments and design science. Observations can be either overt (the people know they are observed) or covert (the people are unaware that they are being observed) [Oates, 2006]. An example of a covert observation could have been to create a fictive website, invited people to register for it, and then observed their use of the sites privacy policy through the registration process. Ideally, we could have seen whether a new policy design would have affected their choices in sharing data, and by asking some follow-up questions, either through an interview or a questionnaire, we could have captured their opinion of using the policy. However, given the low percentage of people that actually reads notices, this method would most likely have resulted in poor results.

Alternatively, we could have conducted an overt observation where we in beforehand informed the participant about the purpose of the research. An example of an overt observation could have been to present to the user the same questionnaire and policies as we used, and told him or her to think out loud while answering the information finding questions. While this has the disadvantages of the respondent possibly changing his or hers behavior as a result of being observed (the Hawthorn Effect) [Oates, 2006], we believe this method could have provided better results in terms of contextual feedback on how the user perceived the format.

Based on the above discussion we can conclude that given the similarities between the two designs and the strong focus on design and usability in the report, data generation methods such as focus groups or overt observations could have provided better feedback in terms of getting an understanding of how the respondent perceived the format. They would however not have provided any better results in terms of accuracy and timing data, which also were a part of the initial problem definition. Such methods would also have required a more careful planning of the experiment procedure, and finally we would have been unable to conduct them as a part of the guest lecture. We therefore believe that choosing a questionnaire as a data generation method was the best choice given the expectations regarding the participant recruitment, as well as for the purpose of evaluating the designs in terms

of accuracy, timing data and likeability.

7.2 The proposed design changes

The following section discusses the findings from the evaluation in relation to the proposed design criteria list in Chapter 3. As explained in the section regarding validity of the results in the preceding chapter, we will not use terms such as significant or near-significant through these discussions, but instead refer to the findings as indications of some degree as presented in table 7.1.

Result chapter		Discussion chapter
P<0.05	Significant finding	Strong indication
0,05<P<0.20	Near-significant finding	Weak indication
0.20<P<0.30	An indication that there might be a difference in means	Some indication

Table 7.1: Terms to the describe the findings from the result chapter

A key component throughout the design process of the Privacy Table was the initial criteria list, which was based on the analysis of the potential problems regarding the Nutrition Label. It laid the foundation for the initial design and was also used thoroughly through the planning stages of the evaluation. While we redesigned the Privacy Table through several iterations, all versions remained true to the initial criteria list. In order to discuss how our proposed Privacy Table performed in comparison to the Nutrition Label, the list of criteria therefore plays an important role.

The surrounding structure	The matrix structure
1 Providing a clear statement of what information types are collected	5 Removing the "third dimension" and replace its functionality in an alternative way
2 Making the row and column titles and headers more dominant	6 Decreasing the size and the amount of symbols in the matrix
3 Providing a clear starting point for reading the label	7 Reduce the amount of scrolling needed to understand symbols
4 Making it clearer which titles belongs to which columns and rows	8 Adding more white space
9 Reducing the number of axis in the label	

Table 7.2: The design criteria list from Chapter 3

Following the analysis of the Nutrition Label, we categorized the findings based on an identification of two aspects regarding its structure: the surrounding parts (e.g. the row and column headers and titles) and the core matrix content (e.g. the

symbols and the "third dimension"). As indicated by table 7.2, eight of the criteria were divided between the two, while the final criterion was relevant for both areas.

While all questions used in the evaluation was designed to measure either accuracy, time or likeability, some were also custom designed to test specific design differences between the formats. In addition to this, we also found some questions more relevant for one of the above mentioned structure areas, than for the other. For example, one of the design changes for the surrounding structure was to provide a clear starting point for reading the label, in order to improve the users' first impression. Given the first impression of the label most likely was connected to the first questions the user answered in the experiment, we mapped the simple information finding questions to be most relevant for the design changes regarding the surrounding structure, as seen in table 7.3.

The surrounding structure	The matrix structure
Accuracy and time spent on the simple information finding questions	Accuracy and time on the complex information finding and comparison questions
Mean likeability score in part 1	Mean likeability score in part 2
Likeability question 14 and 16	Likeability question 18, 26 and 28
	To some degree likeability question 14 and 16.

Table 7.3: Related parts and questions to each design area

The remaining questions and parts we found relevant for each area are presented in the same table. Based on the findings from these questions, the following sections discuss each of the structure areas in detail in order to see whether the related design changes had an effect or not on the performance of the formats.

7.2.1 The surrounding structure

The proposed design changes for the surrounding structure was: By adding an extra column for what information was collected (1), making the headers and titles more dominant (2), and making it clearer which titles belongs to which columns and rows (4) we provided a starting point for reading the label (3).

Based on this we expected it to be easier for new users to use the policy, which should result in better accuracy and less time spent on the simplest information finding questions. The more advanced questions were not as relevant, as we expected the respondent to be more used to the structure at that point. We also expected a higher likeability score on these parts, particularly on likeability question 14, which was designed to measure whether the respondent had any difficulties in finding information, and on question 16, which captured the participant's first impression to using the policy.

7.2.1.1 Accuracy and time

We found a weak indication for the Nutrition Label performing better than the Privacy Table in the laboratory experiment on the simple information finding questions. The result was however not backed in the Internet experiment, and the validity of the finding was also threatened by an indication for non-normal distribution in the sample data. The results also had small margins, with all participants in the laboratory experiment answering question 1-3 correctly. As question 4 and 6 also were more advanced than the remaining questions in this part (by also asking how the collected data would be used), we conclude that while there was weak indication for the Nutrition Label performing better in terms of accuracy on part 1.1, it does not indicate any difference in first impression between the formats.

There were no differences in the average time spent on the information finding questions in part 1 between the groups.

7.2.1.2 Likeability

We found a weak indication for that the Privacy Table group had a higher likeability mean than the Nutrition Label group on the single policy information finding questions. This might indicate that the Privacy Table was more enjoyable to use for a first time user, but the finding was not backed by the results in the Laboratory experiment where the formats had equal mean likeability.

Question 16 was more specifically aimed and directly asked whether it was easy to understand what information was collected at first sight. Again we found some indication for that the Privacy Table was more liked than the Nutrition Label in the Internet experiment. Similar results were found for question 14 which asked whether finding information in the policy was hard. We found a weak indication for the Nutrition Label participants finding the policy harder to use than the Privacy Table group. While both these findings indicate that the proposed design changes had an effect, they were not supported in the laboratory experiment where both groups had similar likeability scores.

Based on these indications we can conclude that the Privacy Table and the Nutrition Label was equally liked by both groups in the Laboratory experiment, while there was some indications for that the Privacy Table participants in the Internet experiment found the format more enjoyable to use than the Nutrition Label group.

7.2.1.3 Conclusion

For both experiments combined there was not enough evidence to give any conclusion for whether the design changes to the surrounding structure of the label had any effect. We did see that the Nutrition Label group had a higher accuracy than the Privacy Table group in the laboratory experiment, but these results could be

affected by non-normality in the data, as well as small sample sizes. We did however see that the Privacy Table was more liked than the Nutrition Label in the Internet experiment, with consistent results on both question 14 and 16, as well as on the average results for part 1.1. We therefore conclude with that there was some indications for that the introduction of the "What information is collected" column and the more dominant headers and titles had a positive effect on the Internet participants, and that it improved their first impression and on their overall enjoyability of using the format.

7.2.2 The matrix structure

The proposed design changes for the matrix structure was: By removing the third dimension (5) we automatically reduced complexity in axes' (9), and by replacing the symbols with easy to understand dots (6), we also removed the need to scroll down for legend descriptions (7). The reduced colors and symbols also provided more white space and breathing space in the label (8).

While designing the Privacy Table, we made a compromise between reduced accuracy (in terms of details presented) and improved usability by removing the Nutrition Labels "third dimension". This decision was based on the fact that most of the real life Nutrition Labels we retrieved through our initial search using the Privacy Finder search engine did not utilize the full potential of the structure, as outlined in chapter 4.2.2 (page 43). The Privacy Table would therefore be better suited to present an "average" nutrition label, rather than the more complex Acme and Bell policies used in the experiments.

With the Acme and Bell policies being more complex than the average Nutrition Label policy, we mean that both policies contained different sharing practices within each column. Only 3 out of 22 policies we retrieved from our search utilized this function. Through the same search we also found that the The Acme and Bell policies had a higher amolunt of total elements in the matrix (23 and 19 respectively) compared to the average of 17 on our search. Both the Acme and Bell policy also utilized all four symbols (collected, opt-in, opt-out, not collected), compared to only 2 out of the 22 policies in our search.

As discussed in section 5.3.2 on page 69, we chose to reuse the Acme policy from the study by Kelley et al. [2010] mainly because it was custom designed for several of the question which was reused. Creating a questionnaire based on the content of an "average" Nutrition Label would also have been a very challenging due to the low amount of symbols used, and the minor differences between real-life policies.

The third dimension functionality was replaced by a summarized column as seen in figure 7.1, showing a snippet of the Bell group policy from the laboratory experiment. The summarized column was further developed based on the feedback from the laboratory experiment as seen in table 7.2.

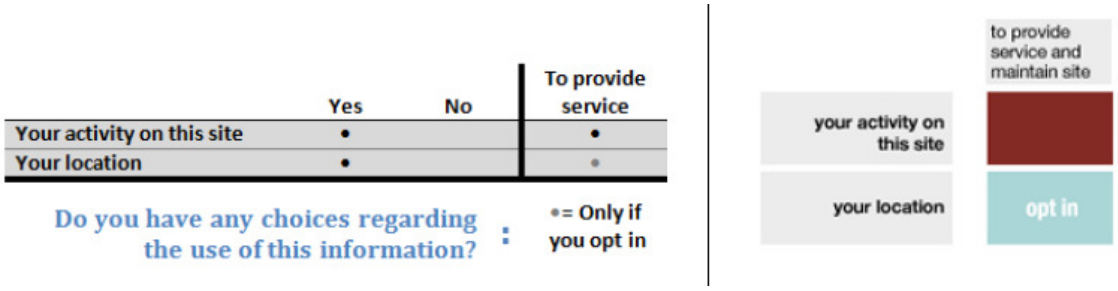


Figure 7.1: The summarized column in the laboratory experiment version of the Privacy Table (left) which replaced the Nutrition Labels "third dimension" (right).

While this system was designed to perform well on policies with similar sharing practices in each column, its usability was reduced when there were two (or more) different sharing practices within a single column, as was the case for both the Acme and Bell policy.

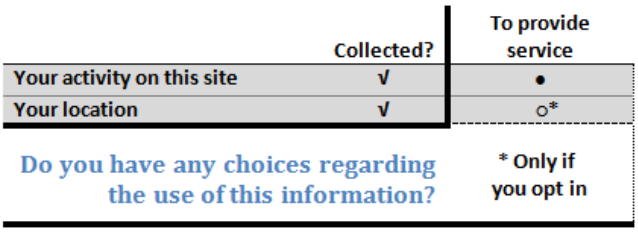


Figure 7.2: The summarized column in the Internet experiment version of the Privacy Table

Based on this, we therefore expected the Nutrition Label to perform marginally better than the Privacy Table on the complex information finding questions and in the comparison parts. We did however expect the Privacy Table to be more enjoyable to use, resulting in a higher mean likeability-score. As for the surrounding structure, we also expected a higher mean likeability score on question 14 and 16 for the matrix structure. We also custom designed three likeability questions (18, 26 and 28) to measure the general opinion of the balance between symbols and text in the policy.

7.2.2.1 Accuracy and time

As expected, there were weak indications for the Nutrition Label group scoring higher than the Privacy Table group on the complex information finding questions in the laboratory experiment. The finding was however not consistent across both experiments, as the Privacy Table scored marginally better in the Internet experiment. This might indicate that the changes we did to the design of the Privacy Table between the experiments was an improvement. As discussed previously, there were indications for a violation of the assumption of normality for the t-tests on these results, which might have biased the outcome.

One interesting finding we did do was that the Nutrition Label group scored higher than the Privacy Table on question 8 and 10, which both asked whether the collected information would be shared to another company. This was the only column to

make use of the "opt-in" symbol, and the difference in scores might indicate that the Privacy Table users misunderstood the meaning of the symbol and regarded it as "opt-out". The difference was however small and the results were not backed up by the findings from the Internet experiment, where both groups scored equally. As mentioned above, the difference between the experiments could again indicate that the changes we did to the Privacy Table (including a different design on the opt-in/out symbol) had an effect.

We did not find any strong indications for differences between the groups on the comparison parts, but we did see that the Privacy Table scored marginally better in both experiments, with the difference being highest in the Internet experiment. This might indicate that the reduced amount of symbols to keep control of, or the increased white space in the label, made it easier to compare the two policies against each other.

We did find a weak indication for that the Nutrition Label group spent less time than the Privacy Table group on the comparison tasks in part 2.1, but as we only measured in entire minutes, the margins were small with most participants spending either 2 or 3 minutes.

7.2.2.2 Likeability

In addition to likeability question 14 and 16 which was addressed in the section regarding the surrounding structure (7.2.1.2), we asked three custom designed likeability questions to measure the participant's opinion regarding the use of symbols in the policy. The first, question 18, asked whether the use of symbols in the Acme policy made it confusing to understand how the information they collect will be used. There were no indications for any differences between the groups, with most participants disagreeing in both experiments.

Question 26 was similar and asked whether it was hard to keep control of the different symbols in the policy. While the results for the Privacy Table groups were consistent across the experiments (means = 3.29/3.33) the mean score for the Nutrition Label groups score differed by over one unit on the likert-scale (means = 3.88/2.83) between the experiments. There were contradictionary results from the corresponding t-tests' for the experiments and therefore hard to conclude whether the reduced use of symbols in the Privacy Table had any effect or not on its performance compared to the Nutrition Label.

While likeability question 18 and 26 did not provide any strong indications for whether the design changes to the matrix structure had any effect, question 28 did provide some evidence. Question 28 asked whether the participant would feel more confident in understanding the privacy practices of Acme or Bell if the policies had more text and less symbols. 38% of the participants using the Nutrition Label in the laboratory experiment agreed to this statement, compared to none in the Privacy Table group. The result was backed by weak indications for a difference in mean

scores between the groups in the Internet experiment, with 42% of the participants in the Nutrition Label group agreeing.

An interesting aspect regarding question 28 is that the results could be interpreted in several ways. The high agreement in the Nutrition Label group could indicate that the participants preferred fewer and less dominant symbols and colors. But it could also be turned the other way around: as the mean score for the Privacy Table group in the Laboratory experiment was 2.29, with a majority disagreeing, it could also indicate that the participants wanted more symbols and less text in the Privacy Table design. As both groups had a mean score on the disagree side in the laboratory experiment, and on the agree side in the Internet experiment, the results could also indicate that the students in the Laboratory experiment would prefer symbols over text, and vice versa for the Internet participants.

While we have no evidence to prove that there was an indication for the latter, we can conclude that there were weak indications for that the Nutrition Label participants in the experiments would feel more confident if the policy had more text and less symbols. There was also a weak indication for the Privacy Table participants disagreeing to the same statement, which could indicate that they would prefer more symbols. This indication was also backed by the open-end questions, where one participant in the Privacy Table group reported that there should be different colors and symbols for the opt-in/out elements. Several other participants also commented that the summarized column did not function well, and that it should either be placed at the top of the table, or replaced by separate opt-in/out elements in the structure.

7.2.2.3 Conclusion

As for the surrounding structure, there is not enough evidence to conclude whether the design changes to the matrix structure had any effect on accuracy and likeability scores. Based on the initial discussion regarding the complexity of policies used, we can say that while the likeability score of the Privacy Table was higher than for the Nutrition Label, the difference in accuracy were not as big as expected. In fact, the Privacy Table scored marginally higher on the comparison tasks in both experiments, indicating that the compromised solution between improved usability and reduced accuracy did not reduce the Privacy Table participants' ability to answer complex information finding questions.

While the likeability score was higher for the Privacy Table, it might seem as a "middle way" between the Privacy Table and the Nutrition Label could be a better solution. The participants seemed to prefer the reduced amount of symbols and complexity of the Privacy Table, whilst finding the summarized columns confusing. A possible merging of the two solutions could therefore be to maintain the simple structure of the Privacy Table, whilst reintroducing a minimized version of the Nutrition Label's third dimension. An idea for such an solution is presented in section 7.5 on page 149.

7.3 Other findings

In addition to the findings related to the design criterias, we also discovered some other patterns in the result data.

7.3.1 Laboratory vs Internet participants

Through analyzing of the evaluation results, we did several observations where there was difference in scores between the two experiment groups. The biggest difference we found was that there was a strong indication for the laboratory participants performing better on question 1, 2, 3 and 5 than the Internet participants. We also found a weak indication for that they performed better on all simple information finding questions combined.

The difference in the information finding results can be due to several reasons. Firstly, the participants in the laboratory experiments were all students from technological study programs, while we had no control over the Internet population. As a majority of the Internet respondents did report having a bachelor-degree or higher, these differences might however not be due to education, but rather due to a motivational issue. It is possible that the participants in the laboratory experiments took the task more serious than the Internet respondents. Being in a silent laboratory environment they could also focus solely about the tasks, while the Internet participants could have been distracted or multitasking between the questionnaire and other tasks. A final reason might also have been that the questionnaire was more usable on paper than on Internet, where it was required to scroll up and down between the policy and the questions.

There were also differences between the experiments on the likeability questions for part 2. This could again indicate that the students in the laboratory experiment understood the format better than the participants in the Internet experiment, resulting in a higher satisfaction of using the policies.

It is important to highlight that these claims are solely based on general comparisons of the result data, and not based on any statistical finding. Given the small sample sizes across both experiments, and the fact that we had no control over the population in the Internet experiment, the findings are most likely due to other causes than suggested in the discussion above.

7.3.2 Difficult terms

One question stood out from the rest when we analyzed the results from the simple information finding questions. Question 4 asked whether the policy allowed Acme to use information about gender for marketing purposes, which very few participants answered correctly compared to question 1-3 and 5-6 in the same part. 54% of

the participants in the laboratory experiment and 36% of the Internet participants answered this correctly, which is a strong indication for a difference when the average rate of correct answers for the other questions in part 1.1 was 92% for the laboratory participants and 80% for the Internet experiment.

The most probable reason for the low score on this question is that the participants were unable to make the connection between gender and demographic information. The term "demographic information" might not be well known among non-native English speakers, and this suspicion was confirmed through the open-end questions were several participants reported that they did not understand what "demographic information" was.

A second term or concept the participants seemed to be unfamiliar with was profiling. Only 5 out of 39 participants in the experiments answered question 9 correctly. As with the demographic information question, this question also differed from the rest in its part. A closer analysis of the responses showed that most participants answered "yes" or "yes, unless i tell them not to" (opt-out) to this question, while the correct answer was "no".

7.4 Comparison to previous findings

An important reason for reusing questions in the questionnaire was to be able to directly compare our results to those found in previous experiments on the Nutrition Label. This allowed us to see how our Privacy Table performed in comparison to the previous findings, and also to see whether the previous findings on the Nutrition Label were consistent with our results. As we only compared the Privacy Table against the Nutrition Label in our experiments, it also enabled us compare it to the textual policy as well.

Since some questions were excluded, and some modified from the original questionnaires, we were only able to partly compare our results. No detailed results were published regarding the likeability questions from Kelley et al. [2009], and we were therefore only able to compare the information finding questions.

The figures in the following subsections presents the averaged results across both experiments for the groups using the Privacy Table and the Nutrition Label, as well as the results for the Nutrition Label and the full text policy in the study by Kelley et al. [2010]. We averaged the scores for each format across experiments for displaying purposes. As the study by Kelley et al. [2010] utilized two different policies per format (similar format, different content), we also averaged these results.

Grey and blue colored columns indicate the average results for the Privacy Table and the Nutrition Label groups from our experiment respectively. The orange and green colored columns indicates the result from the Nutrition Label and Full text policy groups from the study by Kelley et al. [2010] respectively.

7.4.1 Part 1.1

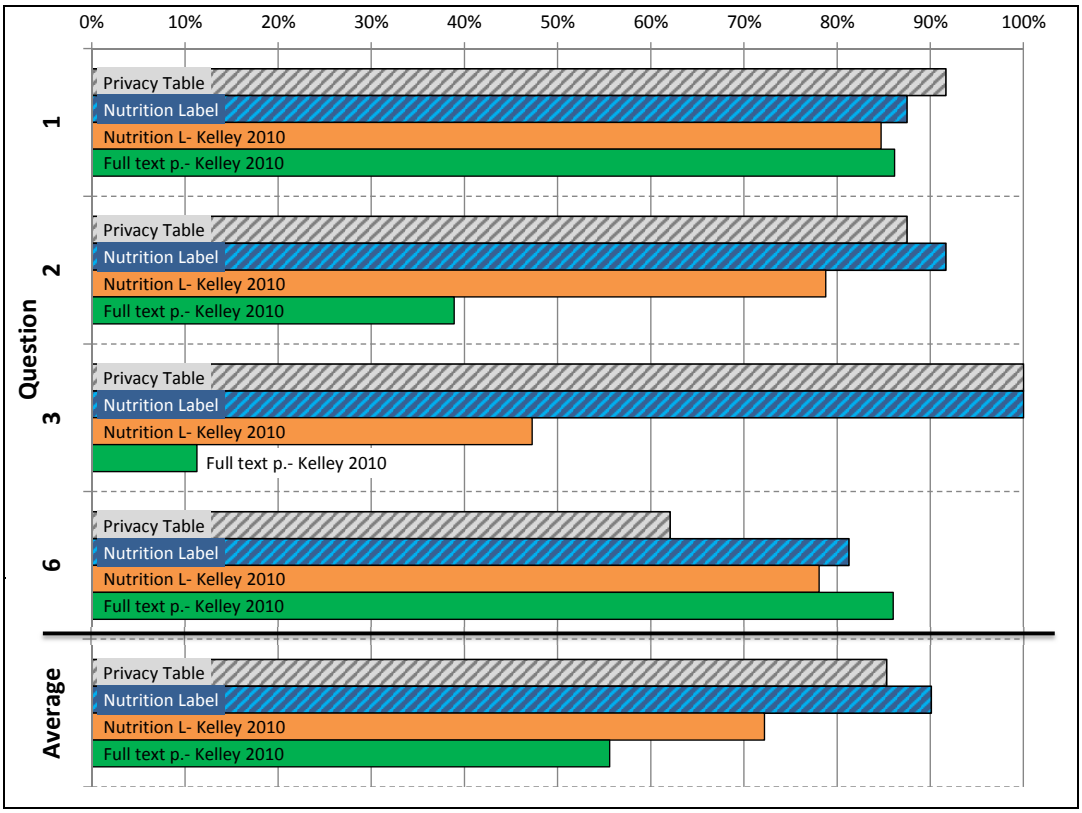


Figure 7.3: Comparison of questions in part 1.1

We reused four simple information finding questions from the Kelley et al. [2010] study, where question 2 was slightly modified (from "Acme might want to use your information to improve their website. Does this policy allow them to use your information to do so?" to "Does the policy allow Acme to use your information to improve their website?") as we wanted all questions in this part on a similar form.

As seen in the final row in figure 7.3, the average scores for the two formats in our experiment were higher than for the Nutrition Label and full text policy in the study by Kelley et al. [2010]. With the exception of question 6, the Privacy Table performed better than both groups in the [Kelley et al., 2010] study, and with the exception of question 3, the Nutrition Label group in our experiments performed similarly to the Nutrition Label group in the Kelley et al. [2010] study. Based on these findings we can assume that our Privacy Table most likely would perform better on simple information finding questions if compared to a full text policy, while there is too little evidence to say something about the comparison to the Nutrition Label data from [Kelley et al., 2010].

7.4.2 Part 1.2

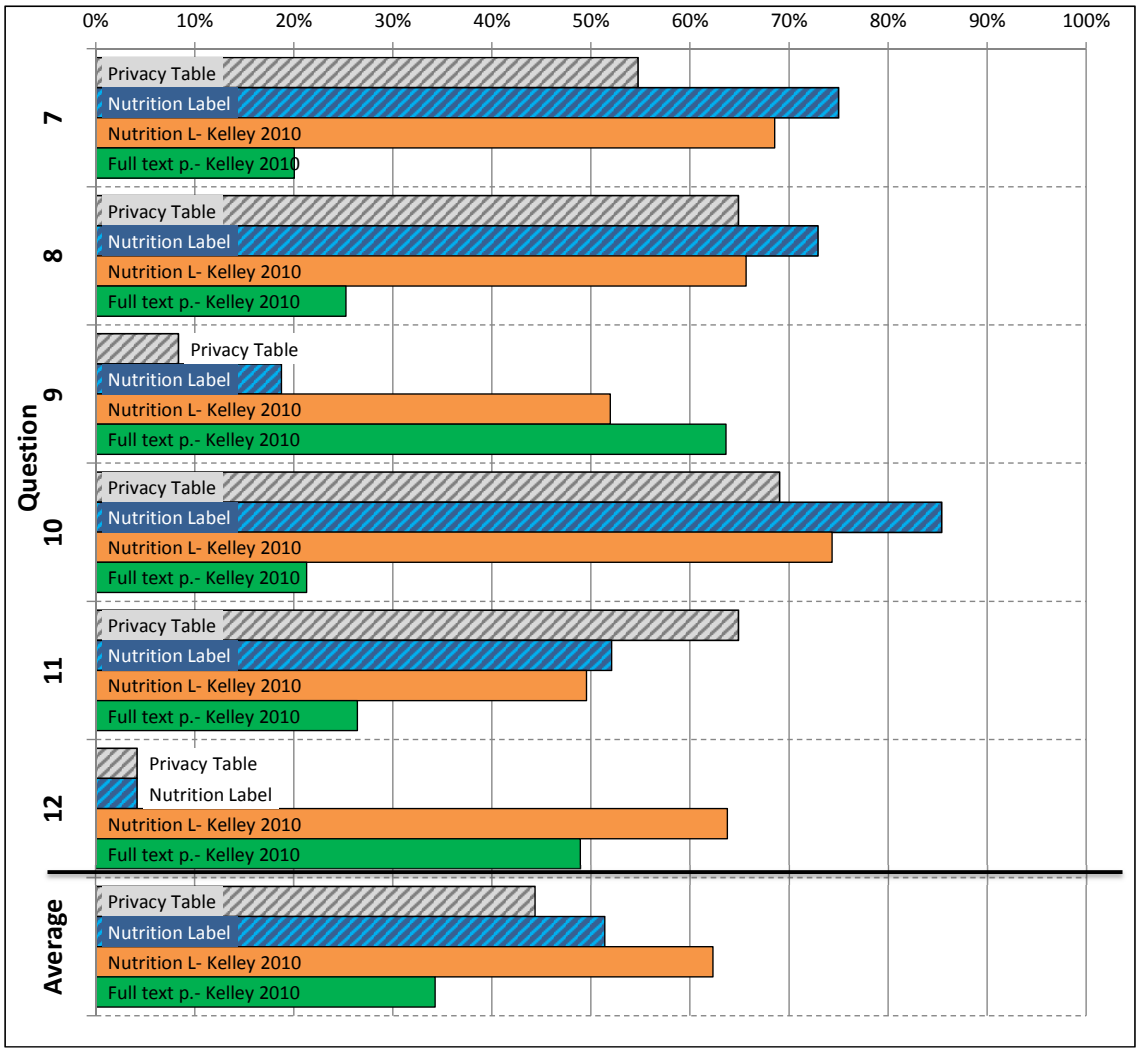


Figure 7.4: Comparison of questions in part 1.2

Figure 7.4 summarizes the findings from part 1.2 (the complex information finding questions). As for the simple information finding questions in part 1.1, we can see that the Privacy Table and the Nutrition Label groups in our experiment performed more or less similar to the Nutrition Label group in the Kelley et al. [2010] study on question 7, 8, 10 and 12 (averages: 63% (NL), 71% (PT), 65%(NL -Kelley) and 23% (Full text)). We did some changes to the terms in question 7 (public bulletin boards was changed to public forums) and question 8 (home phone number was changed to mobile number). Question 9 was discussed in section 7.3.2 on page 145, and we concluded that the participants did not understand the concept of profiling. Higher scores on this question in the Kelley et al. [2010] study indicates that "profiling" might be better known among native English speakers.

There were no issues with the terminology in question 12, and the only reason we can find for the low scores is that the question might have been vague. "Does Acme

give you control regarding their sharing of your personal data" could be interpreted in several ways depending on what information is regarded as "personal data". It is however interesting that a higher percentage of participants in the Kelley et al. [2010] study seemed to understand the question, compared to those in our experiments.

Based on this discussion we make the same assumptions for part 1.2 as we did for part 1.1; that the results from our Nutrition Label group was consistent with those from the Kelley et al. [2010] study, and that our Privacy Table most likely would perform better in a comparison study to a full text policy.

7.4.3 Part 2.1

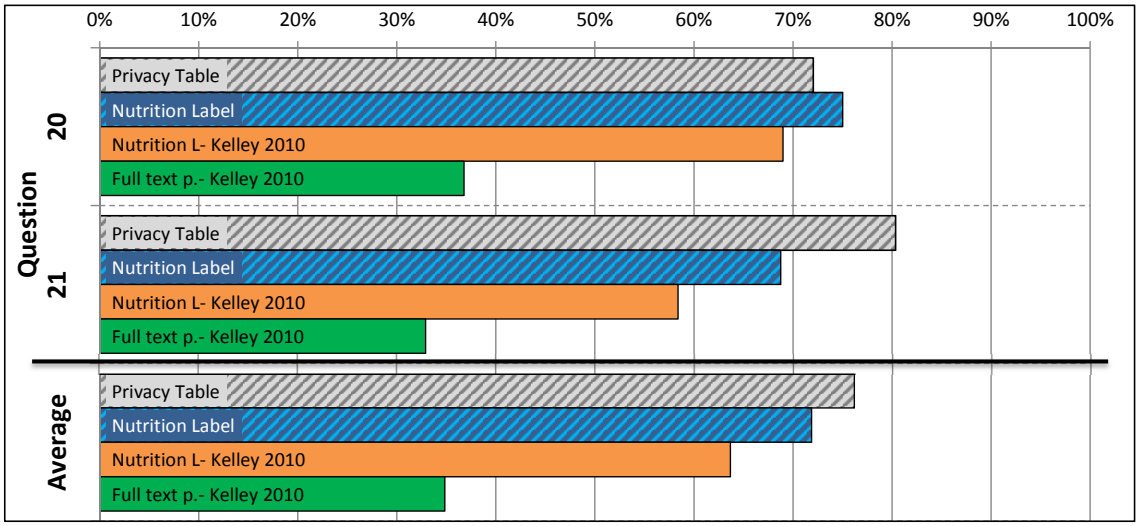


Figure 7.5: Comparison of questions in part 2.1.

Figure 7.5 shows the comparison for the two reused comparison questions. The results were consistent for both questions and provides for similar assumptions as described in the above comparison of part 1.1 and 1.2.

7.5 Further development

As a final part of the design cycle, and to summarize the findings from the evaluation, we present a merged version between the Privacy Table and the Nutrition Label in figure 7.6. This version is based on the following findings:

- For the surrounding structure there was little evidence to base any conclusion on. The indication we did find was that Privacy Table was generally more liked than the Nutrition Label in the internet experiment. In the same experiment there was also some indication for that the participants had a better first impression (question 14) and found the policy easier to use (question 16). We

therefore conclude with that there was some indications for that the changes the surrounding structure improved the respondents first impression of the label.

- For the matrix structure we concluded that it might seem as a middle way between the two formats could be a better solution. The participants seemed to prefer the reduced amount of symbols and complexity of the Privacy Table, whilst finding the summarized columns confusing.

Based on this we chose to keep the surrounding structure, and re-introduce the symbols from the Nutrition Label. The symbols have been made smaller than in the original Nutrition Label to save space and most importantly: to ensure the column and row headers and titles are still dominant.

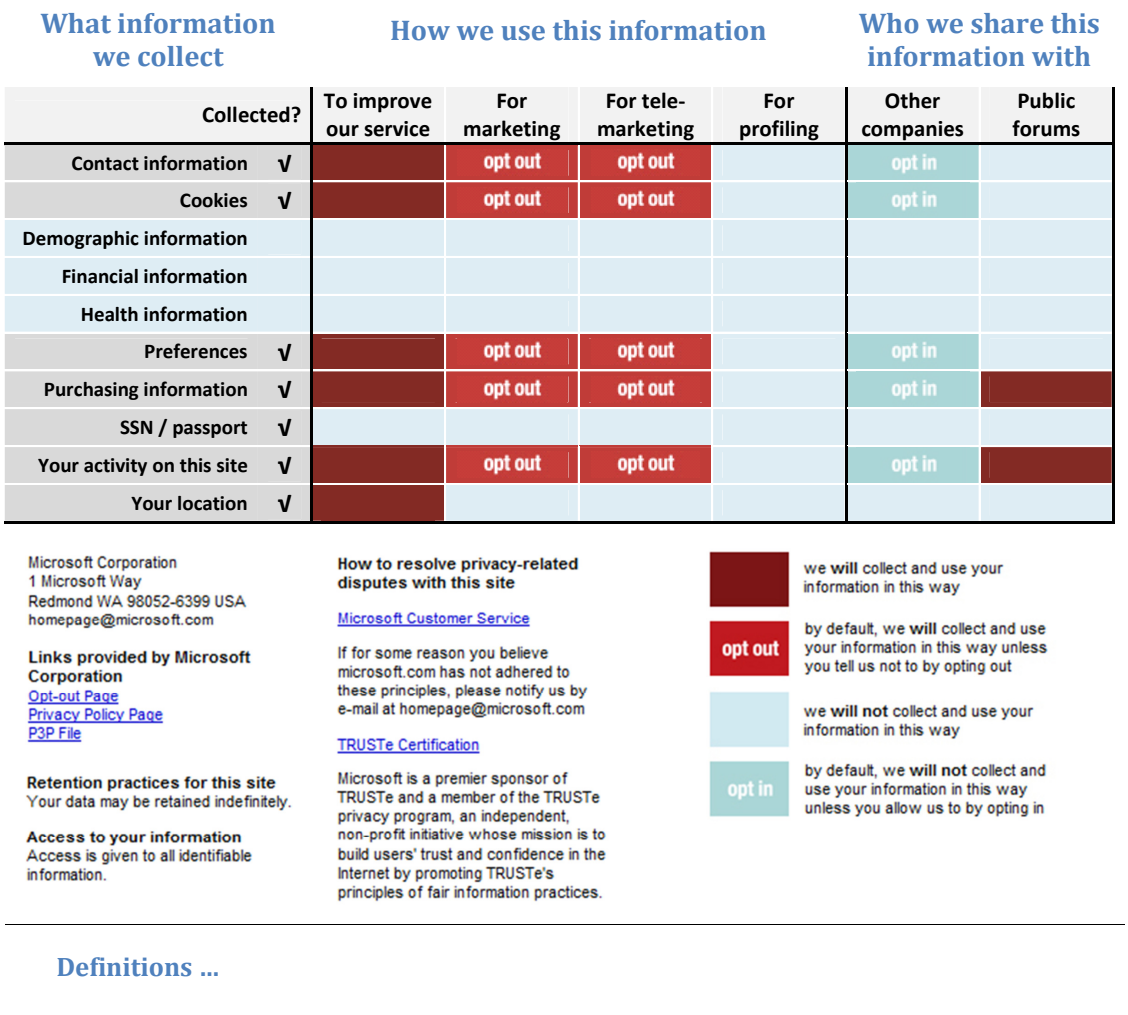


Figure 7.6: The merged version of the Nutrition Label and the Privacy Table

As seen in figure 7.6, we still provide a clear statement of what information types is collected or not (1), which also acts as a clear starting point for reading the label (3). The row and column headers are still dominant (2), which also aids in making it clearer which titles belongs to which columns (4).

We have also partly followed three of the design criteria for the matrix structure: By decreasing the size of the symbols in the matrix (6), we have reduced the amount of scrolling to understand the symbols (5), and the reduced matrix size has also resulted in more white space around the structure itself (8). As indicated in figure 7.6, there is more space below the structure, which could be used to provide contextual definitions on the same page as the policy itself.

We have not removed the "third dimension" (5), nor have we reduced the amount of symbols (7), and finally we have not reduced the number of axis in the label (9).

7.6 Findings in relation to the problem definition

We end this chapter by discussing whether we have addressed the initial problem definition, outlined in Chapter 1.

Our main goal of this report was to:
Find an alternative way of presenting a privacy policy, which could be more usable for the users, and aid in quickly and accurately finding the information they are looking for. To be useful for the COPE project, this solution should be suitable for use with a privacy agent, and be able to present important parts of the policy to the user in a simple but concrete way, based on the users' preferences.

We also outlined the following criteria for such a solution:

1. Make the general policy content more understandable,
2. Make the policy easier and more enjoyable to read
3. Highlight the most important privacy practices, while preserving details and nuances
4. Enable for easier comparison of privacy practices across sites
5. Be useful for the COPE project as well as current and future Privacy Enhancing Technologies (PETs)

Given our design strategy was to choose the current solution which best suited our list of criteria; we can split this discussion in two parts. Primarily, we need to discuss to which degree the Nutrition Label matched the outlined criteria listed above. Secondarily, we need to answer whether the design process which was based on

what we believed was the Nutrition Label's weaknesses, has resulted in an improved or weakened solution in relation to the same criteria list.

The discussion on whether the Nutrition Label is a good solution regarding the outlined list of criteria, is closely related to the work done to get an overview of current proposals and technologies for presenting privacy policies. We believe that the solid foundation which was laid through our in-depth study on privacy policies, together with a pre-study on current solutions as a part of this master thesis, provided us with enough knowledge to make a reasonable good decision.

The arguments for why we believed the Nutrition Label best addressed our problem definition was discussed in Chapter 3, and was mainly based on the Nutrition Label being an ideal "middle-way" between icons and text, and the fact that it has been repeatable evaluated with good results through several research projects.

Assuming we made a reasonable decision with the choice of the Nutrition Label, the second discussion is to address whether our newly proposed design is an improved version in relation to our criteria list. The development of a new solution was initiated by an analysis of what we believed was the weaknesses of the Nutrition Label structure. Based on this we created an alternative design, the Privacy Table, which was re-designed through several iterations and resulted in the merged proposal presented in the previous section. As the evaluation did not provide any clear evidence, and the validity of the results was threatened by few participants, it is hard to conclude with whether our design changes had the proposed effect. What we can say however is that, given the similar performance of the two formats in both experiments, and the consistent results compared to previous research, is that we assume the Privacy Table would perform better than a textual policy if compared in an experimentual setting.

In other words, based on the two discussions above, we believe that both the Privacy Table and the Nutrition Label represents good solutions to the initial problem definition. They both make the policy content more understandable, the policy itself easier and more enjoyable to read, and they enable for easier comparison of privacy practices across sites. Based on the thorough research behind the Nutrition Label, we also believe that both solutions captures the most important privacy practices, while preserving the details and nuances in each policy. We also believe we have contributed with useful knowledge for the further development of the Nutrition Label, as well as for other similar privacy enhancing technologies.

Finally, based on the discussion in Chapter 3 regarding the separation of the policy in three layers as proposed by the Creative Commons, we also believe that both the Privacy Table and the Nutrition Label could be useful as top layers for future proposals in privacy enhancing technologies. Given that both solutions are static representations of the privacy policy content, further work must however be conducted in order to develop top layer visualizations for preference-based privacy agents, which could be relevant for the COPE project.

CHAPTER 8

Conclusion and further work

We started this report with a problem definition of finding a new way of presenting privacy policies which could make the policy content more understandable, highlight the most important privacy practices while preserving details and nuances, enable for easier comparison of privacy practices and finally be useful for current and future Privacy Enhancing Technologies.

Through an initial pre-study, followed by an assess and refine cycle process based on a design science research strategy, we presented a final proposal for a merged version of the Nutrition Label and the Privacy Table, based on a discussion of our findings from the evaluation. While we were unable to provide any statistical evidence for the performance of our final solution, we still believe our work has resulted in findings which could be useful for further development on privacy enhancing technologies as a whole.

Through the pre-study, we found that a majority of the current approaches for presenting privacy policies matched the three-layer "Privacy Commons" proposed by Bickerstaff [2009]. We believe future research should take advantage of this structure, and develop unified solutions within each layer, rather than cross-layer solutions as the situation is today. A majority of the issues we discovered through the analysis of the Nutrition Label springs from its roots to the P3P language, which is outdated, unnecessarily complex and has resulted in the current prototype displaying labels that are far from as intuitive as the ones used to demonstrate the performance through previous research.

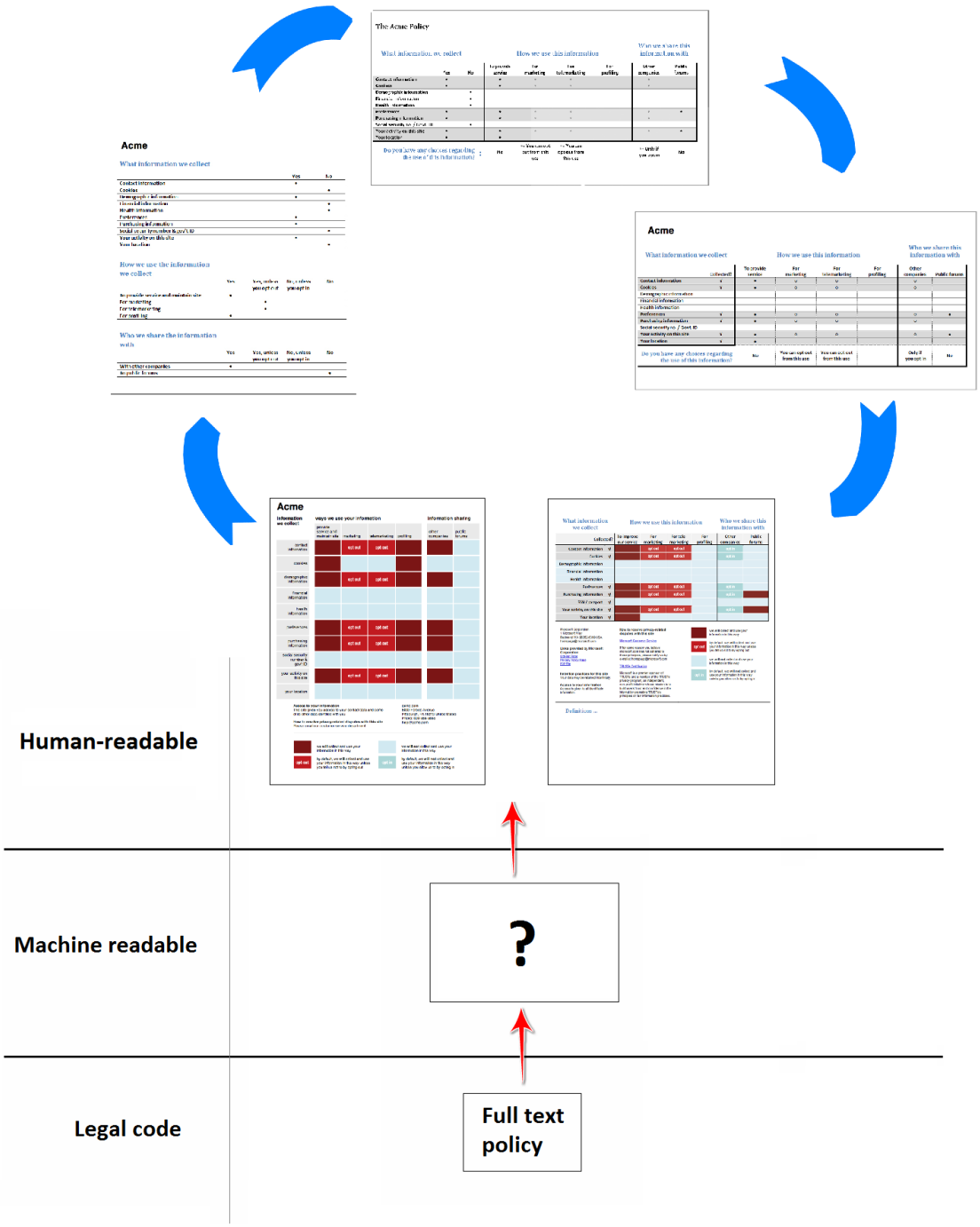
Being based on an analysis of these issues, we therefore believe our proposal for the Privacy Table could better serve the purpose of displaying P3P policies through the Privacy Finder search engine than the Nutrition Label itself, but that a combined solution could be advantageous if the Nutrition label were to free itself from its underlying technology. Such an independent visual representation could be further developed to be useful for both current and future privacy enhancing technologies, including preference based solutions. This would require a further analysis on how the current structure of the Privacy Table or the Nutrition Label could be modified to match a user's preference. Again, how these preferences are decided or which

technology they rely on, belongs to the machine-readable layer. By clearly separating these fields while at the same time ensuring interoperability, we believe the focus on future development on the Nutrition Label should be usability-related. In this setting, we believe our findings could be of high relevance for its further development.

We argue this by the findings from our evaluation which provided indications for that several of the assumed issues were correct. Most importantly we found that a more dominant header and title area, and less focus on the symbols, could improve usability of the label. This way, space is freed up, which in addition to improved usability also enables more space for contextual information. Several participants indicated that they would like more such contextual aid, which shows that the current development of the Nutrition Label is heading in a right direction. We are however unsure whether the current strategy of expanding the structure is the best solution, and suggest that future focus should be on further developing the core structure itself.

Based on this, we therefore believe we have provided an adequate response to our problem definition, and that our contributions could be useful not only for the further development of the Nutrition label, but also for other solutions based on the machine-readable layer, including the preference based privacy agent under development by Sintef ICT.

We end this master thesis by presenting the above discussed conclusion in terms of a figure which sums up the design process of the Privacy Table, as well as its position together with the Nutrition Label in the current privacy enhancing technology landscape.


Human-readable
Machine readable
Legal code
?
Full text policy

List of Figures

List of Tables

References

Amazon (2011). Amazon Mechanical Turk FAQ. "https://www.mturk.com/mturk/welcome".

Analyseit (2008). Blog - testing the assumption of normality. http://www.analyse-it.com/blog/2008/8/testing-the-assumption-of-normality.aspx.

Analyseit (2011). Documentation, summary statistics, histogram, box/-dot/mean/normal plot. http://www.analyse-it.com/support/documentation/220/standard/summary_statistics.htm.

Antón, A., Bertino, E., Li, N., and Yu, T. (2007). A roadmap for comprehensive online privacy policy management. *Communications of the ACM*, 50(7):116.

Beard, T., Nowson, C., and Riley, M. (2007). Traffic-light food labels. *Med J Aust.*

Bickerstaff, R. (2009). Towards a Commons Approach to Online Privacy for Social Networking Services, a Privacy Commons. "http://www.ico.gov.uk/upload/documents/pio_conference_2009/roger_bickerstaff_birdandbird_presentation.pdf".

Chaparro, B., Shaikh, D., and Baker, R. (2005). Reading Online Text with a Poor Layout: Is Performance Worse? *Usability News*, 7(1):1–4.

Cohen, L., Manion, L., Morrison, K., and Morrison, K. (2007). *Research methods in education*. Routledge.

Concort (2010). Consort statement 2010. http://www.consort-statement.org/consort-statement/overview0/.

Creative-Commons (2011). Creative commons licenses. http://creativecommons.org/licenses/.

CSPA (2011). Cspa's online courses: Wading through the data swamp. http://pathwayscourses.samhsa.gov/eval201/eval201_toc.htm.

Drichoutis, A., Lazaridis, P., and Nayga, R. (2006). Consumers use of nutritional labels: a review of research studies and issues. *Acad Marketing Sci Rev, no. 9.*

Field, A. and Hole, G. (2003). *How to design and report experiments*. Sage publications Ltd.

Fink, A. and Litwin, M. (1995). *The Survey Kit: How to measure survey reliability and validity*. Sage Publications.

Fischer-Hubner, S., Uld, K. S., Bussard, L., and Verdicchio, M. (2010). Ui prototypes: Policy administration and presentation version 1 members of the primelife consortium. *Framework*, pages 1–67.

Flanagan-Hyde, P. (2011). Is it normal? http://apcentral.collegeboard.com/apc/members/courses/teachers_corner/37042.html?type=print.

Gomez, J., Pinnick, T., and Soltani, A. (2009). KnowPrivacy. http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf (Retrieved: April 27 2010).

Govani, T. and Pashley, H. (2005). Student awareness of the privacy implications when using facebook. http://lorrie.cranor.org/courses/fa05/tubzhlp.pdf (Retrieved: April 27 2010).

Hansen, M. (2009). Putting privacy pictograms into practice - a european perspective. In Fischer, S., Maehle, E., and Reischuk, R., editors, *GI Jahrestagung*, volume 154 of *LNI*, pages 1703–1716. GI.

Helton, A. (2009). Privacy Commons Icon Set. "http://aaronhelton.wordpress.com/2009/02/20/privacy-commons-icon-set/".

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1):75–105.

Holtz, L.-E., Nocun, K., and Hansen, M. (2011). Towards displaying privacy information with icons. In Fischer-Hbner, S., Duquenoy, P., Hansen, M., Leenes, R., and Zhang, G., editors, *Privacy and Identity Management for Life*, volume 352 of *IFIP Advances in Information and Communication Technology*, pages 338–348. Springer Boston. 10.1007/978-3-642-20769-3_27.

Hunton and Williams (2005). Ten steps to develop a multilayered privacy notice. "http://www.informationpolicycentre.com/files/Uploads/Documents/Centre/Ten_Steps_whitepaper.pdf".

Jensen, C. and Potts, C. (2004). Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 471–478. ACM.

J.S., D., G., L., , and J., W. (2009). Strategies for Promoting Healthier Food Choices. *American Economic Review*.

KCG (2006). Evolution of a Prototype Financial Privacy Notice. http://www.ftc.gov/privacy/privacyinitiatives/ftcfinalreport060228.pdf (Retrieved: May 8 2010).

Kelley, P. (2009). Designing a privacy label: assisting consumer understanding of online privacy practices. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 3347–3352. ACM.

Kelley, P. G. (2011). Designing Privacy Interfaces - Thesis proposal. "http://patrickgagekelley.com/dpi/proposal-main.pdf".

Kelley, P. G., Bresee, J., Cranor, L. F., and Reeder, R. W. (2009). A "nutrition label" for privacy. *Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09*, 2009. http://portal.acm.org/citation.cfm?doid=1572532.1572538 (Retrieved: May 1 2010).

Kelley, P. G., Cesca, L., Bresee, J., and Cranor, L. F. (2010). Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 1573–1582, New York, NY, USA. ACM.

Kelley, P. G., Cesca, L., Bresee, J., and Cranor, L. F. (2011). Presentation of an online study of the nutrition label approach. http://cups.cs.cmu.edu/privacyLabel/files/CHI-privacyFinal2010.

Kitchenham, B. A. and Pfleeger, S. L. (2008). Personal opinion surveys. In Shull, F., Singer, J., and Sjøberg, D. I. K., editors, *Guide to Advanced Empirical Software Engineering*, pages 63–92. Springer London.

Kitzinger, J. (1995). Qualitative research. introducing focus groups. *BMJ British Medical Journal*, 311:299–302.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1):537–567.

Lillebo, O. K. (2010). Privacy policies role in todays privacy policies landscape, and how they address users privacy concers.

Lin, D.-Y. M. (2004). Evaluating older adults' retention in hypertext perusal: impacts of presentation media as a function of text topology. *Computers in Human Behavior*, 20(4):491–503.

March, S. T. and Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4):251 – 266.

McDonald, A. M., Reeder, R. W., Kelley, P. G., and Cranor, L. F. (2009). A comparative study of online privacy policies and formats. *Privacy Enhancing Technologies*, 1998:1.

Mehldau, M. (2007). Iconset for Data-Privacy Declarations v0.1. "http://netzpolitik.org/wp-upload/data-privacy-icons-v01.pdf".

Oates, B. (2006). *Researching Information Systems and computing*. Sage Publications Ltd.

Pinnick, T. (2011). Blog post - Layered Policy and Short Notice Design. "http://www.truste.com/blog/?p=1594".

PrivacyFinder (2011). Privacy Finder FAQ. "http://www.privacyfinder.org/faq".

Reeder, R. W., Kelley, P. G., McDonald, A., and Cranor, L. F. (2008). A user study of the expandable grid applied to P3P privacy policy visualization. *Privacy in the electronic*.

Reips, U.-D. (2002). Internet-based psychological experimenting: Five dos and five donts. *Social Science Computer Review*, 20(3):241–249.

Rifon, N., LaRose, R., and Choi, S. (2005). Your privacy is sealed: Effects of web privacy seals on trust and personal disclosures. *Journal of Consumer Affairs*, 39(2):339–362.

Rundle, M. (2006). International Data Protection and Digital Identity Management Tools. "http://identityproject.lse.ac.uk/mary.pdf".

Seymore, J., Lazarus Yaroch, A., M., S., Blanck, H., and Khan, L. (2004). Impact of nutrition environmental interventions on point-of-purchase behavior in adults a review. *Preventative Medicine*.

Simon, H. A. (1996). *The Sciences of the Artificial*, volume 3rd. MIT Press.

Solutions, S. (2011). Testing of assumptions. http://www.statisticssolutions.com/resources/directory-of-statistical-analyses/testing-of-assumptions.

Tang, Z., Hu, J., and Smith, M. (2005). Protecting online privacy: self-regulation, mandatory standards, or caveat emptor. In *Proceedings of the Fourth Annual Workshop on Economics and Information Security*.

Trochim, W. M. K. (2006). *The Research Methods Knowledge Base*. Atomic Dog Publishing.

W3C (2007). P3P 1.0 Implementations. http://www.w3.org/P3P/implementations.html (Retrieved: April 29 2010).

Wang, Y. and Kobsa, A. (2008). *Privacy-Enhancing Technologies*, pages 352–375. Idea Group Inc (IGI).

APPENDIX A

Nutrition Labels from the Privacy Finder search

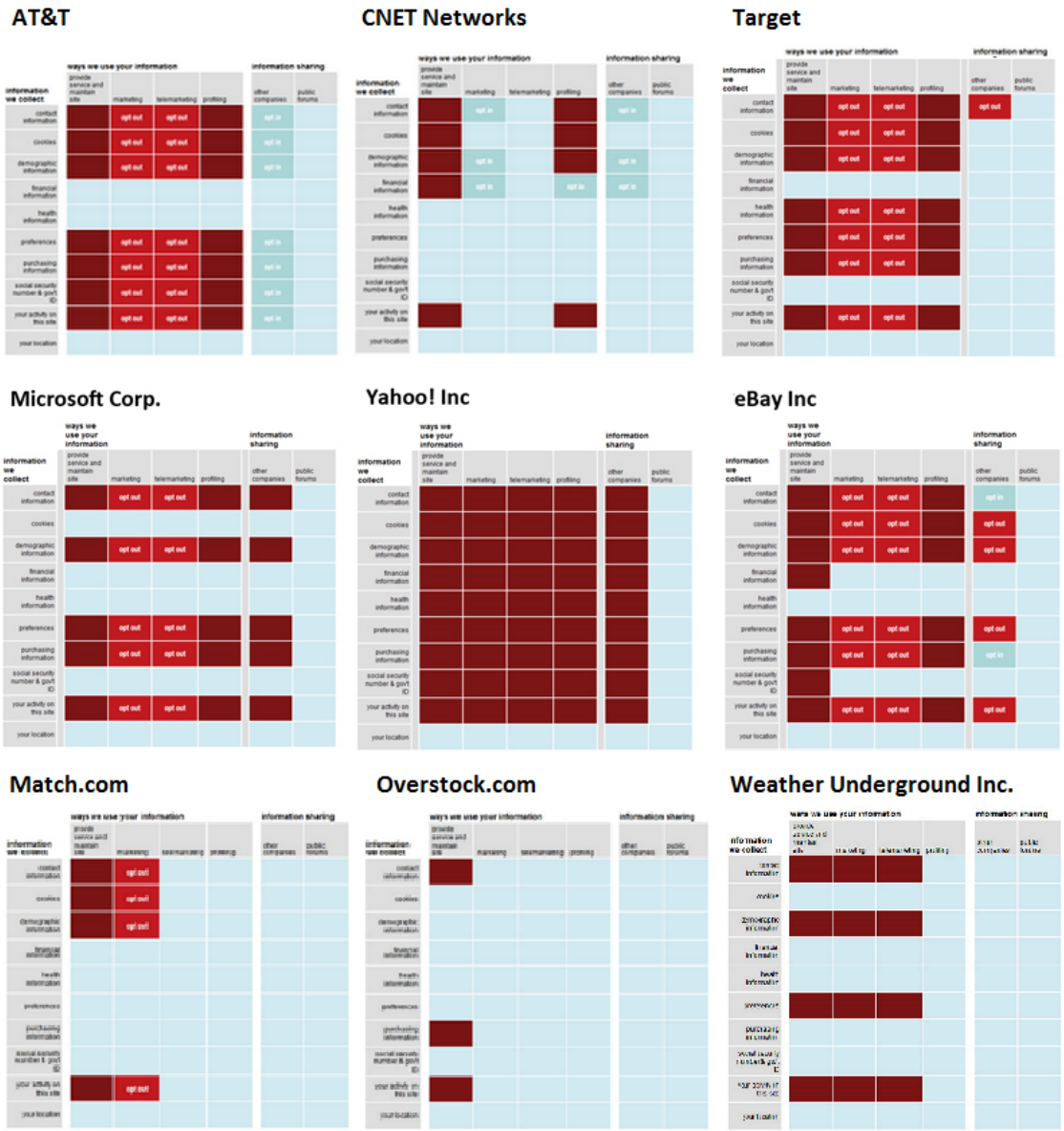


Figure A.1: Nutrition Labels from the Privacy Finder search (1/2) - The Microsoft policy was the same for Microsoft.com, Windows.com and Live.com. while the Yahoo! Inc policy was the same for Yahoo.com and Flickr.com

Figure A.2: Nutrition Labels from the Privacy Finder search (2/2)

APPENDIX B

Related experiments

- **Table B.1** - Overview of related experiments on the P3P expandable grid / Nutrition Label
- **Table B.2** - Questions from Kellet et al.[2009]
- **Table B.3** - Questions from Kellet et al.[2010]

Paper	Type	Type of design	Policy types	Procedure	Variables
Reeder2008	Web-study	Between-participant	P3P Expandable grid	786 participants	Accuracy
Expandable grid P3P			Natural language -	7 comprehension q's	Time
			(Short, medium, long)	8 satisfaction q's	Enjoyment
				Two factors:	
				- presentation format	
				- policy length	
Reeder2008	Lab-study	Within-participant	P3P Expandable grid	12 participants	Accuracy
Expandable grid P3P			Natural language -	7 comprehension q's	
			(Short, long)	Half viewed P3P first,	
				then the natural lang.	
				Participants thinking loud	
Kelley2009	Lab-study	Within-participant	Nutrition label	24 participants	Accuracy
A "Nutrition Label"			Natural language	45 min study	Time
for privacy				8 information finding q's	Enjoyment
				6 single policy likert q's	
				6 comparison q's	
				4 comparison likert q's	
Cranor2009	Web-study	Between-participant	Natural language	749 participants	Accuracy
Study of policy			Privacy finder	15 conditions -	Time
formats			Layered policies	6 NL, 6 PF, 3 layered	Enjoyment
				Questions randomized	
Kelley2010	Web-study	Between-participant	Nutrition label	761 participants	Accuracy
An online study of			Short nutr. label	5 categories	Time
the nutrition label			Natural language	- 6 simple tasks (single)	Enjoyment
approach			Stand. short text	- 6 complex tasks (single)	
			Layered text	- 7 single policy likert	
				- 5 comparison tasks	
				- 5 comparison likert	

Table B.1: Overview of related experiments on the P3P expandable grid / Nutrition Label

	Part 1	Question reused?	
	Information finding questions	**Pre-test**	**Final**
1.	Does the policy allow the Acme website to use cookies?	–	–
2.	Does the policy allow the Acme website to share your information on public bulletin boards?	●	●
3.	By default, does the policy allow the Acme website to collect your email address and use it for marketing?	–	–
	Part 2	**Question reused?**	
	Perceived Privacy Policy Understanding	**Pre-test**	**Final**
4	I feel secure about sharing my personal information with Acme after viewing their privacy practices	●	●
5	I feel that Acmes privacy practices are explained thoroughly in the privacy policy I read	●	●
6	Finding information in Acmes privacy policy was a pleasurable experience	●	–
7	I feel confident in my understanding of what I read of Acmes privacy policy	●	–
8	It was hard to find information in Acmes policy	●	●
9	If all privacy policies looked just like this I would be more likely to read them	●	●
	Part 3	**Question reused?**	
	Policy Comparison Questions	**Pre-test**	**Final**
10	By default, Button Co. can share information about your purchases with other companies, but Acme cannot.	●	●
11	Which company will better protect your information online? ccc cccc cccc ccc	–	–
12	Youre looking to buy a gift online. At which company would you prefer to shop?	–	–
	Part 4	**Question reused?**	
	Policy Comparison Enjoyment & Ease	**Pre-test**	**Final**
13	Looking at policies to find information was an enjoyable experience	–	–
14	Looking at policies to find information was easy to do	–	–
15	Comparing two policies was an enjoyable experience	●	–
16	Comparing two policies was easy to do	●	●

Table B.2: Questions from Kellet et al.[2009]

	Part 3	Question reused?	
	Single policy simple tasks	**Pre-test**	**Final**
1.	Does the policy allow Acme to collect information about which pages you visited on this web site?		
2.	Acme might want to use your information to improve their website. Does this policy allow them to use your information to do so?		
3.	Does the policy allow Acme to collect information about your current location?		
4.	Based on the policy will Acme register their secure certificate with VeriSign or some other company?		
5.	Based on the policy may Acme store cookies on your computer?		
6.	Does the policy allow Acme to collect information about your medical conditions, drug prescriptions, or family health history?		

	Part 4	Question reused?	
	Single policy complex tasks	**Pre-test**	**Final**
7.	Does the policy allow Acme to share some of your information on public bulletin boards?		
8.	Does the policy allow Acme to share your home phone number with other companies?		
9.	Does the policy allow Acme to use your buying history to design custom functionality targeted at you?		
10.	Does the policy allow Acme to share your cookie information with other companies?		
11.	Will Acme contact you with advertisements?		
12.	Does Acme give you control regarding their sharing of your personal data?		

	Part 6	Question reused?	
	Policy Comparison tasks	**Pre-test**	**Final**
14.	Does either company give you options with regards to cookies?		
15.	Does either company collect sensitive information (such as banking or medical records)?		
16.	By default, Acme can collect information about your age and gender in order to market to you by email, but the Bell Group cannot.		

Table B.3: Questions from Kellet et al.[2010]

APPENDIX C

Pre-test

C.1 Questionnaire

Part 1

In this part you will be answering 3 sets of questions regarding "Acme's" privacy policy, found on the next page.

Please remember to fill in the exact time for when you start and finish part 1.1 and part 1.2.

Part 1.1

Please answer the following questions regarding the Acme policy

Start time: :

	Yes	No	Does not say
1. Does the policy allow Acme to collect information about which pages you visited on this web site?	❑	❑	❑
2. Acme might want to use your information to improve their website. Does this policy allow them to use your information to do so?	❑	❑	❑
3. Does the policy allow Acme to collect information about your current location?	❑	❑	❑
4. Based on the policy, may Acme store cookies on your computer?	❑	❑	❑
5. Does the policy allow Acme to collect information about your medical conditions, drug prescriptions, or family health history?	❑	❑	❑
6. Does the policy allow Acme to collect information regarding your household income?	❑	❑	❑

End time: :

Part 1.2

Please answer the following questions regarding the Acme policy

Start time: :

	Yes	Yes, unless I tell them not to	Only if I allow them to	No	The policy does not say
7. Does the policy allow Acme to share some of your information on public bulletin boards?	❑	❑	❑	❑	❑
8. Does the policy allow Acme to share your mobile number with other companies?	❑	❑	❑	❑	❑
9. Does the policy allow Acme to use your buying history to design custom functionality targeted at you?	❑	❑	❑	❑	❑
10. Does the policy allow Acme to share your cookie information with other companies?	❑	❑	❑	❑	❑

11. Will Acme contact you with advertisements?	❑	❑	❑	❑	❑
12. Does Acme give you control regarding their sharing of your personal data?	❑	❑	❑	❑	❑

End time: :

Part 1.3

To what extent do you agree or disagree with each of the following statements?

	Strongly agree	**Agree**	**Neither agree nor disagree**	**Disagree**	**Disagree strongly**
13. I feel secure about sharing my personal information with Acme after viewing their privacy practices	❑	❑	❑	❑	❑
14. I feel that Acme's privacy practices are explained thoroughly in the privacy policy I read	❑	❑	❑	❑	❑
15. Finding information in Acme's privacy policy was a pleasurable experience	❑	❑	❑	❑	❑
16. I feel confident in my understanding of what I read of Acme's privacy policy	❑	❑	❑	❑	❑
17. It was hard to find information in Acme's policy	❑	❑	❑	❑	❑
18. If all privacy policies looked just like this I would be more likely to read them	❑	❑	❑	❑	❑

Part 2

In this part you will be answering questions which require you to compare two policies against each other. The two policies (Bell and Acme) can be found on the last page of this survey.

Please remember to fill in the exact time for when you start and finish part 2.1.

Part 2.1

Please answer the following questions regarding the Acme policy

 Start time: :

	Yes	No	Does not say
19. Does either company give you options with regards to cookies?	❑	❑	❑
20. Does either company collect sensitive information (such as banking or medical records)?	❑	❑	❑
21. By default, Acme can collect information about your age and gender in order to market to you by email, but Bell cannot.	❑	❑	❑
22. By default, Bell can share information about your purchases with other companies, but Acme cannot	❑	❑	❑
23. By default, both companies collects your default location and can use this information to improve their services	❑	❑	❑

 End time: :

Part 2.2

To what extent do you agree or disagree with each of the following statements?

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Disagree strongly
24. Comparing two policies was an enjoyable experience	❑	❑	❑	❑	❑
25. Comparing two policies was easy to do.	❑	❑	❑	❑	❑
26. If all policies looked like this I would compare privacy practices across websites more often	❑	❑	❑	❑	❑

C.2 Policies

Figure C.1: The Nutrition Label for part 1 in the pre-test

The Acme policy

What information we collect

	Yes	No
Contact information	•	
Cookies	•	
Demographic information		•
Financial information		•
Health information		•
Preferences	•	
Purchasing information	•	
SSN / passport	•	
Your activity on this site	•	
Your location		•

How we use the information we collect

	Yes	Yes, unless you opt out	No, unless you opt in	No
To provide service and maintain site	•			
For marketing		•		(*)
For telemarketing		•		(*)
For profiling				•

(*) SSN/passport

Who we share the information with

	Yes	Yes, unless you opt out	No, unless you opt in	No
With 3rd parties for marketing purposes			•	(*)
To public forums	(**)			•

(*) SSN/passport

(**) Preferences and your activity on this site

Contact us: call 1 888-888-8888
www.acme.com

Figure C.2: The Privacy Table label for part 1 in the pre-test

The Bell Policy

What information we collect

	Yes	No
Contact information	•	
Cookies	•	
Demographic information		•
Financial information	•	
Health information		•
Preferences	•	
Purchasing information	•	
SSN / passport	•	
Your activity on this site	•	
Your location	•	

How we use the information we collect

	Yes	Yes, unless you opt out	No, unless you opt in	No
To provide service and maintain site	•			
For marketing	•	(*)		(**)
For telemarketing	•	(*)		(***)
For profiling				•

(*) Contact information and cookies
(**) Financial information, purchasing information and SSN / Passport
(***) Contact information, Financial information, purchasing information and SSN / Passport

Who we share the information with

	Yes	Yes, unless you opt out	No, unless you opt in	No
With 3rd parties for marketing purposes			•	(*)
To public forums	(**)			•

(*) SSN/passport

(**) Preferences and your activity on this site

Contact us: call 1 888-888-8888
www.acme.com

The Acme policy

What information we collect

	Yes	No
Contact information	•	
Cookies	•	
Demographic information		•
Financial information		•
Health information		•
Preferences	•	
Purchasing information	•	
SSN / passport	•	
Your activity on this site	•	
Your location		•

How we use the information we collect

	Yes	Yes, unless you opt out	No, unless you opt in	No
To provide service and maintain site	•			
For marketing		•		(*)
For telemarketing		•		(*)
For profiling				•

(*) SSN/passport

Who we share the information with

	Yes	Yes, unless you opt out	No, unless you opt in	No
With 3rd parties for marketing purposes			•	(*)
To public forums	(**)			•

(*) SSN/passport

(**) Preferences and your activity on this site

Contact us: call 1 888-888-8888
www.acme.com

Figure C.3: The Privacy Table labels for part 2 in the pre-test

Figure C.4: The Nutrition Labels for part 2 in the pre-test

APPENDIX D

Laboratory experiment

D.1 Questionnaire

Instructions

Your task in this experiment is to simply answer a series of questions regarding the privacy policies of two fictive companies: “Acme Inc.” and “Bell Group”. The questions are divided into two parts, where part one requires you to look for answers in the Acme policy, and part two requires you to compare the policies of Acme and Bell to find the answer. Both parts also contain a section of statements to which you answer by selecting a degree of agreement/disagreement.
Each question has just one correct answer, and you also have the option to answer “No answer” in case you don’t understand the question.

You will see a clock symbol () followed by a text box (:) before and after some questions in both parts. Please write the exact current time (of when you reach that point) in these boxes. You are allowed to look at and actively use the privacy policies WHILE answering the questions, but please only do so after you have written the current time in the box.

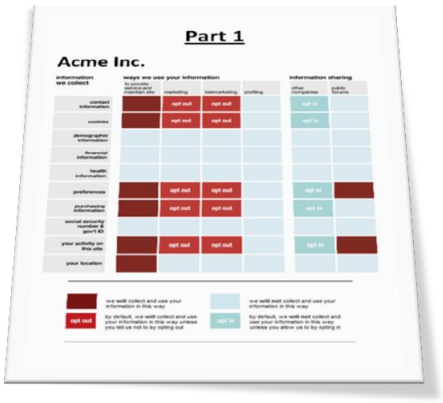
No personal information will be collected during this experiment and all your answers will be treated confidentially and anonymously. Participating is voluntary and you may stop participating at any time. By completing and handing in this survey you accept that your answers will be used for the purpose of answering the research questions in my thesis.

Thank you for your contribution!

Part 1

In this part you will be answering questions regarding the "Acme Inc." privacy policy.

Use the sheet labeled **Part 1** for answering the questions in this part!



Part 1.1

Start time: :

Please answer the following questions regarding the "Acme Inc." policy

	Yes	No	Does not say	I don't know
1. Does the policy allow Acme to collect information about which pages you visited on this web site?	❑	❑	❑	❑
2. Does the policy allow Acme to collect information about your medical conditions, drug prescriptions, or family health history?	❑	❑	❑	❑
3. Does the policy allow Acme to collect information about your current location?	❑	❑	❑	❑
4. Does the policy allow Acme to use information about your gender for marketing purposes?	❑	❑	❑	❑
5. Does the policy allow Acme to collect information regarding your household income?	❑	❑	❑	❑
6. Does the policy allow Acme to use your information to improve their website?	❑	❑	❑	❑

Part 1.2

Please answer the following questions regarding the Acme Inc. policy

	Yes	Yes, unless I tell them not to	Only if I allow them to	No	The policy does not say	I don't know
7. Does the policy allow Acme to share your personal information on public forums?	❑	❑	❑	❑	❑	❑
8. Does the policy allow Acme to share your mobile number with other companies?	❑	❑	❑	❑	❑	❑
9. Does the policy allow Acme to use your buying history to design custom functionality targeted at you?	❑	❑	❑	❑	❑	❑
10. Does the policy allow Acme to share your cookie information with other companies?	❑	❑	❑	❑	❑	❑
11. Will Acme contact you with advertisements?	❑	❑	❑	❑	❑	❑
12. Does Acme give you control regarding their sharing of your personal data?	❑	❑	❑	❑	❑	❑

End time: :

Part 1.3

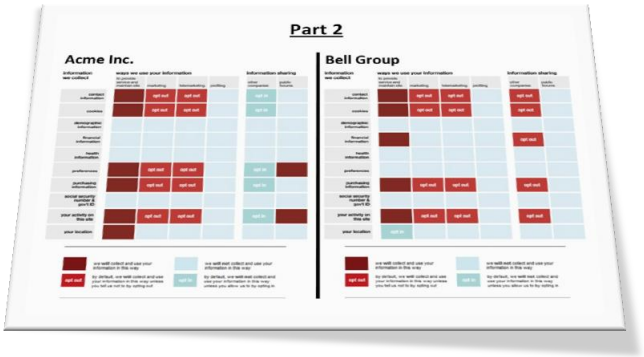
To what extent do you agree or disagree with each of the following statements?

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree
13. I feel that Acme's privacy practices are explained thoroughly in the privacy policy I read	❑	❑	❑	❑	❑
14. I had problems finding the information I was looking for in Acmes policy	❑	❑	❑	❑	❑
15. I feel secure about sharing my personal information with Acme after viewing their privacy practices	❑	❑	❑	❑	❑
16. When I first looked at Acme's policy, it was easy to understand what information they will collect from me	❑	❑	❑	❑	❑
17. If all privacy policies looked just like this I would be more likely to read them	❑	❑	❑	❑	❑
18. The use of symbols (e.g. icons, characters or colors) in Acmes policy made it confusing to understand how the information they collect will be used	❑	❑	❑	❑	❑

Part 2

In this part you will be answering questions which require you to compare two policies against each other. The "Acme Inc." policy is the same as in part 1, and the other policy belongs to "Bell Group".

Use the sheet labeled **Part 2** for answering the questions in this part!



Part 2.1

Start time: :

Please answer the following questions by comparing the two policies

	True	False	Does not say	I don't know
19. Both Acme and Bell lets you control whether non-sensitive information (such as name and address) is shared with other companies	❑	❑	❑	❑
20. Neither Acme nor Bell collects any sensitive information (such as banking or medical records, or passport numbers)	❑	❑	❑	❑
21. By default, Acme can collect information about your use of their website in order to market to you by email, but Bell cannot.	❑	❑	❑	❑
22. By default, Bell can share your history of purchased items with other companies, but Acme cannot	❑	❑	❑	❑
23. By default, both companies collects your default location and can use this information to improve their services	❑	❑	❑	❑

End time: :

Part 2.2

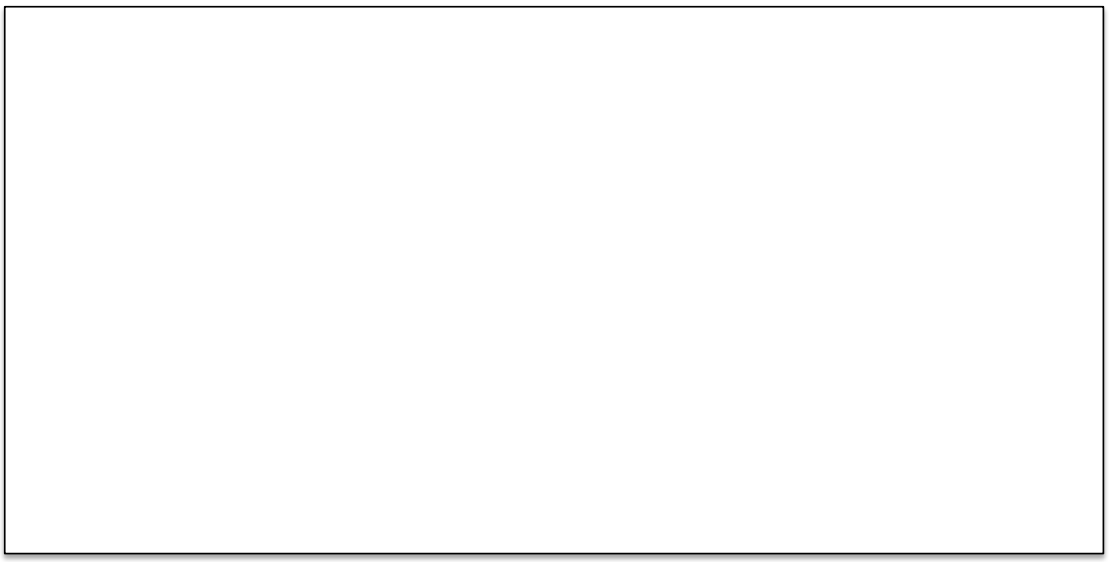
To what extent do you agree or disagree with each of the following statements?

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
24. I found comparing two policies an easy task to do	❑	❑	❑	❑	❑
25. If all policies looked like this I would compare privacy practices across websites more often	❑	❑	❑	❑	❑
26. It was easy to keep control of what the different symbols (e.g. icons, characters or colors) in the policy stood for	❑	❑	❑	❑	❑
27. I feel that these policies would cover most of the privacy concerns I would have if I were to sign up for Acme or Bell in real life	❑	❑	❑	❑	❑
28. I would feel more confident in understanding the privacy practices of Acme or Bell if the policies had more text and less symbols (e.g. icons, characters or colors).	❑	❑	❑	❑	❑

Part 2.3

29. Do you have any comments or suggestions regarding the privacy policy format?

30. Do you have any feedback regarding the format of this survey?



D.2 Policies

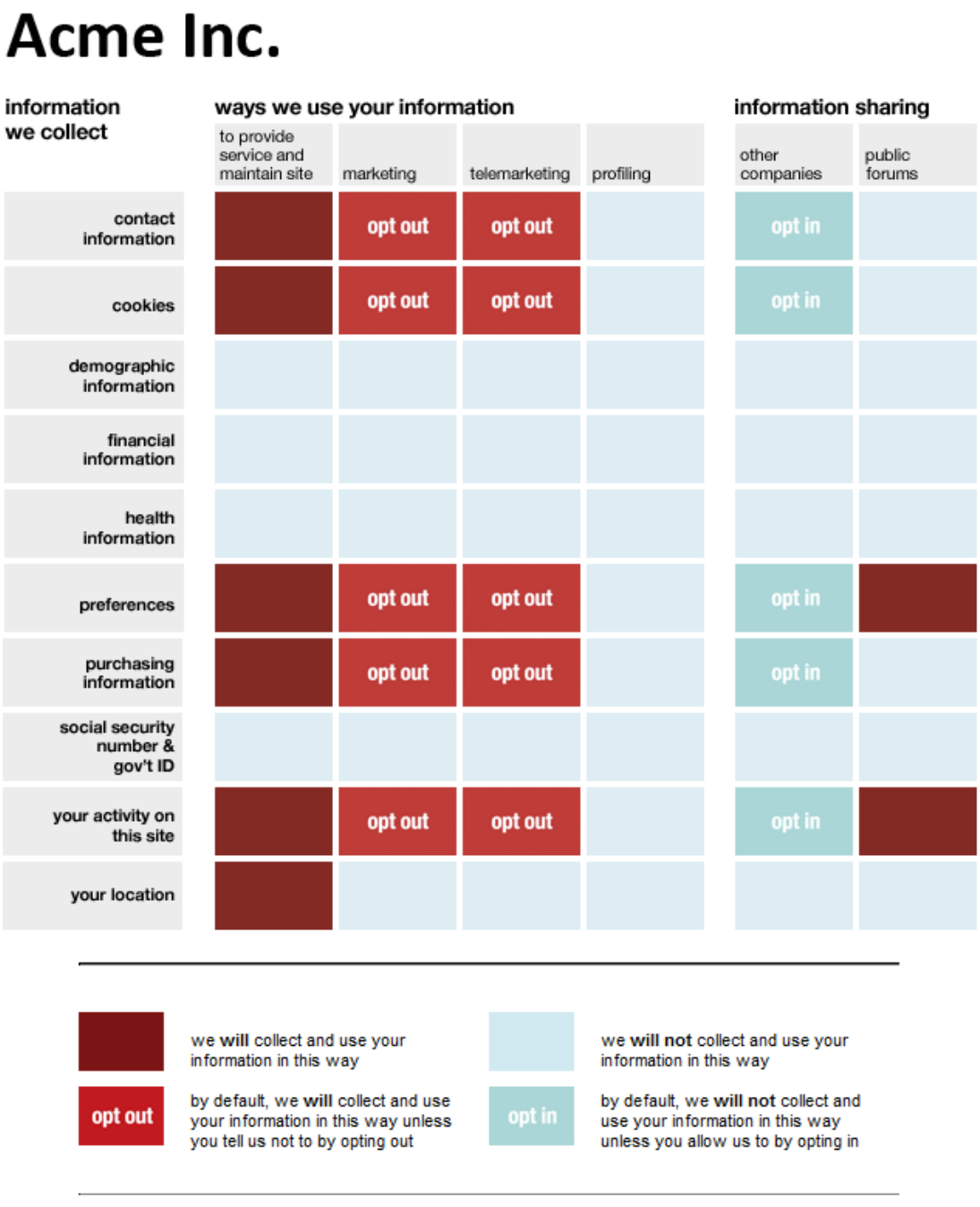
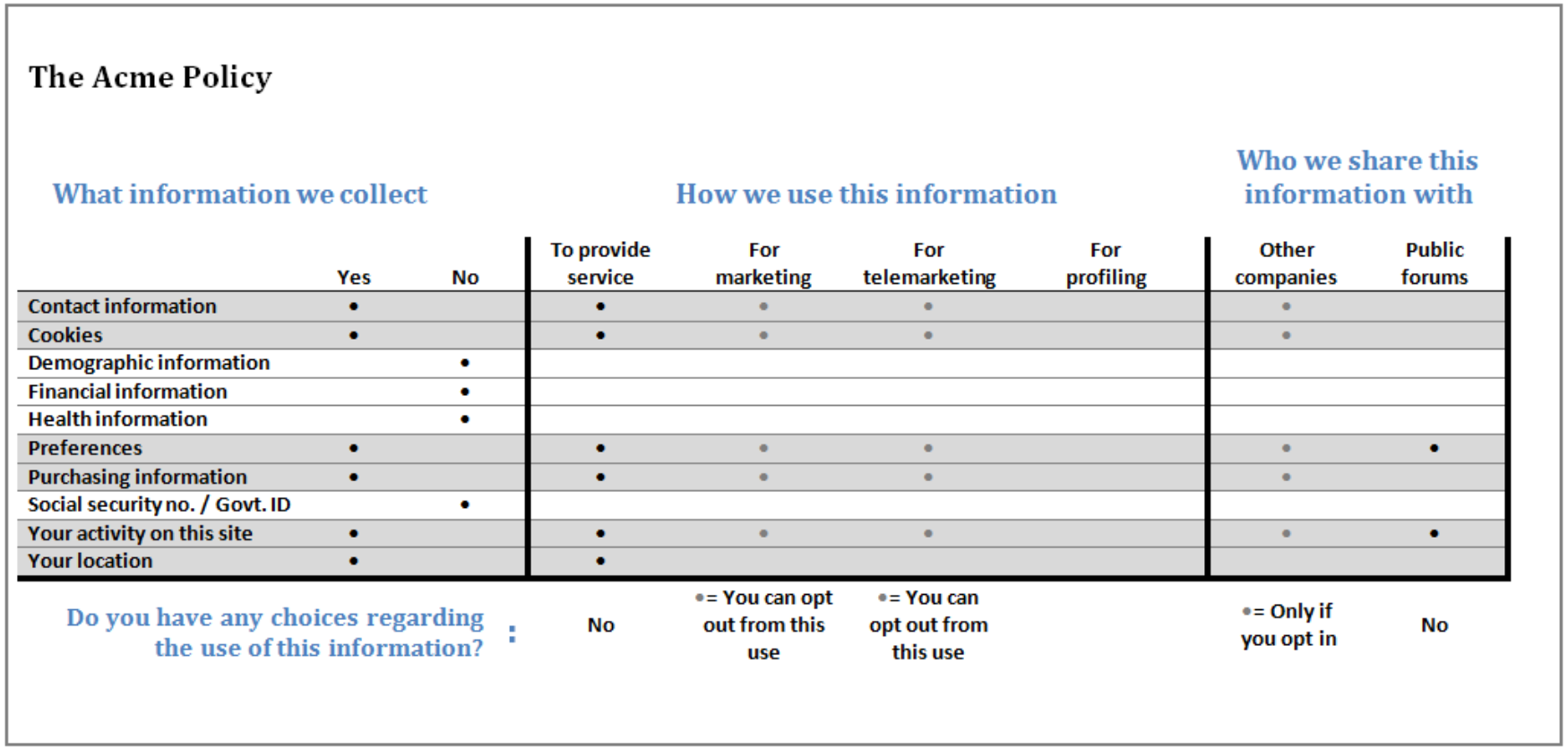


Figure D.1: The Nutrition Label for part 1 in the laboratory experiment



The Acme Policy

What information we collect			How we use this information				Who we share this information with	
	Yes	No	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	•		•	◦	◦		◦	
Cookies	•		•	◦	◦		◦	
Demographic information		•						
Financial information		•						
Health information		•						
Preferences	•		•	◦	◦		◦	•
Purchasing information	•		•	◦	◦		◦	
Social security no. / Govt. ID		•						
Your activity on this site	•		•	◦	◦		◦	•
Your location	•		•					

Do you have any choices regarding the use of this information? :	No	◦= You can opt out from this use	◦= You can opt out from this use		◦= Only if you opt in	No

Figure D.2: The Privacy Table for part 1 in the laboratory experiment

The Acme Inc. Policy

What information we collect			How we use this information				Who we share this information with	
	Yes	No	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	•		•	◦	◦		◦	
Cookies	•		•	◦	◦		◦	
Demographic information		•						
Financial information		•						
Health information		•						
Preferences	•		•	◦	◦		◦	•
Purchasing information	•		•	◦	◦		◦	
Social security no. / Govt. ID		•						
Your activity on this site	•		•	◦	◦		◦	•
Your location	•		•					
Do you have any choices regarding the use of this information?			No	◦= You can opt out from this use	◦= You can opt out from this use		◦= Only if you opt in	No

The Bell Group Policy

What information we collect			How we use this information				Who we share this information with	
	Yes	No	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	•		•	◦	◦		◦	
Cookies	•		•	◦	◦		◦	
Demographic information		•						
Financial information	•		•				◦	
Health information		•						
Preferences		•						
Purchasing information	•		•	◦	◦		◦	
Social security no. / Govt. ID		•						
Your activity on this site	•		•	◦	◦		◦	
Your location	•		◦					
Do you have any choices regarding the use of this information?			◦= Only if you opt in	◦= You can opt out from this use	◦= You can opt out from this use		◦= You can opt out from this use	

Figure D.3: The Privacy Table for part 2 in the laboratory experiment

Figure D.4: The Nutrition Labels for part 2 in the laboratory experiment

D.3 Guest lecture slides

Guest lecture NTNU, 01-04-2011

Privacy Agents for End-Users

Karin Bernsmed
PhD – Research scientist

SINTEF Information and Communication Technology
Department of Software Engineering, Safety and Security
Trondheim, Norway

SINTEF SINTEF ICT 1

Agenda

PART I (08:15-09:00)
- Privacy and the Internet
- Personal data management
- Privacy agents

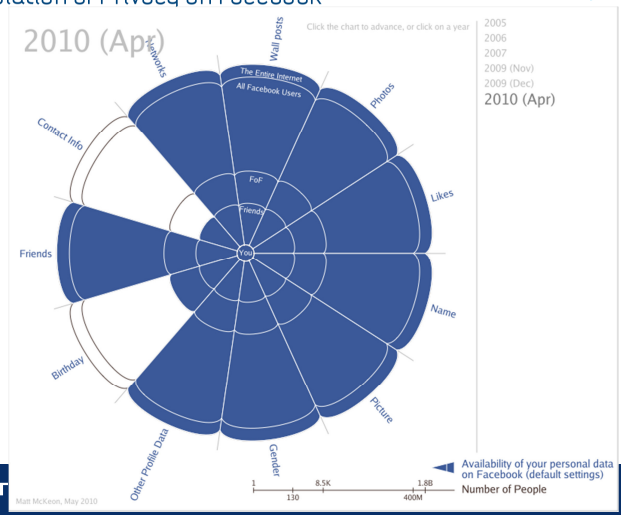
PART II (09:15-10:00)
- Practical experiment

Privacy and the Internet

- Users "pay" for "free" services with their personal information
- "The privacy paradox": Users often claim to be concerned about privacy – a statement not reflected by their actions.
- Context aware applications are here to stay!

The Evolution of Privacy on Facebook

Source: http://mattmckeon.com/facebook-privacy/



Privacy Enhancing Technologies (PETs)

PETs for the Internet may be divided into 5 categories:
- Identity management:
  The user receive personalized services without being identified (using pseudonyms)
- Anonymous communication:
  Preventing traffic analysis using proxies (MIX networks, DC networks, "Crowds")
- Anonymous access
  The users can access data and applications anonymously
- Privacy –preserving authorization
  Users are authorized without being authenticated (e-cash)
- Personal data management
  Personal data minmization & self-determination (privacy agents)

Personal Data

- Any information that can be used to uniquely identify, contact, or locate a single person
  - Full name
  - Telephone number
  - Street address
  - Email address
  - Date of birth
  - Credit card number
  - IP address
  - Digital identity
  - Face, fingerprint or handwriting
  - .....
- The term "Personally Identifiable Information" (PII) is often used in the U.S.
- The collection of personal data has increased enormously lately!

Sharing Personal Data Online

Privacy policies are used to inform users how their personal data will be colleted, processed and shared with 3rd parties

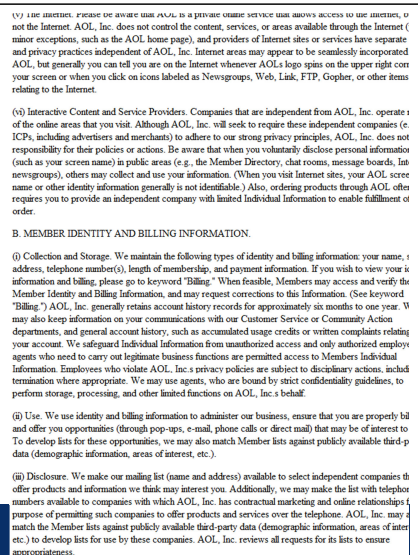
- What data do they collect?
- What do they use it for?
- Who do they share it with?

Most users do not read (or understand) privacy policies!

Why Don't People Read Privacy Policies?

Problem with policies
- They are too long
- They are difficult to find
- The legal jargong is confusing
- Lack of standardized format

Users' attitude
- The users don't have time
- The users don't care
- Users believe they don't have choices



(v) The Internet. Please be aware that AOL is a private online service that allows access to the Internet, b not the Internet. AOL, Inc. does not control the content, services, or areas available through the Internet (minor exceptions, such as the AOL home page), and providers of Internet sites or services have separate and privacy practices independent of AOL, Inc. Internet areas may appear to be seamlessly incorporated AOL, but generally you can tell you are on the Internet whenever AOLs logo spins on the upper right corn your screen or when you click on icons labeled as Newsgroups, Web, Link, FTP, Gopher, or other items relating to the Internet.

(vi) Interactive Content and Service Providers. Companies that are independent from AOL, Inc. operate s of the online areas that you visit. Although AOL, Inc. will seek to require these independent companies (e. ICPs, including advertisers and merchants) to adhere to our strong privacy principles, AOL, Inc. does not responsibility for their policies or actions. Be aware that when you voluntarily disclose personal informatio (such as your screen name) in public areas (e.g., the Member Directory, chat rooms, message boards, Int newsgroups), others may collect and use your information. (When you visit Internet sites, your AOL scree name or other identity information generally is not identifiable.) Also, ordering products through AOL offer requires you to provide an independent company with limited Individual Information to enable fulfillment o order.

B. MEMBER IDENTITY AND BILLING INFORMATION.

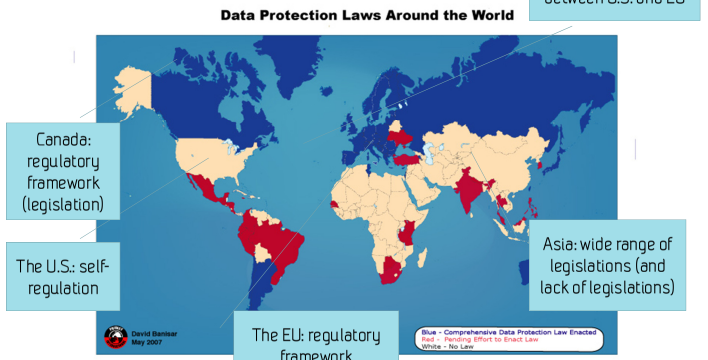
(i) Collection and Storage. We maintain the following types of identity and billing information: your name, s address, telephone number(s), length of membership, and payment information. If you wish to view your ic information and billing, please go to keyword "Billing." When feasible, Members may access and verify the Member Identity and Billing Information, and may request corrections to this Information. (See keyword "Billing.") AOL, Inc. generally retains account history records for approximately six months to one year. W may also keep information on your communications with our Customer Service or Community Action departments, and general account history, such as accumulated usage credits or written complaints relating your account. We safeguard Individual Information from unauthorized access and only authorized employe agents who need to carry out legitimate business functions are permitted access to Members Individual Information. Employees who violate AOL, Inc.s privacy policies are subject to disciplinary actions, includi termination where appropriate. We may use agents, who are bound by strict confidentiality guidelines, to perform storage, processing, and other limited functions on AOL, Inc.s behalf.

(ii) Use. We use identity and billing information to administer our business, ensure that you are properly bil and offer you opportunities (through pop-ups, e-mail, phone calls or direct mail) that may be of interest to To develop lists for these opportunities, we may also match Member lists against publicly available third-p data (demographic information, areas of interest, etc.).

(iii) Disclosure. We make our mailing list (name and address) available to select independent companies th offer products and information we think may interest you. Additionally, we may make the list with telephor numbers available to companies with which AOL, Inc. has contractual marketing and online relationships f purpose of permitting such companies to offer products and services over the telephone. AOL, Inc. may s match the Member lists against publicly available third-party data (demographic information, areas of inter etc.) to develop lists for use by these companies. AOL, Inc. reviews all requests for its lists to ensure appropriateness.

Privacy Rules Around the World

Safe Harbour: bridges the differences between U.S. and EU



Canada: regulatory framework (legislation)

The U.S.: self-regulation

Asia: wide range of legislations (and lack of legislations)

The EU: regulatory framework (legislation)

http://www.privacyinternational.org

The EU Data Protection Directive (Directive 95/46/EC)

- Based on seven principles:
  - Notice—data subjects should be given notice when their data is being collected;
  - Purpose—data should only be used for the purpose stated and not for any other purposes;
  - Consent—data should not be disclosed without the data subject's consent;
  - Security—collected data should be kept secure from any potential abuses;
  - Disclosure—data subjects should be informed as to who is collecting their data;
  - Access—data subjects should be allowed to access their data and make corrections to any inaccurate data; and
  - Accountability—data subjects should have a method available to them to hold data collectors accountable for following the above principles.
- The E-Privacy directive adds requirements on data retention, spam and cookies.
- New data protection rules will be presented summer 2011
  - The right to be forgotten
  - Transparency
  - Privacy by default
  - Protection regardless of data location

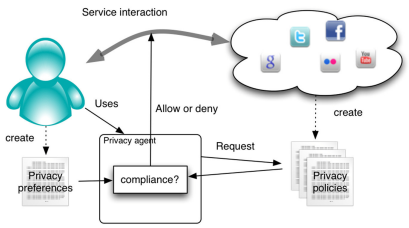
The U.S. FTC Privacy Model

- Notice/awareness:
  Consumers should be informed of information practices before personal data is collected from them
- Choice/consent:
  Giving cosumers options as to how personal data collected from them may be used
- Access/participation:
  The right to view and contest the accuracy of collected data
- Integrity/security:
  Data must be accurate and protected
- Enforcement/redress:
  Mechanisms to ensure that privacy protection principles are applied

The FTC privacy model says nothing about limits (or usage) on the amount of data collected

The Basic Privacy Agent Model

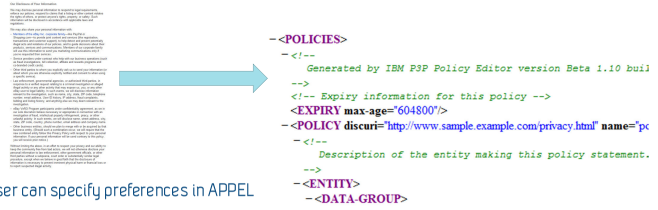
- A privacy agent helps the user with his privacy management
- The agent can display and/or match privacy policies with user preferences
- The agent relies on machine-redable policies



- The Platform for Privacy Preferences (P3P) Project was an early approach

The Platform for Privacy Preferences (P3P)

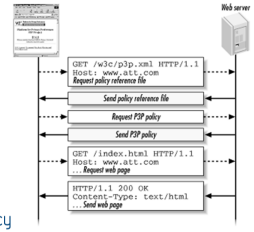
- P3P is a standard for communicating privacy policies for web sites to their clients
- P3P was developed by the World Wide Web Consortium (W3C). It includes:
  - a standard *vocabulary* for describing a web site's data practices
  - a set of *base data elements* that web sites can refer to in their P3P privacy policies
  - a protocol for requesting and transmitting web site privacy policies
- P3P policies are encoded using XML
- P3P "compact policies" covers the use of cookies



- User can specify preferences in APPEL

P3P: How It Works

- Service providers use their textual privacy policy to create the P3P policy file(s)
- A policy reference file lists the P3P policies for the site and the parts to which they apply
- The policy references file is put in a well-known location (*/w3c/p3p.xml*)
- Privacy agents can then fetch and process P3P policy files



P3P Vocabulary: Assertions

General assertions:

‹POLICY›	location of textual policy and opt-out mechanisms
‹TEST›	indication that policy is for testing purposes (optional)
‹ENTITY›	web-site contact information
‹ACCESS›	access information
‹DISPUTES›	information about disputes resolution (optional)

Data specific assertions:

‹CONSEQUENCE›	consequence of providing data (optional)
‹NON-IDENTIFIABLE›	no identifiable data will be collected (optional)
‹PURPOSE›	how data will be used
‹RECIPIENT›	with whom data may be shared
‹RETENTION›	data-retention policy
‹DATA›	what kind of data is collected

An Example P3P Policy

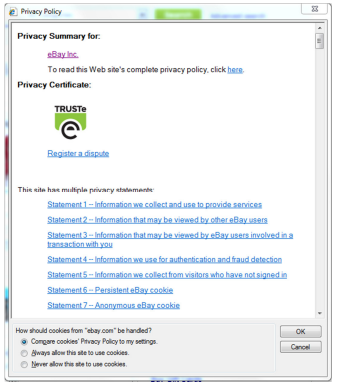
This is the web site for the book Web Privacy with P3P by Lorrie Faith Cranor. We do not currently collect any information from visitors to this site except the information contained in standard web server logs (your IP address, referrer, information about your web browser, information about your HTTP requests, etc.). The information in these logs will be used only by us and the server administrators for website and system administration, and for improving this site. It will not be disclosed unless required by law. We may retain these log files indefinitely. Please direct questions about this privacy policy to privacy@p3pbook.com.

```
<POLICIES>
<POLICY discuri="http://p3pbook.com/privacy.html"
name="policy">
<ENTITY>
<DATA-GROUP>
<DATA ref="#business.contact-
info.online.email">privacy@p3pbook.com </DATA>
<DATA ref="#business.contact-
info.online.uri">http://p3pbook.com/
</DATA>
<DATA ref="#business.name">Web Privacy With
P3P</DATA>
</DATA-GROUP>
</ENTITY>
<ACCESS>
<nonident/>
</ACCESS>
<STATEMENT>
<CONSEQUENCE>
Our Web server collects access logs containing this
information.
</CONSEQUENCE>
<PURPOSE>
<admin/>
<current/>
<develop/>
</PURPOSE>
<RECIPIENT>
<ours/>
</RECIPIENT>
<RETENTION>
<indefinitely/>
</RETENTION>
<DATA-GROUP>
<DATA ref="#dynamic.clickstream"/>
<DATA ref="#dynamic.http"/>
</DATA-GROUP>
</STATEMENT>
</POLICY>
</POLICIES>
```

Policy obtained from http://p3pbook.com/privacy.html

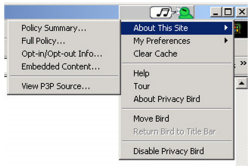
Browser Support

- Internet Explorer (IE) 6+ can display and match P3P policies
  - limited to P3P compact policies (cookies)
- P3P user agents were also implemented in early Mozilla Firefox and Netscape web browsers
- User agents can also be implemented as add-ons



"The AT&T Privacy Bird"

- One of the first P3P user agents
- Implemented as a browser helper object for IE 5.01, 5.5 and 6.0
- Graphical support for preferences specification
- A bird icon in browser window changes shape and color
- Policy summary explains reason for mismatch



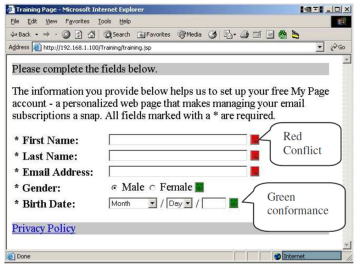
http://www.privacybird.org/

"Integrated Privacy View"

- Designed to match privacy preferences and policies on a fine-grained level
- Based on an "policy anchors" (an extension of P3P) and "p3pdataelement" (a new attribute for HTTP elements)

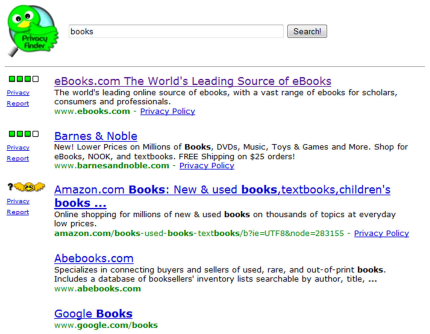
```
<input name="registration_email"
size="40"
value="levysn@us.ibm.com"
p3pdataelement="#user.business-info.online.email" />
```

- Policy statements can be linked to input fields
- An HTTP proxy fetches and modifies pages to indicate conformance



"Privacy Finder"

- Privacy-enhanched online search engine
- Orders search results based on P3P policies
- Privacy meter indicates matching results
- Graphical privacy policy report

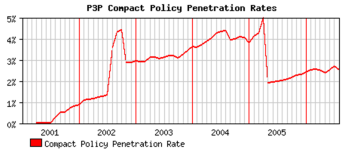


http://www.privacyfinder.org/

The Rise and Fall of P3P

- 1997: P3P kickoff meeting at W3C
- 1998: First draft of P3P 1.0 specification
- 1999: W3C removes data transfer from P3P
- 2000: P3P workhop with user agent and policy editor demonstrations.
- 2001: Microsoft announces P3P support in Internet Explorer 6.0
- 2002: W3C issues P3P 1.0 as a recommendation
- .....
- 2006: W3C work on P3P suspended

Most of the major service providers do not use P3P anymore



Why Did P3P Fail?

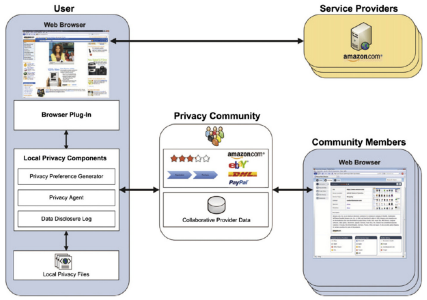
Technical obstacles

- P3P addresses only parts of privacy concepts ("notice" and "choice" but not "enforcement")
- It is difficult to manage complex privacy policies and preferences
- The P3P vocabulary is cumbersome to use
- The limited availability of user agents

Polictical critiques

- P3P has been heavility criticized by parts of the privacy community (the self-regulation vs legislation debate)

Other Privacy Agent Efforts (Research)



Privacy Beaver Icon	Colour	Meaning
	Grey	The PIPWatch system has been turned off.
	Grey	The privacy officer not been identified or contacted
	Grey	The privacy officer has not responded to the questionnaire
1/10	Red	The current website's privacy practices DOES NOT match the user's preferences
5/10	Yellow	The current website's privacy practices PARTIALLY matches the user's preferences
10/10	Green	The current website's privacy practices matches the user's preferences

What is Missing in the Privacy Agent Approach?

- Service provider support
  - They need a reason to create machine-readable policies
- Web browser support
  - Privacy agents must be a core browser functionality
- Privacy policy enforcement mechanisms
  - The user should know that the provider adheres to the policy

Our Research

SINTEF ICT does research on privacy, with focus on

- Privacy agents design and implementation
- Machine-readable policy languages
- Privacy preferences specification

We supervise 5th year project and master students on these and related topics.

IngerAnne.Tondel@sintef.no
Karin.Bernsmed@sintef.no

Next up: practical experiment on privacy policy presentation

APPENDIX E

Internet experiment

E.1 Questionnaire

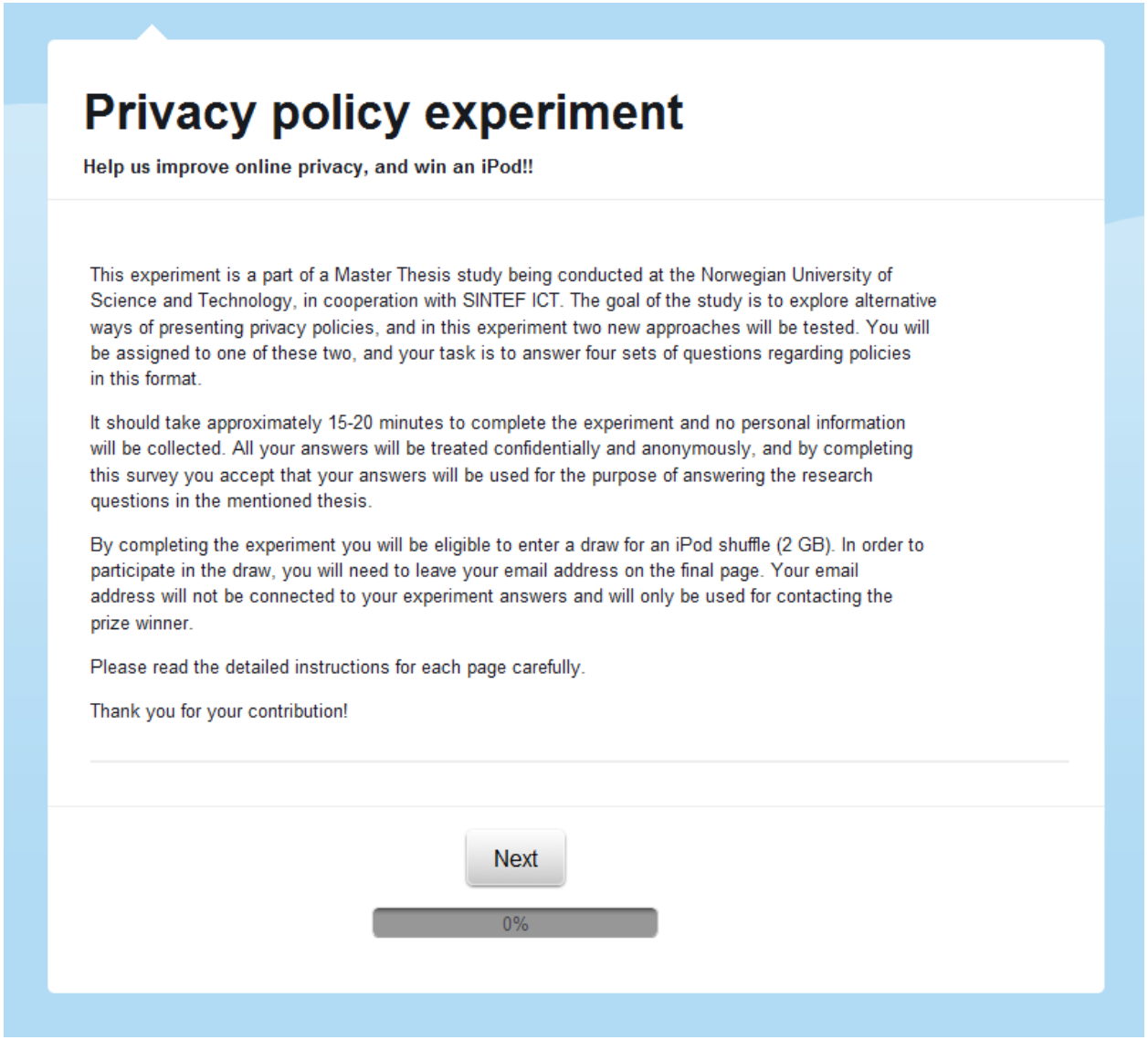


Figure E.1: Introduction for the Internet experiment

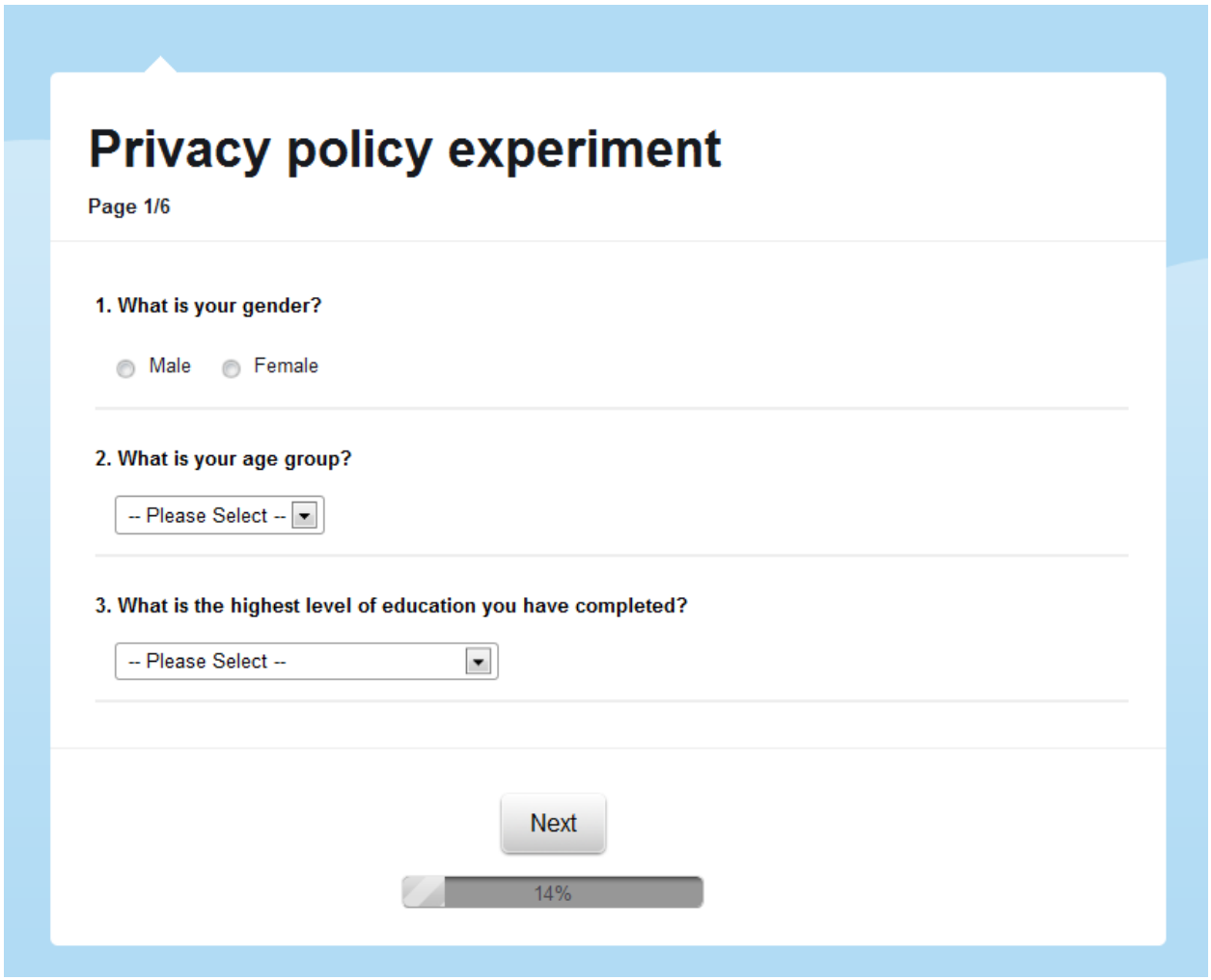
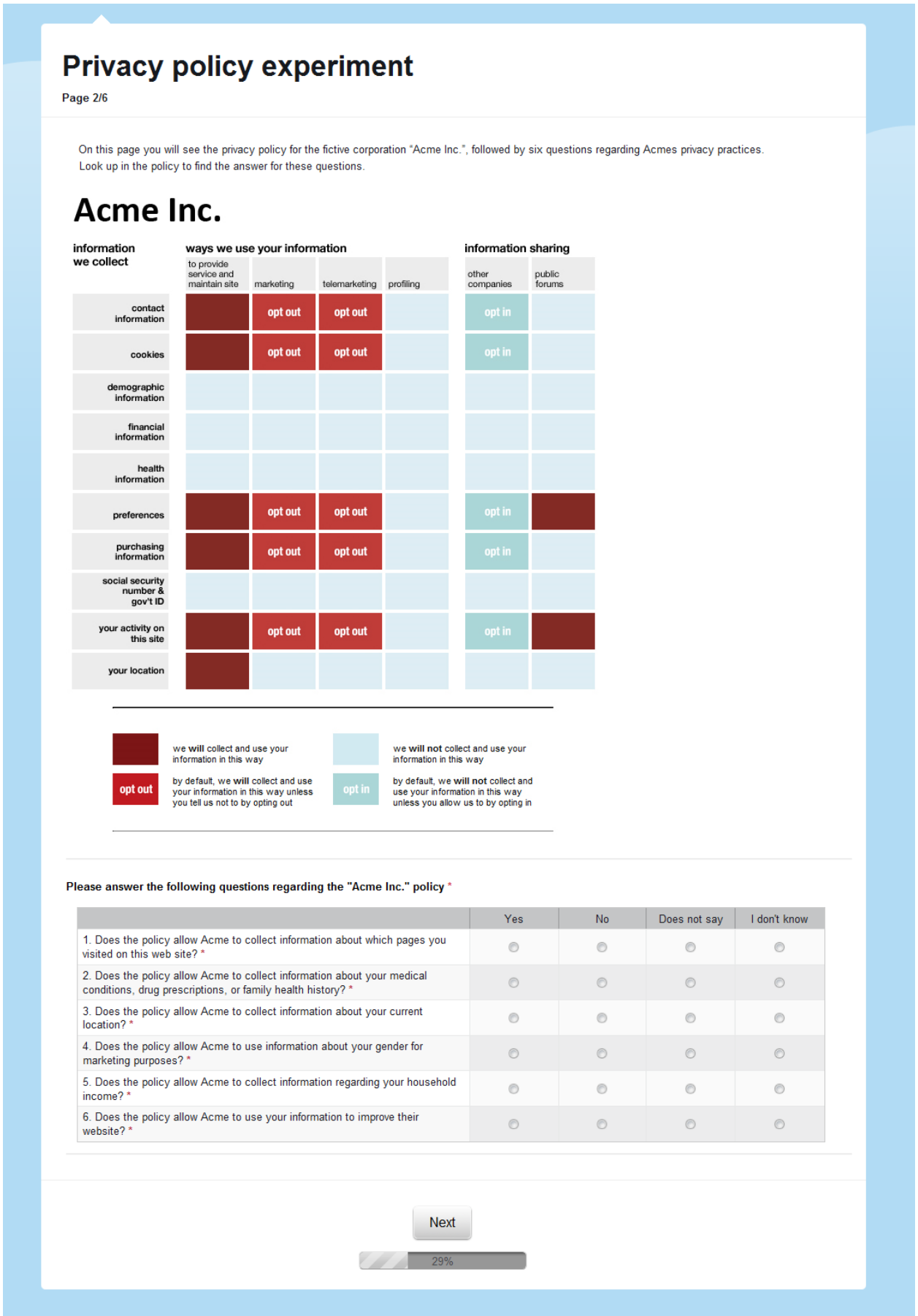


Figure E.2: Demographic questions in the Internet experiment



E–3

Figure E.3: The nutrition label version of part 1.1 in the Internet experiment. The Privacy Table version was identical

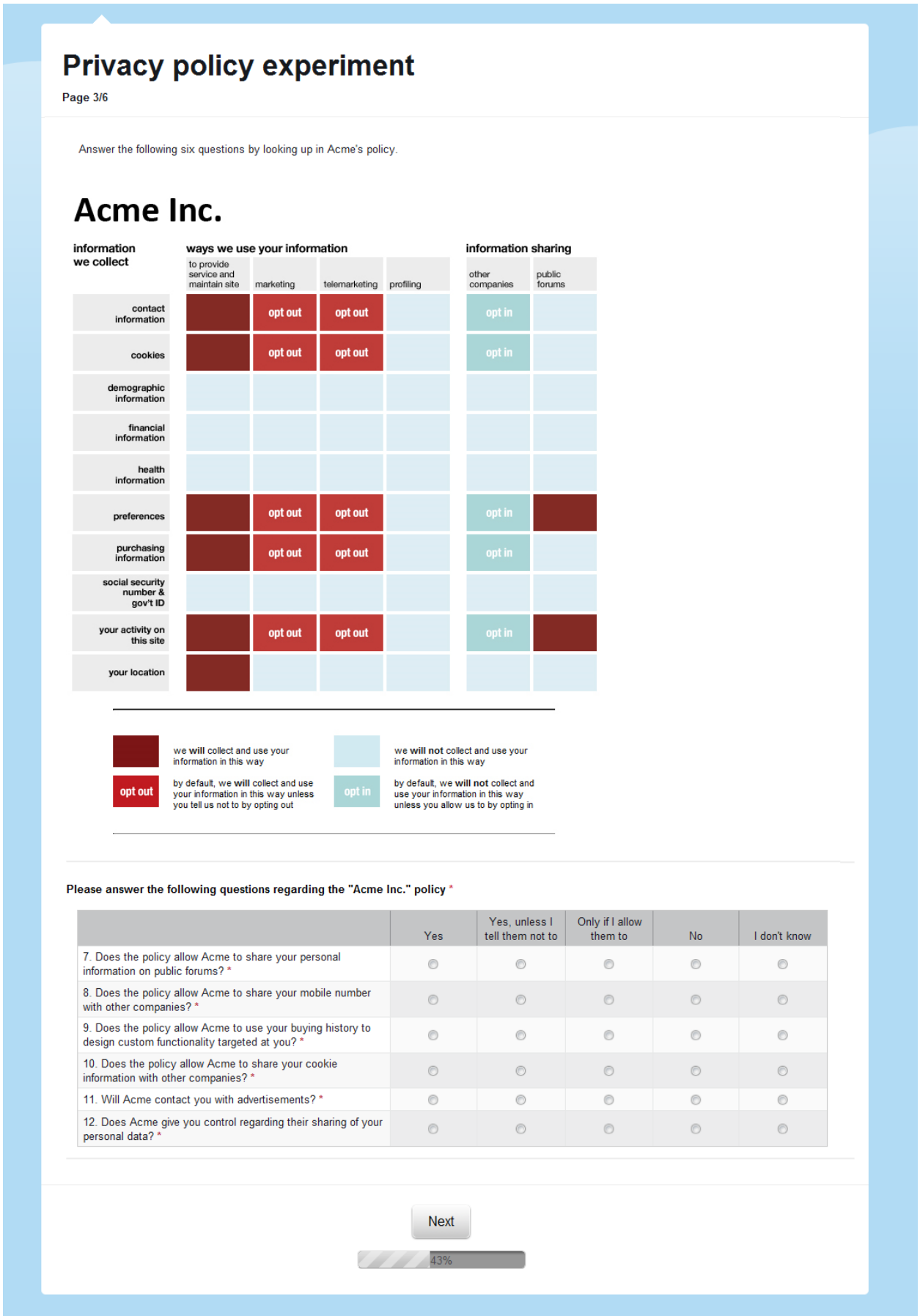


Figure E.4: The nutrition label version of part 1.2 in the Internet experiment. The Privacy Table version was identical

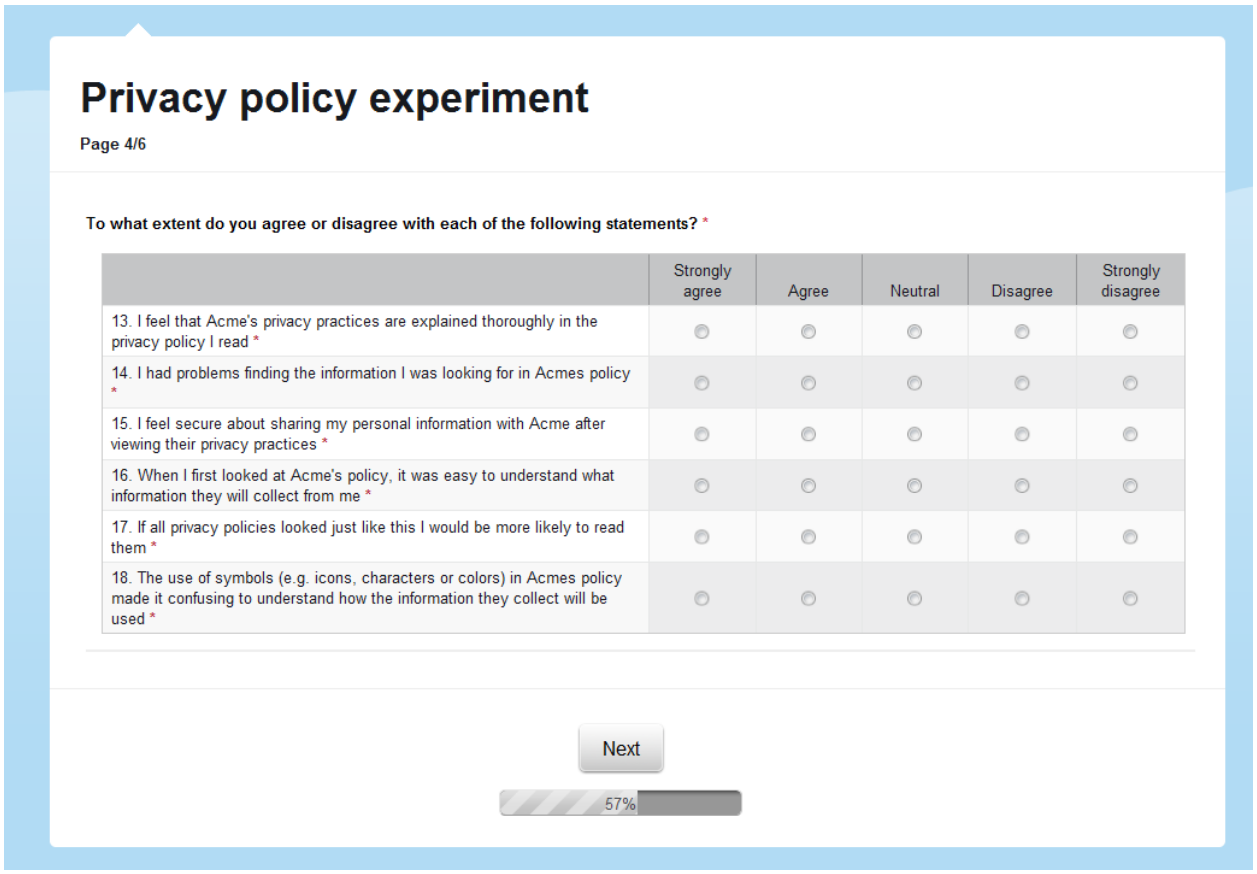


Figure E.5: The single policy likeability questions (part 1.3) in the Internet experiment.

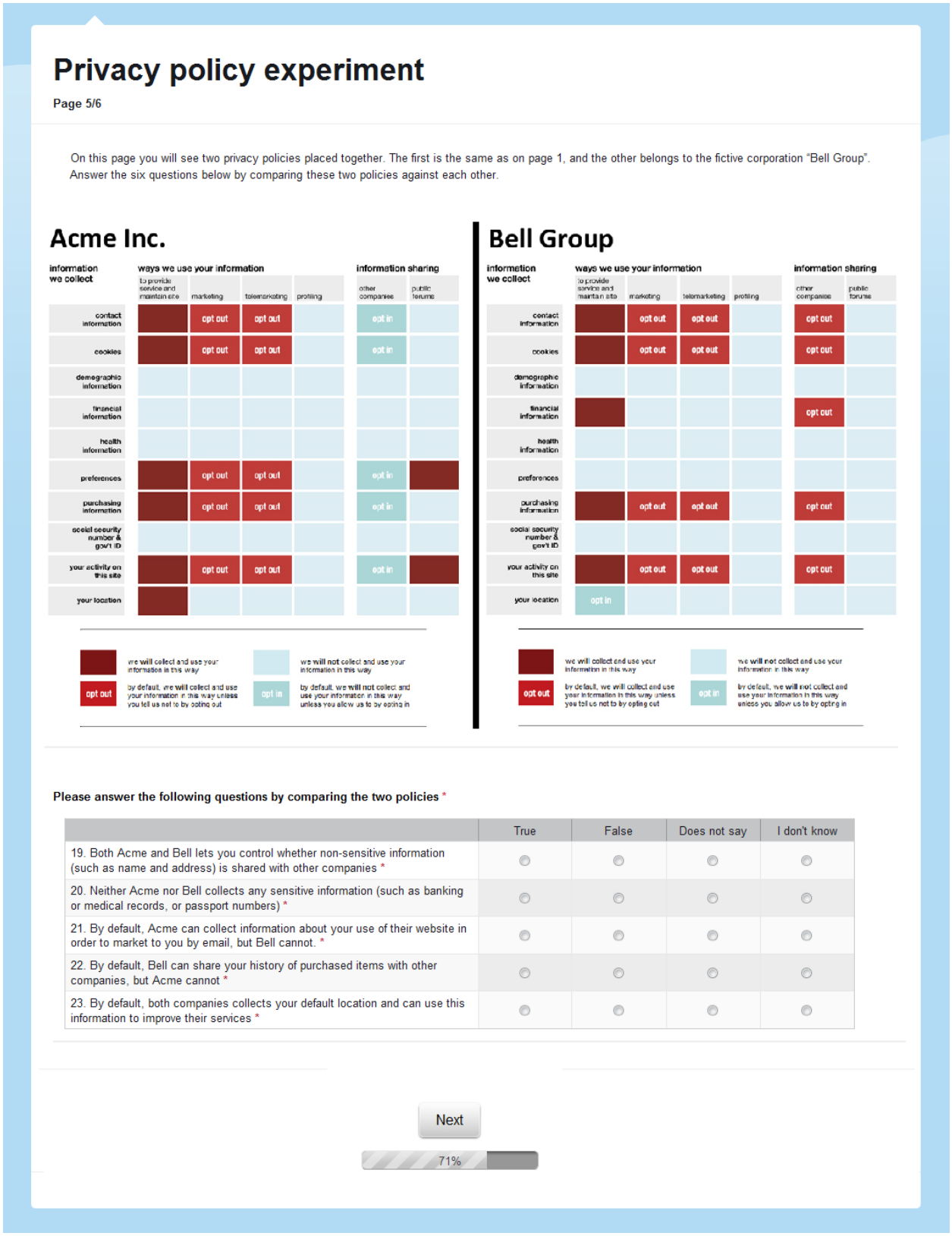


Figure E.6: The nutrition label version of part 2.1 in the Internet experiment. The Privacy Table version was identical. The two policies has been descaled in this screenshot to fit within a single page.

Privacy policy experiment

Page 6/6

To what extent do you agree or disagree with each of the following statements? *

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
24. I found comparing two policies an easy task to do *					
25. If all policies looked like this I would compare privacy practices across websites more often *					
26. It was easy to keep control of what the different symbols (e.g. icons, characters or colors) in the policy stood for *					
27. I feel that these policies would cover most of the privacy concerns I would have if I were to sign up for Acme or Bell in real life *					
28. I would feel more confident in understanding the privacy practices of Acme or Bell if the policies had more text and less symbols (e.g. icons, characters or colors) *					

29. Do you have any comments or suggestions regarding the privacy policy format?

30. Do you have any feedback regarding the format of this survey?

Win an iPod Shuffle!

To enter the prize drawing, please enter your email address.

Note: This email-address will not be connected to your answers in this experiment, and will only be used for the purpose of contacting the prize winner. We will not spam or share your information. We will use the randomresult.com service to ensure a fair and random drawing process.

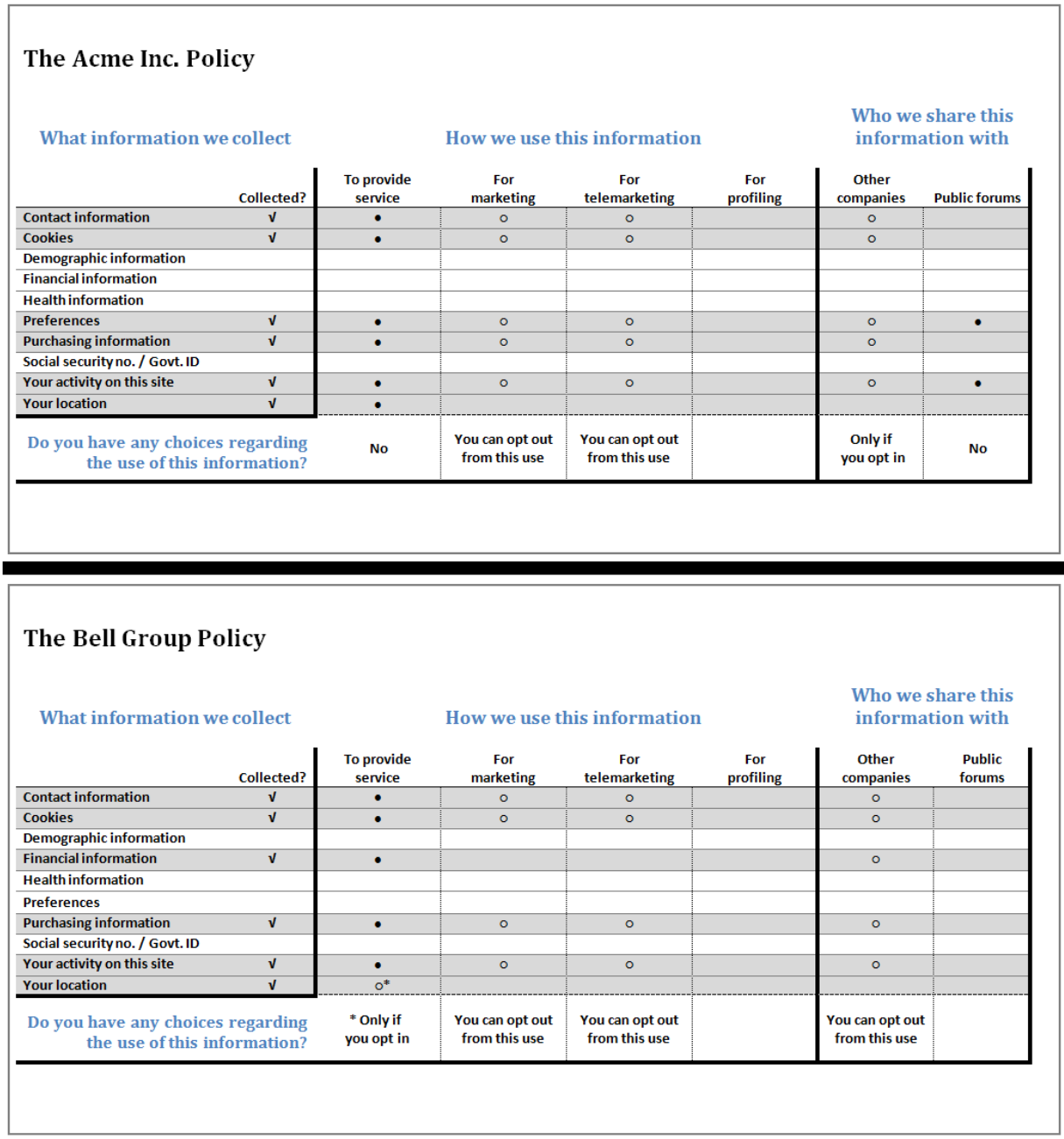
Submit

86%

Figure E.7: The policy comparison likeability questions (part 1.3) and the open end questions in the Internet experiment.

E.2 Policies

Se appendix D.2 for the nutrition label policies.



The Acme Inc. Policy

What information we collect		How we use this information				Who we share this information with	
	Collected?	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	√	•	○	○		○	
Cookies	√	•	○	○		○	
Demographic information							
Financial information							
Health information							
Preferences	√	•	○	○		○	•
Purchasing information	√	•	○	○		○	
Social security no. / Govt. ID							
Your activity on this site	√	•	○	○		○	•
Your location	√	•					
Do you have any choices regarding the use of this information?		No	You can opt out from this use	You can opt out from this use		Only if you opt in	No

The Bell Group Policy

What information we collect		How we use this information				Who we share this information with	
	Collected?	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	√	•	○	○		○	
Cookies	√	•	○	○		○	
Demographic information							
Financial information	√	•				○	
Health information							
Preferences							
Purchasing information	√	•	○	○		○	
Social security no. / Govt. ID							
Your activity on this site	√	•	○	○		○	
Your location	√	○*					
Do you have any choices regarding the use of this information?		* Only if you opt in	You can opt out from this use	You can opt out from this use		You can opt out from this use	

Figure E.8: The Privacy Table labels for part 2 in the Internet experiment

The Acme Inc. Policy

What information we collect		How we use this information				Who we share this information with	
	Collected?	To provide service	For marketing	For telemarketing	For profiling	Other companies	Public forums
Contact information	√	•	o	o		o	
Cookies	√	•	o	o		o	
Demographic information							
Financial information							
Health information							
Preferences	√	•	o	o		o	•
Purchasing information	√	•	o	o		o	
Social security no. / Govt. ID							
Your activity on this site	√	•	o	o		o	•
Your location	√	•					
Do you have any choices regarding the use of this information?		No	You can opt out from this use	You can opt out from this use		Only if you opt in	No

Figure E.9: The Privacy Table for part 1 in the Internet experiment

APPENDIX F

Results

F.1 Histogram and normality plots

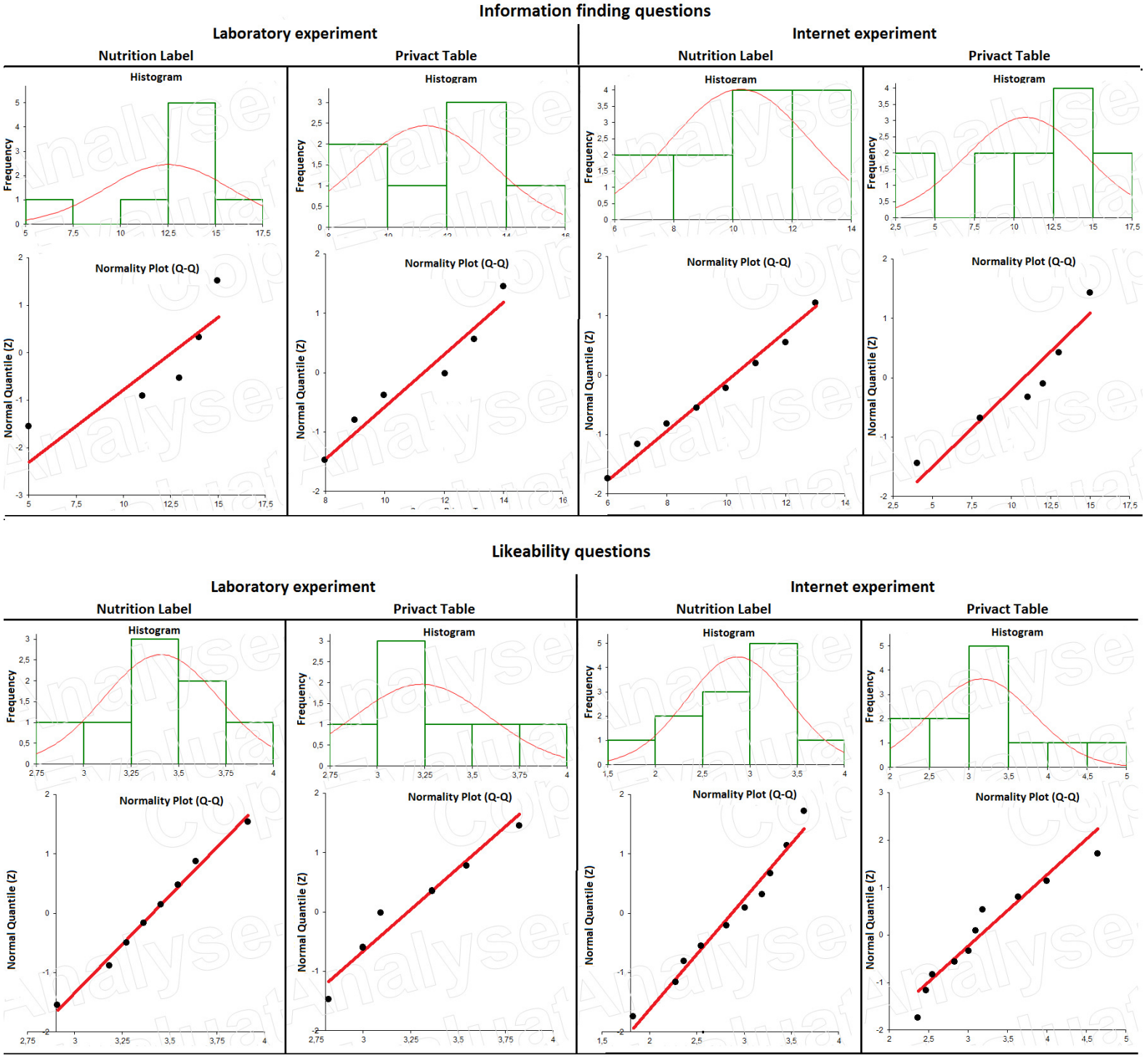


Figure F.1: Histograms and Normality plots for the normality test on the sample datas. Charts created with "Analyse-it".

F.2 Information finding questions

- **Table F.1** - Descriptive statistics for the information finding questions in Part 1.1, 1.2, 2.1, and overall statistics for all parts combined for the labroatory and Internet experiment.
- **Table F.2** - Two-sample t-tests performed on the average participant score in the information findings sections.

Laboratory experiment **NL: Nutrition Label** **PT: Privacy Table**

Part:	Part 1.1		Q1,2,3,5	Q1-6	Part 1.2		Part 2.1		Total	
Group:	*NL*	*PT*	*NL+PT*	*NL+PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*
Mean	5,25	4,71	3,87	5,00	3,50	2,57	3,75	4,00	12,50	11,29
Standard Error	0,250	0,286	0,091	0,195	0,423	0,429	0,648	0,378	1,150	0,865
Median	5	5	4	5	4	3	4,5	4	14	12
Mode	5	4	4	5	4	3	5	3	14	13
Standard Deviation	0,707	0,756	0,352	0,756	1,195	1,134	1,832	1,000	3,251	2,289
Sample Variance	0,500	0,571	0,124	0,571	1,429	1,286	3,357	1,000	10,571	5,238
Kurtosis	-0,229	-0,350	4,349	-1,077	2,576	-0,743	1,699	-2,600	5,005	-1,686
Skewness	-0,404	0,595	-2,405	0,000	-1,339	-0,725	-1,556	0,000	-2,194	-0,372
Range	2	2	1	2	4	3	5	2	10	6
Minimum	4	4	3	4	1	1	0	3	5	8
maximim	6	6	4	6	5	4	5	5	15	14
Sum	42	33	58	75	28	18	30	28	100	79
Count	8	7	15	15	8	7	8	7	8	7
Confidence Level (95,0%)	0,591	0,699	0,195	0,419	0,999	1,049	1,532	0,925	2,718	2,117

Internet experiment

Part:	Part 1.1		Q1,2,3,5	Q1-6	Part 1.2		Part 2.1		Total	
Group:	*NL*	*PT*	*NL+PT*	*NL+PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*
Mean	4,67	4,33	3,29	4,50	2,67	2,75	2,92	3,67	10,25	10,75
Standard Error	0,256	0,396	0,175	0,233	0,188	0,446	0,499	0,432	0,687	1,115
Median	5	4,5	4	5	3	3	3	4	10,5	12,5
Mode	5	6	4	5	3	4	5	4	13	13
Standard Deviation	0,888	1,371	0,859	1,142	0,651	1,545	1,730	1,497	2,379	3,864
Sample Variance	0,788	1,879	0,737	1,304	0,424	2,386	2,992	2,242	5,659	14,932
Kurtosis	-0,254	-1,226	-1,357	-0,597	-0,337	-0,271	-1,217	-0,224	-0,847	-0,483
Skewness	-0,139	-0,217	-0,632	-0,382	0,439	-1,099	-0,227	-1,068	-0,453	-0,861
Range	3	4	2	4	2	4	5	4	7	11
Minimum	3	2	2	2	2	0	0	1	6	4
maximim	6	6	4	6	4	4	5	5	13	15
Sum	56	52	79	108	32	33	35	44	123	129
Count	12	12	24	24	12	12	12	12	12	12
Confidence Level (95,0%)	0,564	0,871	0,363	0,482	0,414	0,982	1,099	0,951	1,511	2,455

Table F.1: Descriptive statistics for the information finding questions in Part 1.1, 1.2, 2.1, and overall statistics for all parts combined (grey column) for the labroatory experiment (top) and Internet experiment (bottom). The orange shaded columns presents aggregated data for both groups (Nutrition label and Privacy table) within the experiment for Part 1 and Question 1,2,3,5.

	Part 1.1							
	Laboratory		Internet		Question 1,2,3,5		Question 1-6	
	NL	*PT*	*NL*	*PT*	*Laboratory*	*Internet*	*Laboratory*	*Internet*
Mean	5,25	4,71	4,67	4,33	3,87	3,29	5,00	4,50
Variance	0,500	0,571	0,788	1,879	0,124	0,737	0,571	1,304
Observations	8	7	12	12	15	24	15	24
P (F-test - equality of variance)	0,427		0,083		0,001		0,056	
Pooled Variance	0,533		1,333		x		1,027	
Hypothesized Mean Diff.	0		0		0		0	
df	13		22		33		37	
t Stat	1,418		0,707		2,913		1,499	
P(T<=t) one-tail	0,090		0,243		0,003		0,071	
T Critical one-tail	1,771		1,717		1,692		1,687	
P(T<=t) two-tail	0,180		0,487		0,006		0,142	
T Critical Two-tail	2,160		2,074		2,035		2,026	

	Part 1.2				Part 2.1				Part 1.1, 1.2 and 2.1			
	Laboratory		Internet		Laboratory		Internet		Laboratory		Internet	
	NL	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*
Mean	3,50	2,57	2,67	2,75	3,75	4,00	2,92	3,67	12,50	11,29	10,25	10,75
Variance	1,429	1,286	0,424	2,386	3,357	1,000	2,992	2,242	10,571	5,238	5,659	14,932
Observations	8	7	12	12	8	7	12	12	8	7	12	12
P (F-test - equality of variance)	0,457		0,004		0,081		0,320		0,205		0,061	
Pooled Variance	1,363		x		2,269		2,617		8,110		10,295	
Hypothesized Mean Diff.	0		0		0		0		0		0	
df	13		15		13		22		13		22	
t Stat	1,537		-0,172		-0,321		-1,136		0,824		-0,382	
P(T<=t) one-tail	0,074		0,433		0,377		0,134		0,212		0,353	
T Critical one-tail	1,771		1,753		1,771		1,717		1,771		1,717	
P(T<=t) two-tail	0,148		0,866		0,754		0,268		0,425		0,706	
T Critical Two-tail	2,160		2,131		2,160		2,074		2,160		2,074	

Table F.2: Two-sample t-tests performed on the average participant score in the information findings sections. The purpose pf the two additional t-tests in Part 1.1 ("Question 1,2,4,5" and "Question 1-6") was to see wether there was any difference between the internet and laboratory participants. Based on initial F-tests to determine equality of variance, we assumed equal variance in all t-tests except in "Question 1,2,3,5" and the laboratory-test in part 1.2, where we assumed unequal variances. The grey rows indicates mean and p-values.

F.3 Likeability question

- **Table F.3** - Descriptive statistics for the 5-point likert questions in part 1.3 and 2.2 in the **laboratory experiment**.
- **Table F.4** - Descriptive statistics for the 5-point likert questions in part 1.3 and 2.2 in the **internet experiment**
- **Table F.5** - Two-sample t-tests with unequal sample size, performed on the average participant score in the likeability-questions.
- **Table F.6** - Two-sample t-tests with unequal sample size, performed on question 13-18 in both experiments.
- **Table F.7** - Two-sample t-tests with unequal sample size, performed on question 24-28 in both experiments.

Laboratory experiment **NL: Nutrition Label** **PT: Privacy Table**

Question:	*13*		*14*		*15*		*16*		*17*		*18*		***Tot: 13-17***	
Group:	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT
Mean	3,50	2,86	3,50	3,43	3,00	2,71	3,50	3,86	3,75	4,57	2,50	2,57	3,45	3,49
Standard Error	0,327	0,404	0,327	0,429	0,327	0,360	0,423	0,340	0,313	0,202	0,267	0,528	0,152	0,190
Median	4	2	3,5	4	3	2	3,5	4	4	5	2	2	4	4
Mode	4	2	3	4	2	2	2	4	4	5	2	2	4	4
Standard Deviation	0,926	1,069	0,926	1,134	0,926	0,951	1,195	0,900	0,886	0,535	0,756	1,397	0,959	1,121
Sample Variance	0,857	1,143	0,857	1,286	0,857	0,905	1,429	0,810	0,786	0,286	0,571	1,952	0,921	1,257
Kurtosis	0,000	-2,800	0,000	-1,227	-2,100	-1,687	-1,456	4,287	1,851	-2,800	0,875	0,167	-0,953	-1,366
Skewness	-1,440	0,374	0,000	-0,235	0,000	0,764	0,000	-1,569	-1,026	-0,374	1,323	1,079	-0,309	-0,294
Range	2	2	3	3	2	2	3	3	3	1	2	4	3	3
Minimum	2	2	2	2	2	2	2	2	2	4	2	1	2	2
maximim	4	4	5	5	4	4	5	5	5	5	4	5	5	5
Sum	28	20	28	24	24	19	28	27	30	32	20	18	138	122
Count	8	7	8	7	8	7	8	7	8	7	8	7	40	35
Conf. Level (95,0%)	0,774	0,989	0,774	1,049	0,774	0,880	0,999	0,832	0,741	0,494	0,632	1,292	0,307	0,385

Question:	*24*		*25*		*26*		*27*		*28*		*24-27*		***13-17, 24-27***	
Group:	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT
Mean	4,13	3,57	3,50	3,71	3,88	3,29	3,38	2,71	2,88	2,29	3,72	3,32	3,57	3,41
Standard Error	0,295	0,369	0,327	0,360	0,295	0,474	0,375	0,474	0,398	0,286	0,163	0,212	0,112	0,141
Median	4	4	4	4	4	4	4	2	3	2	4	4	4	4
Mode	4	4	4	4	3	4	4	4	4	2	4	4	4	4
Standard Deviation	0,835	0,976	0,926	0,951	0,835	1,254	1,061	1,254	1,126	0,756	0,924	1,124	0,947	1,116
Sample Variance	0,696	0,952	0,857	0,905	0,696	1,571	1,125	1,571	1,268	0,571	0,854	1,263	0,897	1,246
Kurtosis	-1,392	0,042	0,000	1,245	-1,392	0,521	3,937	-2,071	-0,989	-0,350	1,388	-0,597	-0,255	-0,976
Skewness	-0,277	-0,277	-1,440	-0,863	0,277	-1,450	-1,960	0,029	-0,488	-0,595	-0,958	-0,697	-0,564	-0,456
Range	2	3	2	3	2	3	3	3	3	2	4	4	4	4
Minimum	3	2	2	2	3	1	1	1	1	1	1	1	1	1
maximim	5	5	4	5	5	4	4	4	4	3	5	5	5	5
Sum	33	25	28	26	31	23	27	19	23	16	119	93	257	215
Count	8	7	8	7	8	7	8	7	8	7	32	28	72	63
Conf. Level (95,0%)	0,698	0,903	0,774	0,880	0,698	1,159	0,887	1,159	0,941	0,699	0,333	0,436	0,222	0,281

Table F.3: Descriptive statistics for the 5-point likert questions in part 1.3 and 2.2 in the laboratory experiment. The results for question 14 was inverted as it was negatively posed. The three grey columns presents the aggregate results for part 1 (question 13-17), part 2 (question 24-27) and both parts combined (question 13-17 and 24-27). Question 18 and 28 (the orange columns) was not included in the aggregate result as they were designed for spesific purposes.

Internet experiment **NL: Nutrition Label** **PT: Privacy Table**

Question:	*13*		*14*		*15*		*16*		*17*		*18*		***Tot: 13-17***	
Group:	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT
Mean	2,83	3,42	2,42	3,00	2,17	3,08	2,42	3,00	2,58	3,33	2,83	2,67	2,48	3,17
Standard Error	0,271	0,288	0,288	0,302	0,366	0,313	0,398	0,302	0,358	0,333	0,297	0,284	0,149	0,135
Median	3	3	2	3	2	3	2	3	2,5	3,5	2,5	2,5	2	3
Mode	3	3	2	2	1	3	1	2	4	4	2	2	2	2
Standard Deviation	0,937	0,996	0,996	1,044	1,267	1,084	1,379	1,044	1,240	1,155	1,030	0,985	1,157	1,044
Sample Variance	0,879	0,992	0,992	1,091	1,606	1,174	1,902	1,091	1,538	1,333	1,061	0,970	1,339	1,090
Kurtosis	-0,298	-0,654	-0,654	-0,856	-1,367	-0,238	-0,828	-0,856	-1,677	-1,473	-0,022	-0,980	-1,183	-1,041
Skewness	-0,412	0,274	0,274	0,574	0,596	0,837	0,582	0,574	-0,056	0,063	0,988	0,127	0,018	0,394
Range	3	3	3	3	3	3	4	3	3	3	3	3	4	3
Minimum	1	2	1	2	1	2	1	2	1	2	2	1	1	2
maximim	4	5	4	5	4	5	5	5	4	5	5	4	5	5
Sum	34	41	29	36	26	37	29	36	31	40	34	32	149	190
Count	12	12	12	12	12	12	12	12	12	12	12	12	60	60
Conf. Level (95,0%)	0,596	0,633	0,633	0,664	0,805	0,689	0,876	0,664	0,788	0,734	0,654	0,626	0,299	0,270

Question:	*24*		*25*		*26*		*27*		*28*		*24-27*		***13-17, 24-27***	
Group:	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT	NL	PT
Mean	3,33	3,25	3,08	3,50	2,83	3,33	3,17	3,17	3,92	3,00	3,10	3,31	2,76	3,23
Standard Error	0,333	0,279	0,229	0,289	0,297	0,256	0,322	0,297	0,358	0,369	0,147	0,137	0,109	0,096
Median	4	3	3	3,5	3	3	4	3	4	3	3	3	3	3
Mode	4	4	3	3	3	3	4	3	5	2	4	3	4	3
Standard Deviation	1,155	0,965	0,793	1,000	1,030	0,888	1,115	1,030	1,240	1,279	1,016	0,949	1,135	1,001
Sample Variance	1,333	0,932	0,629	1,000	1,061	0,788	1,242	1,061	1,538	1,636	1,031	0,900	1,287	1,002
Kurtosis	-0,055	-0,770	-1,261	-0,764	1,099	-0,254	-0,872	-0,022	-0,877	-0,856	-0,643	-0,773	-1,098	-0,965
Skewness	-0,787	0,136	-0,161	0,000	0,388	0,139	-0,858	0,810	-0,845	0,313	-0,343	0,256	-0,098	0,315
Range	4	3	2	3	4	3	3	3	3	4	4	3	4	3
Minimum	1	2	2	2	1	2	1	2	2	1	1	2	1	2
maximim	5	5	4	5	5	5	4	5	5	5	5	5	5	5
Sum	40	39	37	42	34	40	38	38	47	36	159	159	298	349
Count	12	12	12	12	12	12	12	12	12	12	48	48	108	108
Conf. Level (95,0%)	0,734	0,613	0,504	0,635	0,654	0,564	0,708	0,654	0,788	0,813	0,295	0,276	0,216	0,191

Table F.4: Descriptive statistics for the 5-point likert questions in part 1.3 and 2.2 in the Internet experiment. The results for question 14 was inverted as it was negatively posed. The three grey columns presents the aggregate results for part 1 (question 13-17), part 2 (question 24-27) and both parts combined (question 13-17 and 24-27). Question 18 and 28 (the orange columns) was not included in the aggregate result as they were designed for spesific purposes.

	Question 13-17				Question 24-27				Question 13-17+24-27			
	Laboratory		Internet		Laboratory		Internet		Laboratory		Internet	
	NL	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*
Mean	3,45	3,49	2,48	3,17	3,72	3,32	3,10	3,31	3,57	3,41	2,76	3,23
Variance	0,329	0,465	0,811	0,777	0,204	0,723	0,596	0,717	0,153	0,188	0,559	0,681
Observations	8	7	12	12	8	7	12	12	8	7	12	12
P (F-test - equality of variance)	0,328		0,473		0,061		0,382		0,393		0,374	
Pooled variance	0,391		0,794		0,444		0,657		0,169		0,620	
Hypothesized Mean Diff.	0		0		0		0		0		0	
df	13		22		13		22		13		22	
t Stat	-0,110		-1,879		1,152		-0,630		0,736		-1,469	
P(T<=t) one-tail	0,457		0,037		0,135		0,268		0,237		0,078	
T Critical one-tail	1,771		1,717		1,771		1,717		1,771		1,717	
P(T<=t) two-tail	0,914		0,074		0,270		0,535		0,475		0,156	
T Critical Two-tail	2,160		2,074		2,160		2,074		2,160		2,074	

The corresponding data used for the t-tests above. Each cell corresponds to a participants average score in the related part/experiment.

2,60	3,00	1,8	3,2	3,25	3,75	2,5	2,75	2,89	3,33	2,11	3,00
4,20	4,20	2,4	2,8	4,25	3,50	3,75	4	4,22	3,89	3,00	3,33
3,60	3,80	4	2,4	3,25	2,75	4	2,25	3,44	3,33	4,00	2,33
3,60	2,40	2,8	4,4	3,50	4,25	3,5	3,75	3,56	3,22	3,11	4,11
2,60	4,00	2,2	2,6	4,25	4,25	2,5	3,25	3,33	4,11	2,33	2,89
3,80	3,00	1,6	5	3,25	2,75	4	5	3,56	2,89	2,67	5,00
3,40	4,00	2,6	3,4	4,00	2,00	2,75	3,25	3,67	3,11	2,67	3,33
3,80		3,6	2	4,00		3,5	2	3,89		3,56	2,00
		2,4	3,4			3,75	3,25			3,00	3,33
		3,6	2,2			3	2,75			3,33	2,44
		1,8	3,6			2,5	4,25			2,11	3,89
		1	3			1,5	3,25			1,22	3,11

Table F.5: Two-sample t-tests with unequal sample size, performed on the average participant score in the likeability-questions. We initially conducted F-tests on the sample data to determine equality of variance (with results in the orange shaded line), and based on this we assumed equal variance for all t-tests (P >0.05 for all questions). Average results for the participants, as seen in the bottom section of the table, was measured across question 13-17 (Part 1.3), 24-27 (Part 2.2) and finally across 13-17 + 24-27 together. Question 14 was inverted as it was originally negatively posed. The grey shaded rows indicates mean and p-values.

	Question 13				Question 14				Question 15			
	Laboratory		Internet		Laboratory		Internet		Laboratory		Internet	
	NL	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*
Mean	3,50	2,86	2,83	3,42	2,50	2,57	3,58	3,00	3,00	2,71	2,17	3,08
Variance	0,857	1,143	0,879	0,992	0,857	1,286	0,992	1,091	0,857	0,905	1,606	1,174
Observations	8	7	12	12	8	7	12	12	8	7	12	12
P (F-test - equality of variance)	0,355		0,422		0,302		0,439		0,465		0,306	
Pooled variance	0,989		0,936		1,055		1,042		0,879		1,390	
Hypothesized Mean Diff.	0		0		0		0		0		0	
df	13		22		13		22		13		22	
t Stat	1,249		-1,477		-0,134		1,402		0,589		-1,904	
P(T<=t) one-tail	0,117		0,077		0,448		0,088		0,283		0,035	
T Critical one-tail	1,771		1,717		1,771		1,717		1,771		1,717	
P(T<=t) two-tail	0,234		0,154		0,895		0,175		0,566		0,070	
T Critical Two-tail	2,160		2,074		2,160		2,074		2,160		2,074	

	Question 16				Question 17				Question 18			
	Laboratory		Internet		Laboratory		Internet		Laboratory		Internet	
	NL	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*
Mean	3,50	3,86	2,42	3,00	3,75	4,57	2,58	3,33	2,50	2,57	2,83	2,67
Variance	1,429	0,810	1,902	1,091	0,786	0,286	1,538	1,333	0,571	1,952	1,061	0,970
Observations	8	7	12	12	8	7	12	12	8	7	12	12
P (F-test - equality of variance)	0,253		0,185		0,119		0,409		0,066		0,442	
Pooled variance	1,143		1,496		0,555		1,436		1,209		1,015	
Hypothesized Mean Difference	0		0		0		0		0		0	
df	13		22		13		22		13		22	
t Stat	-0,645		-1,168		-2,203		-1,533		-0,126		0,405	
P(T<=t) one-tail	0,265		0,128		0,024		0,070		0,451		0,345	
T Critical one-tail	1,771		1,717		1,782		1,717		1,771		1,717	
P(T<=t) two-tail	0,530		0,255		0,048		0,139		0,902		0,689	
T Critical Two-tail	2,160		2,074		2,179		2,074		2,160		2,074	

Table F.6: Two-sample t-tests with unequal sample size, performed on question 13-18 in both experiments. (I.e. two t-tests were performed for each question: the first on the laboratory experiment data (NL vs PT) and the other on the internet experiment data (NL vs. PT). We initially conducted F-tests on the sample data to determine equality of variance (results in the orange shaded line), and based on this we assumed equal variance for all t-tests (P >0.05 for all questions). The data for question 14 was not inverted.

	Question 24				Question 25				Question 26			
	Laboratory		Internet		Laboratory		Internet		Laboratory		Internet	
	NL	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*
Mean	4,13	3,57	3,33	3,25	3,50	3,71	3,08	3,50	3,88	3,29	2,83	3,33
Variance	0,696	0,952	1,333	0,932	0,857	0,905	0,629	1,000	0,696	1,571	1,061	0,788
Observations	8	7	12	12	8	7	12	12	8	7	12	12
P (F-test - equality of variance)	0,343		0,281		0,465		0,227		0,156		0,315	
Pooled variance	0,815		1,133		0,879		0,814		1,100		0,924	
Hypothesized Mean Diff.	0		0		0		0		0		0	
df	13		22		13		22		13		22	
t Stat	1,185		0,192		-0,442		-1,131		1,085		-1,274	
P(T<=t) one-tail	0,129		0,425		0,333		0,135		0,149		0,108	
T Critical one-tail	1,771		1,717		1,771		1,717		1,771		1,717	
P(T<=t) two-tail	0,257		0,850		0,666		0,270		0,297		0,216	
T Critical Two-tail	2,160		2,074		2,160		2,074		2,160		2,074	

	Question 27				Question 28				Question 28 total	
	Laboratory		Internet		Laboratory		Internet		Laboratory + Internet	
	NL	*PT*	*NL*	*PT*	*NL*	*PT*	*NL*	*PT*	NL	PT
Mean	3,38	2,71	3,17	3,17	2,88	2,29	3,92	3,00	3,50	2,74
Variance	1,125	1,571	1,242	1,061	1,268	0,571	1,538	1,636	1,632	1,316
Observations	8	7	12	12	8	7	12	12	20	19
P (F-test - equality of variance)	0,334		0,399		0,175		0,460		0,326	
Pooled variance	1,331		1,152		0,946		1,587		1,478	
Hypothesized Mean Diff.	0		0		0		0		0	
df	13		22		13		22		37	
t Stat	1,107		0,000		1,170		1,782		1,959	
P(T=t) one-tail	0,144		0,500		0,131		0,044		0,029	
T Critical one-tail	1,771		1,717		1,771		1,717		1,687	
P(T<=t) two-tail	0,289		1,000		0,263		0,089		0,058	
T Critical Two-tail	2,160		2,074		2,160		2,074		2,026	

Table F.7: Two-sample t-tests with unequal sample size, performed on question 24-28 in both experiments. (I.e. two t-tests were performed for each question: the first on the laboratory experiment data (Nutriton L vs Privacy T) and the other on the internet experiment data (Nutrition L vs. Privacy T). We initially conducted F-tests on the sample data to determine equality of variance (results in the orange shaded line), and based on this we assumed equal variance for all t-tests (P >0.05 for all questions).

F.4 Detailed statitistics for the timing results

	Part 1		Part 2		Total	
	NL	PT	NL	PT	NL	PT
Mean	5,25	5,00	2,00	2,83	3,86	4,00
Variance	4,786	0,333	0,800	0,167	5,670	1,500
Observations	8	7	7	6	15	13
P (F-test - equality of variance)	0,002		0,055		0,014	
Pooled Variance	x		0,483		x	
Hypothesized Mean Difference	0		0		0	
df	8		10		20	
t Stat	0,311		-2,076		-0,198	
P(T<=t) one-tail	0,382		0,032		0,423	
T Critical one-tail	1,860		1,812		1,725	
P(T<=t) two-tail	0,764		0,065		0,845	
T Critical Two-tail	2,306		2,228		2,086	

	Part 1		Part 2		Total	
	NL	PT	NL	PT	NL	PT
Mean	5,25	5,00	2,00	2,83	3,86	4,00
Standard Error	0,773	0,218	0,365	0,167	0,636	0,340
Median	4,5	5	2	3	3,5	4
Mode	4	5	3	3	4	5
Standard Deviation	2,188	0,577	0,894	0,408	2,381	1,225
Sample Variance	4,786	0,333	0,800	0,167	5,670	1,500
Kurtosis	3,224	3,000	-1,875	6,000	2,350	-1,354
Skewness	1,651	0,000	0,000	-2,449	1,274	0,000
Range	7	2	2	1	9	4
Minimum	3	4	1	2	1	2
maximim	10	6	3	3	10	6
Sum	42	35	12	17	54	52
Count	8	7	7	6	15	13
Conf. Level (95,0%)	1,829	0,534	0,939	0,428	1,375	0,740

Table F.8: Two-sample t-tests and corresponding descriptive statistics for the mean time score in the laboratory experiment. We initially conducted F-tests on the sample data to determine equality of variance (with results in the orange shaded line), and based on this we assumed unequal variance for the t-tests on Part 1 and both parts combined (Total), and equal variances for Part 2.