

GRAPHICAL REPRESENTATION OF ONLINE PRIVACY RISKS WITHIN  
TRUSTED CONTEXTS

A Thesis  
Presented to  
The Faculty of Graduate Studies  
of  
The University of Guelph

by

ANDREW BERRY

In partial fulfilment of requirements  
for the degree of  
Master of Science  
December, 2010

©Andrew Berry, 2010

## ABSTRACT

# GRAPHICAL REPRESENTATION OF ONLINE PRIVACY RISKS WITHIN TRUSTED CONTEXTS

**Andrew James Berry**

**University of Guelph, 2010**

**Advisor:**

**Professor Judi McCuaig**

This thesis is an investigation of the use of graphical privacy indicators to represent online privacy risks in trusted contexts. Results indicate that the use of graphical privacy indicators as used in this thesis may not have an effect on an individual's investigation into their privacy risks, or on the amount of data an individual is willing to submit. Most participants chose to ignore the availability of further information about risks to their privacy, while also choosing to submit confidential information. These results indicate that graphical representation of privacy risks as used in this thesis may not be an appropriate method of communication for privacy-aware user agents, and that other methods of communication and graphical representation should be investigated.

## Acknowledgements

Many people were involved in the completion of this thesis. I would like to thank my advisor, Professor Judi McCuaig, for her hours of support and encouragement. My thesis committee of Professor Blair Nonnecke and Professor Michael Wirth provided excellent feedback in the development of this document. Barrow Baldwin also provided assistance in reviewing this document for which I am grateful. My wife, Julia Baldwin, was crucial in helping to organize my experimental sessions and in ensuring that they were executed smoothly. Her patience and support during the critical components of this thesis are greatly appreciated. Finally, I would like to thank all of the volunteers who acted as participants in each experiment for their time and effort.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Thesis Outline . . . . .	5
<b>2</b>	<b>A Review of Electronic Privacy</b>	<b>7</b>
2.1	Introduction . . . . .	7
2.2	Privacy Threats . . . . .	9
2.2.1	Physical and Virtual Privacy . . . . .	10
2.2.2	Manipulating Privacy Perceptions Through User Interface Design . .	13
2.2.3	Data Submission on the Web . . . . .	16
2.2.4	Summary of Privacy Threats . . . . .	16
2.3	Privacy Description Languages . . . . .	17
2.3.1	Platform for Privacy Preferences . . . . .	18
2.3.2	A P3P Preference Exchange Language . . . . .	20
2.3.3	Enterprise Policy Authorization Language . . . . .	20
2.3.4	Extensible Access Control Markup Language . . . . .	21
2.3.5	Summary of Privacy Description Languages . . . . .	22
2.4	Privacy Interactions . . . . .	24

2.4.1	Privacy Policies . . . . .	24
2.4.2	Gender and Online Privacy Attitudes . . . . .	26
2.4.3	Privacy Contexts . . . . .	27
2.4.4	Summary of Privacy Interactions . . . . .	28
2.5	Privacy Representation . . . . .	28
2.5.1	Components of Icons . . . . .	29
2.5.2	Icon Selection in Previous Work . . . . .	30
2.5.3	Icon Testing Methodology . . . . .	32
2.5.4	Summary of Privacy Representation . . . . .	33
2.6	Summary of Electronic Privacy . . . . .	34
<b>3</b>	<b>Experiment Design</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Data Submission on the Web . . . . .	38
3.2.1	Selecting Representative Sites . . . . .	38
3.2.2	Techniques of Data Submission . . . . .	39
3.2.3	Defining Data Submission . . . . .	39
3.2.4	Crawler Results . . . . .	40
3.3	Simplifying Machine-Level Communications . . . . .	42
3.4	Evaluating the Website Trust Component of Context . . . . .	45
3.5	Evaluating Graphical Privacy Indicators . . . . .	46
3.5.1	Participant Selection . . . . .	46
3.5.2	Experiment Procedure . . . . .	46
3.6	Experiment Design Summary . . . . .	50

<b>4</b>	<b>Experimental Results</b>	<b>52</b>
4.1	Evaluating Website Trust . . . . .	52
4.2	Testing Graphical Privacy Indicators . . . . .	54
4.2.1	Results Summary . . . . .	55
4.2.2	Analysis Procedure . . . . .	57
4.2.3	Privacy Investigations Data Analysis . . . . .	58
4.2.4	Submitted Answers Data Analysis . . . . .	65
<b>5</b>	<b>Conclusions and Future Work</b>	<b>70</b>
5.1	Conclusions . . . . .	70
5.1.1	Conclusions in Evaluating Website Trust . . . . .	70
5.1.2	Conclusions in Privacy Investigations . . . . .	71
5.1.3	Conclusions in Submitted Answers . . . . .	73
5.2	Future Work . . . . .	75
5.2.1	Privacy Agents . . . . .	75
5.2.2	Refining Privacy Risk Graphical Representation . . . . .	78
5.2.3	Data Submission on the Web . . . . .	80
5.2.4	Expertise Evaluations . . . . .	82
5.2.5	User comments and Observations . . . . .	88
5.3	Conclusion . . . . .	89
	<b>References</b>	<b>90</b>
<b>A</b>	<b>Survey Questions</b>	<b>98</b>
A.1	Privacy Contexts Survey Questions . . . . .	98
A.2	Graphical Privacy Indicator Pre-Survey . . . . .	100

A.3	University Account Registration . . . . .	102
A.4	Graphical Privacy Indicator Post-Survey . . . . .	104
<b>B</b>	<b>Simple Privacy Framework</b>	<b>105</b>
B.1	Simple Privacy Framework Example . . . . .	105
B.2	Simple Privacy Framework OWL Schema . . . . .	107

# List of Figures

2.1	Browser chrome and viewport. . . . .	15
2.2	Common error icons. . . . .	31
2.3	Privacy progress bar. . . . .	31
3.1	Account registration screen. . . . .	48
3.2	Instant messaging privacy risks. . . . .	49
4.1	Graphical Privacy Indicator Status by Total Privacy Investigations. . . . .	58
4.2	Graphical Indicator Status by Green Privacy Investigations. . . . .	59
4.3	Graphical Indicator Status by Yellow Privacy Investigations. . . . .	59
4.4	Graphical Indicator Status by Red Privacy Investigations. . . . .	59
4.5	Gender by Total Privacy Investigations. . . . .	60
4.6	Gender by Green Privacy Investigations. . . . .	61
4.7	Gender by Yellow Privacy Investigations. . . . .	61
4.8	Gender by Red Privacy Investigations. . . . .	62
4.9	Degree by Total Privacy Investigations. . . . .	62
4.10	Degree by Green Privacy Investigations. . . . .	63
4.11	Degree by Yellow Privacy Investigations. . . . .	63
4.12	Degree by Red Privacy Investigations. . . . .	64



4.13	Age by Total Privacy Investigations. . . . .	64
4.14	Internet Expertise Rating by Total Privacy Investigations. . . . .	65
4.15	Graphical Indicator Status by Total Submitted Questions. . . . .	66
4.16	Total Privacy Investigations by Total Submitted Questions. . . . .	66
4.17	Gender by Total Submitted Questions. . . . .	67
4.18	Age by Total Submitted Questions. . . . .	68
4.19	Major by Total Submitted Questions. . . . .	68
4.20	Expertise Rating by Total Submitted Questions. . . . .	69
5.1	Firefox 3 SSL Certificate Error. . . . .	79
5.2	Total form Tags by Total script Tags. . . . .	81
5.3	Gender by Expertise Rating. . . . .	84
5.4	Degree by Expertise Rating. . . . .	85
5.5	Expertise Rating by Years Online. . . . .	86
5.6	Age by Age First Online. . . . .	86

# Chapter 1

## Introduction

Privacy is a social state wherein an individual is secluded from other individuals (Merriam Webster, 2010). Self-identity is critical to how we understand and interpret our interactions with society. Giving up parts of our private information in certain aspects of our lives is required for society to function. Allowing others into to our private lives provides a safe space for social growth, and is how we define relationships with other individuals (Petronio, 1994).

The development of integrated networked technology forced our society to reevaluate how it interpreted and applied privacy in our daily lives. While many networked technologies are collecting greater amounts of private information, society still expects individuals to be able to control the use of their disclosed information. Such control is difficult for individuals to exert, as they aren't directly interacting with another individual who is collecting that information. Individuals also tend to be less selective about the information they disclose to electronic systems (Berendt, Günther, & Spiekermann, 2005). For proponents and developers of these integrated and networked systems, it is crucial that we build systems that

allow privacy preferences to be defined by individuals and enforced.

Networked systems enable new methods for privacy preferences to be violated. The past decade has seen a proliferation of Global Positioning System (GPS) devices, cameras, and high-speed networks in our daily lives. GPS devices allow for exact locations of individuals to be transmitted over networks. Networked Closed Captioned Television (CCTV) systems allow for individuals not carrying electronic devices with tracking capabilities to be tracked and monitored. High-speed networks reduce or eliminate any delays in transmitting private information. Communication and messaging systems such as Facebook and Twitter are expanding beyond home computers and onto mobile devices. The integration of these networked technologies challenges our traditional notions of privacy. Twitter has become famous for its users posting intimate details of their lives, and for its use by oppressive regimes to track and detain dissidents (Morozov, 2009). Technology now allows for any individual to record and disseminate private information to a wide audience, without concern for social norms or the preferences of those affected. “Smart” buildings are being constructed with integration into networked services for power, connectivity, utilities, and security (Akyildiz, Su, Sankarasubramaniam, & Cayirci, 2002). There is a need for these networked systems to be able to interpret and respect the privacy preferences of individuals.

A privacy policy is a contract used to communicate how an organization will use private data disclosed by an individual. Privacy policies are often used by online organizations, but are also part of many offline contracts between organizations and individuals. For example, most loyalty programs will include a privacy policy. Privacy policies are also used by government agencies to communicate how they handle personal information. As interaction with government agencies is required to access many basic services, such as health care and education, most citizens will have accepted, and grown accustomed to, the privacy policies

of these organizations at many points in their daily lives.

While privacy policies were initially voluntary, many jurisdictions (such as the European Union) now enforce specific requirements in the terms of privacy policies, complicating the creation and application of privacy policies. This is often the case for government agencies, which are required to ensure a specific level of confidentiality of private information (Department of Justice Canada, 2010b). When data is disclosed to a third party, organizations are often required by law to notify and obtain the consent of the individual who is the source of the data (European Parliament, 1995). The requirements and implementation details specified by law varies greatly across different jurisdictions, increasing the difficulty in creating and interpreting privacy policies.

A privacy policy is a legally binding document (Anton, Bertino, Li, & Yu, 2007). This creates a significant barrier for individuals attempting to interpret the privacy policy. In order to be legally binding, privacy policies are typically many tens of pages long, and they may vary slightly between each organization (McDonald & Cranor, 2009). Users often ignore or blindly accept the contents of a privacy policy due to their length (McDonald & Cranor, 2009). While most privacy policies are complex and obtuse, they are a necessary instrument in relations between large organizations and individuals due to the legislation in force requiring such policies. Any user agent representing the meaning of privacy policies must both be accurate in the representation of the policy, and enable users to easily understand the policy.

The difficulty in interpreting a privacy policy means that most users don't bother to enforce their privacy preferences. Privacy policies are too difficult and lengthy to read. Even though the presence of privacy policies gives users the information needed to enforce their privacy

preferences, such information is rarely utilized. This exposes users to unwanted uses of their private information, often without their knowledge.

For an electronic system to be able to enforce privacy preferences, it must be able to express privacy preferences accurately. Individuals usually have privacy preferences they wish to see respected, even if they don't understand privacy policies. Privacy preferences have previously been defined as the behaviours an individual applies in managing solitude, anonymity, intimacy, and reserve (Marshall, 1974). This definition does not match current uses of the term, where each privacy preference is a discrete entity. This thesis defines a privacy preference as a three-part definition including an object, a subject, and an action. For example, a subject ("I am willing to disclose my phone number"), a destination subject ("Acme Anvil Corporation"), and a willingness to disclose ("Only if they give me a coupon") is combined to create a privacy preference. A privacy preference is associated with a willingness to have it violated in specific contexts for a given benefit, resulting in the disclosure of private information. Determining the requirements for information disclosure is critical for electronic systems to accurately enforce privacy preferences.

Much of the difficulty in representing a preference is in the conditions required to allow disclosure of the data. Users can have a large number of conditions for disclosure based on context, and may not be consciously aware of those conditions. Economic factors, previous relations with the organization, or even the time of day may influence the willingness to disclose private information. Until we have a method of communicating privacy policies and their effects clearly to individuals, preference enforcement is a difficult, if not impossible, problem to solve.

A method of allowing individuals to assess their online privacy risks in a quick and concise

manner would give online users the tools needed to enforce their privacy preferences. A system capable of interpreting and displaying privacy policies and their relationship to privacy preferences would be a significant improvement in electronic privacy management.

Graphical representations of privacy policies might allow for greater awareness of privacy policies on the part of individuals. By bringing privacy policies to the forefront of an interaction, users might be able to enforce their privacy preferences with greater accuracy.

Communicating privacy policies to online users is only the first step. Privacy preference violations by online organizations need to be perceived as immediate and real privacy threats. Many online users believe in “practical obscurity”, assuming that their data isn’t important within a group of millions of users (C. N. Davis, 2005). In fact, the development of automated systems to extract and act upon an individuals data allows such companies to act on every datum submitted by an individual. If electronic systems could translate virtual privacy threats into the perception of a concrete risk, then users would be able to identify threats relating to their own preferences as well as protect their own private information.

## **1.1 Thesis Outline**

How can designers, developers, and content creators of websites communicate privacy policies to users in an effective and universal manner? To develop an effective system of representing privacy it is important to understand privacy policies, privacy preferences, and privacy threats. Current methods of electronically representing privacy must be analyzed and explored. Categories of trusted websites must be determined to ensure that the context of the website design used has a controlled level of trust. The use of data collection on the

web (such as submitting forms or dynamically with JavaScript) must be tested to ensure that data collection is a key element of online activities of users. To solve the issues encountered with current privacy representation and transmissions methods, this thesis proposes a prototype schema of the Simple Privacy Framework (SPF), a simplified Extensible Markup Language (XML) framework to describe privacy policies. This allows the generation of indicators of privacy risks from simplified privacy representations. Such indicators could consist of sound, text, or graphical indicators in any combination. Graphical icons may be an effective method of representing privacy risks. It is expected that when presented with graphical representations of privacy risks within trusted contexts, users will be more likely to investigate the effects of such risks on their privacy preferences, giving them the information they need to enforce their preferences.

The remainder of this thesis will discuss the issues and possible solutions in the field of representing privacy risks to users online. Chapter 2 will discuss the current threats to a user's privacy online, the issues with current methods used to describe privacy in a machine-readable language, how users behave during a privacy interaction online, and how privacy representations have been investigated in previous research. Chapter 3 will discuss the experiment design used to test online data submission, machine level communication of privacy policies, the effect of website trust on privacy contexts, and the use of graphical representations of privacy risks. Chapter 4 will provide detailed results and analysis from the experiment. Finally, Chapter 5 will discuss the conclusions from this work and future areas of investigation.

## Chapter 2

# A Review of Electronic Privacy

### 2.1 Introduction

Day to day use of the web has deviated from the vision of the founders of the web. Initially used as a hyperlinked publishing mechanism, the web has become an application delivery platform, allowing individuals to access dynamic, high quality services from anywhere in the world. The use of the web in this manner exposes individuals to many privacy threats. Many online users don't understand the threats to their privacy, or don't have the skills to protect their information. This leads to a situation where privacy disclosure is common, and very difficult to control without severely hampering an online experience.

One reason for this situation is the difficulty in representing privacy risks. Given the dynamic nature of the web, it is very difficult for automatic systems to represent privacy risks. While there are some standards that touch on privacy representation at a machine level, none of them are suitable for use on the web. Without some method to reliably transmit



privacy data online, it is very difficult for a user agent to accurately alert a user of any risks.

Even if there was a usable machine-level method of privacy transmission, privacy preference management is still a difficult problem to solve. A system of preference management would need to understand an individual's preferences, which many individuals don't themselves understand. It is common for online users to not fully understand the risks of information disclosure due to the virtual nature of the transaction. If a user agent is to represent privacy risks using graphical indicators, it must both be aware of the preferences of the user, as well as represent its information in a noticeable, yet not annoying, manner.

Protection of personal data gathered by electronic systems is crucial to the continued acceptance of electronic transactions by individuals (Hoffman, Novak, & Peralta, 1999). Concepts of privacy, including limits on data retention, distribution, and analysis, are recent ideas when compared to property rights and physical safety (Warren & Brandeis, 1890). With the availability of low-cost computational resources, even the smallest of organizations can use private data to analyze, categorize, and customize their interactions with individuals.<sup>1</sup> While these uses of private data can be beneficial to individuals, sometimes the benefits to the organization collecting the data can result in the mishandling of private information, as when information is sold without permission. Common benefits include data mining to research market trends, or selling the data to other organizations.

In order to determine how to optimize online privacy interactions to protect users, three areas must be explored. First, it is important to understand privacy interactions and how they occur in online environments. Second, the encoding of privacy into machine-readable

---

<sup>1</sup>As of August 5th, 2010, the popular SourceForge website listed 683 open source projects in the Customer Relationship Management (CRM) category. Any organization with internet access could use these tools to collect and analyze data from individuals.

formats will be discussed. Finally, methods of representing privacy policies and associated risks to users will be explored.

## 2.2 Privacy Threats

The threats to an individual’s privacy have significantly changed over the past century. With developments in communications technology since the turn of the 20th century, expectations of individual privacy have changed. Newspapers and mass media caused the general public to become concerned about the public distribution of private information (Warren & Brandeis, 1890). Concern about the broadcast of “idle gossip” and “the details of sexual relations” to the public at large led to the definition of privacy as a right, just as inalienable as the right to physical safety (Warren & Brandeis, 1890).

Computer technology has further changed individual expectations of privacy. The decrease in cost for long term data retention and the ability to process archived data for new insights has further the threats to an individual’s privacy.<sup>2</sup> No longer does information gathered from business interactions stay contained within the mind of an individual, or the confines of a single store. Private information is now automatically gathered, catalogued, and potentially exposed to many thousands of people. A legal framework for privacy in business transactions is a modern development. Just as the right to privacy was recognized for those subject to media attention, it subsequently was recognized as a right for all individuals involved in business transactions (Council of Europe, 2010) (United States Department of Justice, 1974). These rights have evolved to address new threats to privacy, especially those

---

<sup>2</sup>In October, 2009, a laptop containing 33,000 patient records was stolen from the United Kingdom’s National Health Service (Public Service, 2010). As storage density continues to increase, the possibility of significant data breaches caused by day-to-day crimes such as car thefts will grow.

threats brought by the development of low cost computing technology. The combination of technology and legislation make managing privacy threats very difficult for most online users.

The specific circumstances surrounding an online interaction can have a significant effect on how users perceive threats to their privacy (Milne & Culnan, 2004). While many websites contain privacy policies, the language used to describe them is often opaque (McDonald & Cranor, 2009). When evaluating the privacy threats of a given website, users draw on a range of factors, including the context (website trustworthiness, the user’s physical location, time of day, and so on) of the interaction, to determine how likely a website is to respect their privacy preferences (Milne & Culnan, 2004).

### **2.2.1 Physical and Virtual Privacy**

The privacy expectations of users in an online environment rarely match their ability to properly recognize threats to their privacy (Berendt et al., 2005). This leads users to assume that their personal information is protected, even when it is not. Even when privacy threats are recognized by users, they often fail to preserve their privacy in an effective manner due to misplaced trust, lack of recognition of privacy threats, complexity of the interaction, or a lack of control (Berendt et al., 2005).

One example of this is Berendt’s use of an avatar to put users into a state of comfort, where they are more likely to disclose information they consider to be private (Berendt et al., 2005). “Luci”, the shopping assistant, was built to ask occasional personal questions unrelated to the shopping experience. By framing them within the context of the user’s actions, participants became more likely to answer such questions.

Another common finding is that users often disclose more information than they intend on social networking websites. For example, Facebook users often post pictures of themselves drinking alcohol, as the site is designed around the promise that only “friends” can view uploaded content (Christofides, Muise, & Desmarais, 2009). A “friend” on Facebook is very different than a friend in real life, but the use of the term helps to draw on those previous physical experiences. By building a website to convey security and privacy through its user interface design, yet authorize further use of the data through a privacy policy, it is possible to convince users to disclose more information than they intend.

It is theorized that the use of a virtual agent to link the current interaction to memories of physical experiences makes the actions more “available” (Berendt et al., 2005). In the physical world, questions such as those asked by Luci pose little harm, as it’s unlikely that a salesperson is actively recording interactions to a database. In the online world, every interaction can easily be recorded and mined, causing the same interaction to have vastly differing consequences. While the social context of interacting with a sales person is the same, the virtual nature of such an interaction presents privacy risks that are not immediately noticeable to the average user due to the persistence and accessibility of the collected data.

There are many methods by which privacy can be maintained within the physical world, but these methods often don’t apply in the virtual world. The most obvious is the concept of a minimal distance between individuals. Increased trust usually results in closer physical boundaries (Little, 1965). Our lives are built around the private spaces we create. Homes are protected by most societies as a private, secure space for individuals. Shared transportation sets a value on privacy through the increased cost for larger or segregated seating. Even our workplaces reserve privacy as a benefit, with enclosed offices being re-

served for specialized personnel. When physical entities are replaced with virtual entities, our privacy expectations and privacy preference enforcement strategies begin to fall apart (Patil & Kobsa, 2005).

Most users have difficulties in mapping their privacy preferences onto actions within the virtual world. For example, users might not want managers at their workplace to use their computer out of privacy concerns, but would be willing to have the same managers access their computer over a network. Patil and Kobasa note that users of instant messaging software are more likely to understand the risk of viewing a monitor displaying private conversations directly than the risk of software logging communications over a network (Patil & Kobsa, 2005). Users often assume that the presence of a privacy policy means that their privacy is protected, without investigating the content of the privacy policy (Resnick & Montania, 2003). Users don't understand how to protect their private information, or how to determine how their information will be used (McDonald & Cranor, 2009). Without these skills, users will continue to be unsuccessful at enforcing their privacy preferences online.

The benefits and consequences of disclosing information online are often unclear to users. Many users are unhappy with the privacy protections offered by online environments, and see greater information security in offline commerce interactions (Burke, 2002). The use of personalized recommendations generated from collected data is often a privacy violation that concerns users about online ecommerce interactions (Chellappa & Sin, 2005). The use of personalized user accounts online has also led to the application of variable pricing by some retailers, which caused some users to feel like they were being gauged (Chellappa & Sin, 2005). In physical interactions, there is often a tangible benefit in disclosing private information. Online, the benefits of information disclosure are either unclear or nonexistent.

For users to feel comfortable in disclosing private information online, they must understand both the benefits of the disclosure and the use of the data they disclose.

A system of describing privacy preferences and controls using “virtual walls” allows users to accurately describe privacy preferences (Kapadia, Henderson, Fielding, & Kotz, 2007) when sensors are present within physical spaces. Participants were presented with an interface allowing them to draw walls around a physical space and indicate how transparent each wall was to different entities. The metaphor of walls matches the intuitive understanding of privacy within the physical world. Applying such a system to online interactions may allow users to describe privacy preferences and ensure their enforcement with greater accuracy and reduced effort. This would require virtual contexts to be accurately and automatically mapped to physical contexts that could be displayed to a user, removing burden of context mapping from the user. While this is an interesting area of research, the mapping of physical to virtual contexts was deemed by the author to be too complicated for most online interactions.

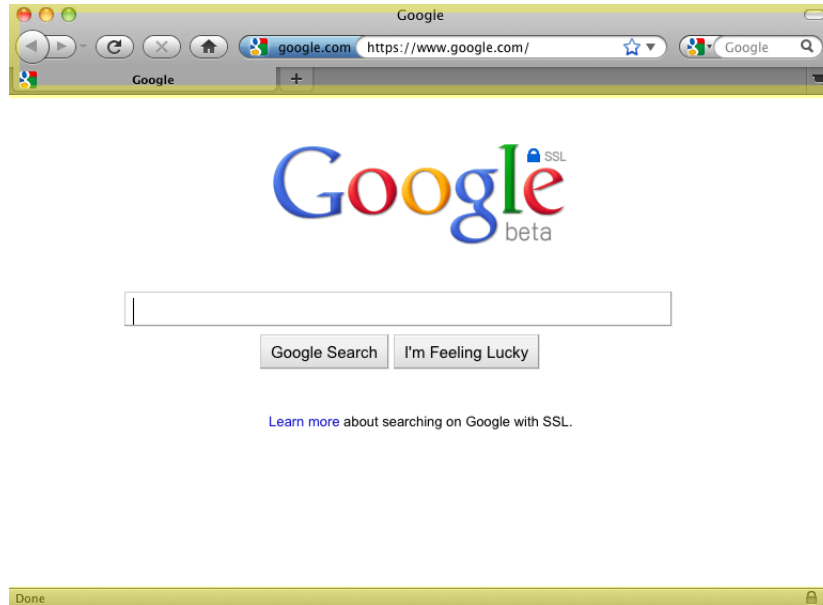
### **2.2.2 Manipulating Privacy Perceptions Through User Interface Design**

Even if interactive systems are able to respect and describe privacy access controls, users do not often understand and behave according to their own privacy preferences. It is common for users of interactive systems to disclose information in contradiction of their own privacy preferences when interacting with professionally designed web sites (Resnick & Montania, 2003). When a user interface is designed to elicit information from a user, it becomes even more difficult for users to enforce their privacy preferences (Downs, Holbrook, & Cranor, 2006).

Social networking websites are at the forefront of exploiting design principles to elicit private information. By including prominent links to customer service, help, and privacy policies, user trust in an organization is greatly increased (Resnick & Montania, 2003). It is important to note that is not the content of linked pages that causes the increase in trust; it is the prominence of the links themselves that is the critical factor (Resnick & Montania, 2003). Most users of social networking websites have the same difficulties as ecommerce users in enforcing their privacy preferences (Christofides et al., 2009).

Websites will often target novice users or those with minimal online literacy, as these users are less likely to understand the risks of information disclosure (Jarmo & Parkkinen, 2001). Using text instead of images can be used to impart trust (Jarmo & Parkkinen, 2001). Appropriate use of white space is another trustworthy element designers can use on web pages (Jarmo & Parkkinen, 2001). Untrustworthy elements include images, cartoon-like graphics, and undefined borders (Jarmo & Parkkinen, 2001). Novice users only consider what is visible in their understanding of a website (Jarmo & Parkkinen, 2001). This can be used to design web pages where the trustworthy elements are in key locations such as the primary navigation or initial viewport. The untrustworthy elements can be placed in low-visibility areas to meet legal requirements while avoiding close inspection by users.

Many trustworthy organizations use elements considered untrustworthy in their interactions with users (Downs et al., 2006). Most commonly, legitimate URLs redirect or point to unrelated content, making them difficult for users to authenticate (Downs et al., 2006). This makes it very difficult for users to determine the trustworthiness of an online entity. Many sites use privacy or security seals within their content to impart trustworthiness, as many users completely ignore browser chrome (Dhamija, Tygar, & Hearst, 2006). This



**Figure 2.1: Browser chrome and viewport.** The browser chrome is highlighted in yellow, and its contents can not be modified by the current web page. Anything within the viewport is under the control of the current web page. Users often trust SSL icons (such as the padlock in the Google logo) within the viewport.

further leads to users trusting the content of the viewport, even though the content of the viewport may not be trustworthy at all. The distinction between the chrome and the viewport, as shown in Figure 2.1, is confusing to most users (Downs et al., 2006). Even privacy researchers can make mistakes when distinguishing a legitimate communication from an attempted scam (Downs et al., 2006). For example, one researcher lost telephone service after disregarding an email as a phishing attempt.

A minimal standard of design (such as colour and shape) for displaying information in a secure and trustworthy manner would be a significant improvement on current methods to communicate privacy risks. Until such a standard of visual design has been defined and implemented, users will continue to place their information at risk due to the poor practices encouraged by legitimate organizations (Downs et al., 2006).



### **2.2.3 Data Submission on the Web**

The work previously discussed relating to privacy interactions evaluates users within controlled situations such as e-commerce or social networking scenarios. A failing of the previous work in this field is that it assumes the web is a publishing mechanism, and that data submission is less common than data retrieval. For example, no previous research testing ecommerce scenarios examines the impact of integrating ecommerce with social networking features, a functionality commonly used today. If efforts are going to be made to protect users from accidentally submitting data to organizations in violation of their privacy preferences, it is important to determine how commonly users encounter web pages that present the potential to collect their data. If the opportunities for data submission are as common, or more common than data retrieval, then the development of methods to protect users privacy is especially important.

Components of the work presented in this thesis depend on the assumption that submitting data is a common, if not required, interaction online. No previous work has confirmed this assumption. In order to ensure that the assumption was correct, a study of the use of data collection technologies was conducted. For details of this study, see Chapter 3.

### **2.2.4 Summary of Privacy Threats**

The threats to an individual's privacy have changed significantly with the development of communications technology. Advances in technology have allowed for the storage and processing of private information cheaply and quickly by most organizations. Significant legislation has been implemented in response to these threats that protect an individual's privacy in their interactions with government and private organizations.

Individuals often have difficulties in identifying the threats to their privacy in a virtual environment. The threats to their privacy may not be obvious, or they may have no prior experience to draw from. Individuals often mistake social interactions online as being identical in threat to social interactions offline, leading to information disclosures.

Website users can be manipulated and encouraged to disclose personal information beyond their privacy preferences. Design elements within a website's content are one of the most powerful methods of generating trust. When websites use cues from the physical world in their user interfaces to manipulate trust, it further distorts users abilities to accurately judge privacy threats.

## **2.3 Privacy Description Languages**

This thesis defines a privacy description language as a machine processable representation of a privacy policy. Such a language allows privacy policies to be expressed, transmitted, and interpreted by electronic systems. Most efforts use XML to define the language. The World Wide Web Consortium (W3C) has played a central role in organizing efforts to develop and deploy XML-based privacy description languages. The W3C has co-ordinated efforts towards the development of Platform for Privacy Preferences (P3P), A P3P Preference Exchange Language (APPEL), and Extensible Access Control Markup Language (XACML). IBM co-ordinated efforts towards the development of Enterprise Policy Authorization Language (EPAL). The evaluation of available privacy description languages is a critical step in determining if any available languages are suitable to use to represent privacy policies on web sites. If an available language is suitable for use, it can be used to generate graphical representations of privacy risks.

It is recognized that current privacy description languages are inadequate for use on the web. At the machine level, there is no widespread standard for communicating privacy information automatically to user agents. While multiple standards exist in relation to privacy and access controls, none of them have seen wide acceptance online (Cranor, Egelman, Sheng, McDonald, & Chowdhury, 2008). The most suitable standard, P3P, has been ignored by browser developers due to issues in the standard, while the most widespread standard, XACML, is used primarily in internal enterprise level applications (McCullagh, 2010). The development of EPAL has been abandoned (W3C, 2007a). Neither P3P or XACML are suitable for use in web applications.

### **2.3.1 Platform for Privacy Preferences**

P3P is a W3C standard that “enables Web sites to express their privacy practices in a standard format” (W3C, 2007c). P3P policies can be transmitted through an extension to a web server, or through the Application Programming Interface (API) of the language used to build a web application. In these methods, information is injected into additional Hypertext Transport Protocol (HTTP) headers. They may also be added as a tag within an XML Hypertext Markup Language (XHTML) document such as the following:

```
<link rel="P3Pv1" href="http://catalog.example.com/P3P/PolicyReferences.xml">
```

P3P is supported at a basic level in Microsoft’s Internet Explorer, but not in other web browsers. There are no current plans to implement P3P by other browser vendors due to a lack of industry adoption and the complexity of the standard (W3C, 2007).

The tools to create P3P policies are difficult to find and cumbersome. For example, IBM’s

P3P editor<sup>3</sup> has not been updated since 2000. There has been minimal adoption of P3P by popular online websites, with no significant growth between 2003 and 2007 (Beatty, Reay, Dick, & Miller, 2007). Since no web browser actively protects users by requiring the presence of P3P policies, or by evaluating them against user preferences, there is little incentive to add them to a website. Carnegie Melon University maintains the “Privacy Bird” P3P extension for Internet Explorer 5 and 6; however, it has not been updated in some years (*Privacy Bird*, 2009).

P3P has not seen wide acceptance, though it has seen some have observed a minimal level of growth in the government sector (Cranor et al., 2008). Many governments in the United States require machine-readable privacy policies to be posted on their websites (Cranor et al., 2008). As P3P is the only viable standard for representing such policies, it is required for United States governments to meet their legal obligations online (Cranor et al., 2008).

Some proponents of P3P argue that it may still become the prevalent method to describe privacy policies online (Cranor et al., 2008). P3P has been compared to the implementation of the Cascading Style Sheets (CSS) standard, which still has varying levels of adoption by web browsers (Cranor et al., 2008). The complexity of the P3P specification may be a significant barrier to adoption. It is important to note that while only Internet Explorer makes any attempt to implement P3P, most graphical web browsers attempt to implement the CSS specifications. Until browsers other than Internet Explorer attempt to implement P3P, it is unlikely that P3P will follow the trend set by the implementation of CSS. Neither the Gecko or WebKit HTML rendering engines support P3P or have any plans to support P3P. Given the multitude of other web standards which have been developed and implemented in the years since the finalization of the P3P standard, it is unlikely P3P will be

---

<sup>3</sup>Available at <http://alphaworks.ibm.com/tech/p3peditor>

included in any future web browser.

### **2.3.2 A P3P Preference Exchange Language**

APPEL is a W3C working draft (W3C, 2007b) that allows for privacy preferences to be described and compared against a P3P policy. APPEL is built upon (P3P), a ratified W3C standard.

APPEL complements P3P by describing a method to allow user agents to apply preferences to a P3P-enabled resource (W3C, 2007b). APPEL allows rules to consider the method of transmission (such as if Secure Sockets Layer (SSL) is enabled), as well as any set of arbitrary behaviours. The current 1.0 draft is quite limited, and only supports a small set of behaviours, meta data, and comparisons. Only simple website policies can be processed successfully.

No reference implementations of an APPEL parser have been released. Due to APPEL's reliance on P3P, it can not be effective without widespread deployment of P3P. No work has occurred on APPEL since 2002. Due to these issues, APPEL is unlikely to become viable as a modern privacy resolution language.

### **2.3.3 Enterprise Policy Authorization Language**

EPAL is a language for allowing an organization to express authorization policies against XML objects (W3C, 2007a). It was initially envisioned as a protocol complimentary to P3P. P3P would be used to allow an organization to disclose its privacy policies to external organizations. EPAL would be used to describe fine-grained policies for internal use. EPAL

allows for policies to be enforced and calculated, while P3P is for expressing the meaning and intent of policies.

EPAL builds on P3P by allowing policies to be attached directly to data. An EPAL policy is transmitted automatically with its data. The allows for increased granularity of preferences as compared to P3P. EPAL was submitted by IBM to the W3C. Upon analysis, it was determined that all of the EPAL features could be expressed within XACML. As well, there are many key features in XACML that cannot be expressed within EPAL (Anderson, 2006). The W3C has not continued the development of the EPAL standard.

#### **2.3.4 Extensible Access Control Markup Language**

XACML is a W3C standard for describing access policies of XML objects (Sun Microsystems, 2003). XACML can be seen as a superset of the functionality of EPAL. Instead of describing a method for communicating with a human being, XACML focuses on providing a framework for electronic systems to communicate privacy and access controls with each other. XACML provides the combined functionality of previous standards such as P3P and APPEL in a single standard. While XACML can be used to describe the specific controls applied to an object, it also describes the method of applying such controls to user preferences and conflict resolutions. A single XACML description can contain both the access controls (“This object can be accessed by managers but not by support staff”) and conflict resolution rules (“The manager role overrides any other roles on an object”).

One advantage of XACML over other standards is that the combining algorithm is a generic object, able to be replaced by other algorithms depending on the needs of the implemen-

tation. This allows a system to use its own set of rules to solve any inconsistencies in the description of access controls against an object.

XACML has some level of acceptance within Enterprise-level digital collections software, such as Fedora Commons (D. Davis, 2010). XACML is not currently used as a method to describe privacy policies to end users online due to its focus on machine-to-machine communication.

### **2.3.5 Summary of Privacy Description Languages**

Current privacy description languages have flaws preventing their adoption as privacy description languages for the web. P3P is not in widespread use and browser vendors are unwilling to implement it (McCullagh, 2010). XACML is best suited for complicated access controls applied to objects represented with XML, and not privacy policies. Without a suitable privacy description language, it is very difficult to implement software to automatically handle online privacy interactions.

Most languages with broad browser support are able to be edited with simple tools such as a text editor. As shown in Table 2.1, this includes common languages such as XHTML and CSS. Both P3P and XACML are too complicated to be edited with simple tools. In the case of P3P, this has likely limited adoption of the language. If a suitable privacy description language can be created, it will be possible to communicate privacy data to user agents, enabling the display of privacy risks using graphical indicators.

**Table 2.1:** Common Web Languages

Language	Licensing	Editor Required	Browser Support
XHTML	Open	Text editor	Broad support
CSS	Open	Text editor	Broad support
JavaScript	Open	Text editor	Broad support
VBScript	Proprietary <sup>a</sup>	Text editor	Some support <sup>b</sup>
Adobe Flash (Plugin)	Mixed (critical components proprietary)	Specialized development tools	Broad support <sup>c</sup>
Java (Plugin)	Mixed (minor components proprietary)	Text editor	Common support <sup>d</sup>

<sup>a</sup>Microsoft makes the documentation and language reference freely available for VBScript. A license is available for the source code for the VBScript engine, however the details of the license are not public (Microsoft, 2010). The patent status for VBScript is unclear and it is likely that VBScript is protected by Microsoft patents, preventing re-implementation by other vendors.

<sup>b</sup>VBScript is included with Windows and Internet Explorer, so it is available for most internet users. However, the lack of cross-platform implementations has hindered its use as a general scripting language on the web.

<sup>c</sup>The Adobe Flash plugin is included by default in Windows, OS X, and many devices such as the Nintendo Wii, making it available to web browsers on those platforms.

<sup>d</sup>The Java virtual machine is included with OS X and many portable devices such as cell phones. It is easily installed on Windows or Unix-like operating systems.



## 2.4 Privacy Interactions

A privacy interaction is defined in this thesis as an online interaction between a user and an online service that involves the personal information of the user. A privacy interaction could be as simple as using a search engine, or as complex as using a social networking website. Privacy policies, physical and virtual privacy threats, user interface design, gender, and the use of data submission on a given web page all affect privacy interactions. It is important to understand these components when building software to be used during a privacy interaction.

### 2.4.1 Privacy Policies

A privacy policy is a legally enforceable document between an organization and an individual describing how the organization will collect, maintain, and use private data collected from the individual. Enforcement processes for privacy policies vary by jurisdiction. In Canada, complaints can be filed with the Office of the Privacy Commissioner, who will produce a report on the matter (Department of Justice Canada, 2010a). The complainant then has the option to take the report to the Federal Court should the force of law be required to ensure compliance or produce remedies (Department of Justice Canada, 2010a). In the United States, the Federal Trade Commission (FTC) Act allows the FTC to sue companies that violate their privacy policies (Federal Trade Commission, 2008). As well, the Lanham Act allows companies to sue their competitors for unfair business practices when privacy policy violations occur (United States District Court, 2007).

Acceptance of a privacy policy is often required to interact with an organization. Privacy policies are commonly bound through legislation to public and private organizations and

are automatically applied when agencies collect personal data. For example, when using a website such as those operated by the Ontario government, the Freedom of Information and Protection of Privacy Act applies to all interactions.

Privacy policies are influenced and complicated by legislation such as the European Union's Data Protection Directive (DPD). The DPD is one of the most significant European Union directives regulating the use of private information. The Directive aims to regulate eight basic principles (Moshell, 2004). The European Convention on Human Rights provides for the right to privacy in "family life, home and correspondence" (Council of Europe, 2010). Unlike regulations in Canada and in the United States, the European Union DPD provides an over-arching set of regulations for all organizations, both private and public, within European states. When creating privacy policies affecting users in Europe, complying with the DPD can add significant complexity to a privacy policy.

Canadian law is in between the approaches of the European Union and the United States, with the Personal Information Protection and Electronic Documents Act (PIPEDA) protecting many, but not all interactions with organizations. Most notably, employees of provincially regulated organizations are excluded by PIPEDA (The Office of the Privacy Commissioner of Canada, 2009), and instead regulated by the relevant provincial legislation. This complicates the management of private information, as each province may have its own laws regarding the use of an employee's information within these organizations. As online organizations often operate over many different jurisdictions, it becomes very difficult to have a single privacy policy that applies equally to all individuals.

### 2.4.2 Gender and Online Privacy Attitudes

Gender correlates significantly with differences in both online attitudes and privacy attitudes. There has been significant work in testing the effect of gender upon online attitudes and privacy attitudes separately, but little work in testing gender differences in online privacy attitudes. Such effects could have serious implications for the development of software to integrate into an online privacy interaction.

A study in the sex differences in offline privacy preferences found significant differences between men and women. Differences were found in two of the six factors of privacy preferences as defined by Marshall and Pedersen (Marshall, 1974) (D. Pedersen, 1987). Women were likely to care more about privacy intimacy with family and friends than men while men were more likely to use isolation as a strategy for managing privacy (D. Pedersen, 1987). It is unknown if these attitudes extend to online privacy interactions.

Previous work has shown that men and women have different attitudes towards the use of computer technology. Most significant is the perception of skill, as those who perceive a lower skill are less likely to continue to pursue careers requiring that skill (Hargittai & Shafer, 2006). Hargittai and Shafer found that men and women tend to have substantially similar skills in online computer use. However, they also found that women are significantly more likely to self-report a lower level of skill than men (Hargittai & Shafer, 2006). This work indicates that previous studies that exclusively used self-reports of skill may not be valid for determining the actual skill of the study participants. Given the gender differences in technology attitudes, studies of privacy attitudes should consider any possible effects of gender.

Research into the gender differences in online bloggers found significant differences in the use

of blogs between men and women (S. Pedersen & Macafee, 2007). Blogs run by men tended to be information focused, while blogs run by women tended to contain more personal content and emphasized social connections (S. Pedersen & Macafee, 2007). Interestingly, women tended to have a greater preference for anonymity out of concerns for physical safety (S. Pedersen & Macafee, 2007). It is possible that gender differences in concerns for physical safety drive many differences in online privacy attitudes.

Overall, there is little work about the effect gender on online privacy attitudes. As gender differences may have a significant effect, it is important to consider them when conducting online privacy research.

### **2.4.3 Privacy Contexts**

The context of a privacy interaction is subject to many variables, such as the location of the individual, the time of day, or the website currently being used. This research focuses specifically on the current website context. Previous trust in an organization or category of website could greatly affect the behaviour of experiment participants. While many previous works test privacy behaviour within contexts such as e-commerce or social networking scenarios, it is common for researchers in the field to generalize their results to include all contexts that an individual may be subject to. In order for the work in this thesis to proceed, an evaluation of trust in various website categories must be conducted. For full details, please see Chapter 3.

#### **2.4.4 Summary of Privacy Interactions**

Privacy policies offer users the ability to accurately determine how their personal information will be handled by an organization. Unfortunately, privacy policies are too long, too complicated, and too difficult to read for most users to use. The differences in legal requirements for privacy policies complicates their implementation for online organizations that operate in many jurisdictions, further complicating understanding by website users.

Gender may contribute significantly to online privacy attitudes. If a system is to be devised to protect the privacy of online users, it will need to be aware of the effect of gender on online interactions.

The use of data collection as a critical component of privacy interactions indicates the importance of privacy communication methods to online users. As most websites present opportunities to submit data, privacy representations may be important and useful to online users.

### **2.5 Privacy Representation**

In order for users to be able to accurately enforce their privacy preferences, privacy policies need to be communicated to them. Most websites include a complete copy of their privacy policy. Some websites (such as aviary.com) (Aviary, 2009) have created two privacy policies; the first being the legal policy, and the second being an interpretation of each clause in plain English. This is effective for those curious about the policies, and makes it simpler to find information about the handling of specific pieces of data. Sound has been tested as a possible notification mechanism, but many participants disabled the sounds due to their annoyance

(Cranor, Guduru, & Arjula, 2006). Other work has shown that the use of graphics and pictograms is often the preferred method to communicate general ideas to a wide audience (Hemenway, 1982). Due to the prevalence of graphical icons in other aspects of computing (such as representing actions or objects (Gittins, 1986)), graphical representation of privacy policies might be an appropriate method to communicate privacy policies to users.

### **2.5.1 Components of Icons**

There are many components available to designers to represent concepts to users with icons. Work investigating the effect of colour, shape, and size in icon comprehension has found that all three categories have significant effects (Nowell, 1997). Colour was found to have the most significant effect, regardless of the type of data represented (Nowell, 1997). Shape and size can also have significant effects, but the specifics depend on the type of information being displayed (Nowell, 1997). Distinguishing messages presented with icons through the use of colour is a simple and effective method for any type of data, and may be critical to the design of graphical privacy representations.

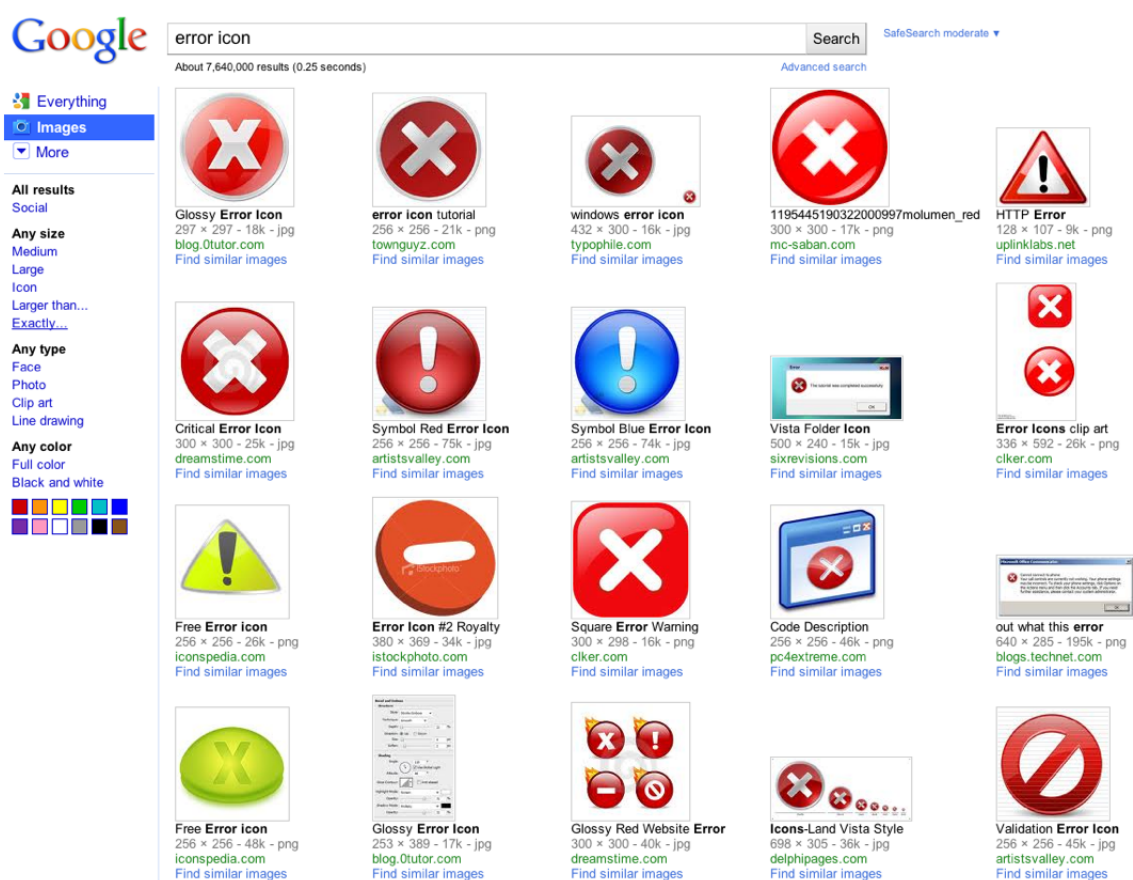
(McDougall & Reppa, 2008) states that complexity, familiarity, and concreteness are the most important factors. Shape, size, and colour may have less of an effect on users who are already familiar with icons for a given domain (McDougall & Reppa, 2008). Icon complexity is a significant factor in icon comprehension, as complex icons require more attention to interpret (McDougall & Reppa, 2008). McDougall and Reppa show that familiarity is the most significant component in user performance and aesthetic appeal. Regardless of other factors, users familiar with an icon are more likely to interpret it quickly and accurately (McDougall & Reppa, 2008). Because of this, it is best to follow previous conventions when

designing icons in an existing domain. For domains that users are unlikely to be familiar with, it is best to draw from existing icon sets where possible.

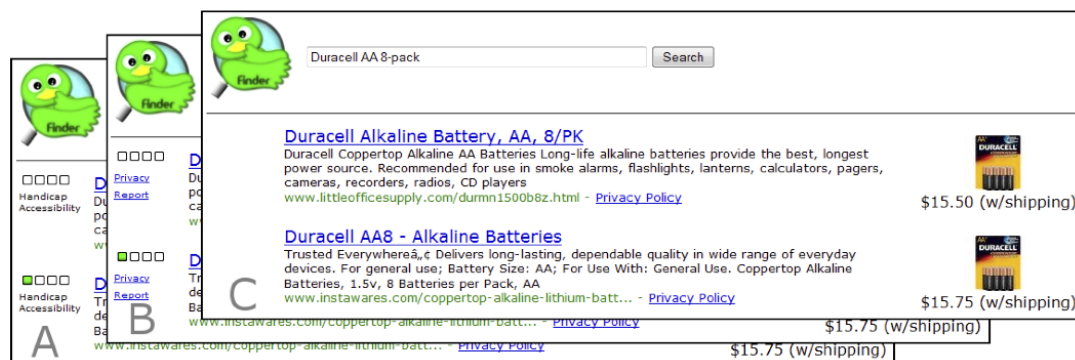
### **2.5.2 Icon Selection in Previous Work**

Communicating complex information to users of electronic systems is a common task in the information technology field. Since the advent of cheap, high resolution graphical displays in the 1980's, icons have become a core component of computing interfaces. Icons can represent a high density of information in a minimum of space, replacing text of many words (Hemenway, 1982). While icons are commonly used in Graphical User Interfaces (GUIs), determining what icons to use is a difficult process. Similar informational messages may have different or conflicting icons. A Google search for “error icon” reveals several conflicting designs for communicating an error message as shown in Figure 2.2. However, many icons do use red as a common colour for warnings, reflecting the importance of colour as noted by Nowell.

Previous privacy icons have used hands with thumbs-up and thumbs-down orientations, window shades, keyholes, and eyes (Cranor et al., 2006). An early version of Netscape Navigator used a chocolate chip cookie as its representation of P3P policies on HTTP cookies. Participants were often confused by these icons, as their relation to privacy was unclear (Cranor et al., 2006). The Privacy Bird P3P agent used a bird icon, in reference to canaries historically being used as warnings in coal mines (Byers, Cranor, Kormann, & McDaniel, 2005). As the bird symbol did not have a direct link to privacy, participants needed to either read the documentation or use the software for period of time to understand the meaning of the various states of the privacy bird (Cranor et al., 2006).



**Figure 2.2: Common error icons.** While there are some distinct themes in the design of error icons, it is apparent that some of those themes are contradictory with each other.



**Figure 2.3: Privacy progress bar.** The privacy progress bar display by Engelman et al. Note the use of both colour and shape to represent privacy risks.



When presented with graphical representations of privacy inline with content, users have shown to be likely to pay a premium for increased privacy (Egelman, Tsai, Cranor, & Acquisti, 2009). As shown in Figure 2.3, Egelman et al. used a graphical indicator inline with search results to show the privacy rating for a website. Similar to the work by Nowell, Egelman et al. also used colour as a primary graphical indicator. When purchasing sensitive items, individuals were more likely to purchase items at sites with a higher privacy rating, even with a higher price. It is noted by the authors that while the “progress bar” style indicator was more successful at communicating privacy than previous indicators such as the privacy bird, work needs to be completed to actually determine the effects of various styles and placements of privacy indicators.

### **2.5.3 Icon Testing Methodology**

Previous work testing the effectiveness of the Privacy Bird P3P agent did not test different icons for the agent itself, or attempt to integrate privacy information displays without the use of icons (Cranor et al., 2006). All evaluation of the icon was based on feedback from participants after being exposed to the icon, and not their behaviours caused by the presence of the icon (Cranor et al., 2006). The Privacy Bird icon consisted of three distinct components: the bird’s eyes, the colour of the bird (green, yellow, and red), and pictograms in a speech bubble emanating from the bird. As participants were exposed to all three elements, it was impossible to determine what element had the greatest effect in the understanding of privacy threats or in user’s behaviour.

Egelman et al.’s study of progress bar style indicators compared indicators labeled with “Privacy Report”, “Handicap Accessibility”, or no privacy information at all (Egelman et

al., 2009). Because they did not test for the effect of a text-only privacy link, it is impossible to determine from their results if it is the presence of a graphical indicator of any kind, or the presence of a graphically-enabled indicator that leads to changes in user's behaviour.

Most privacy icons that have been used have implemented both shape and colour as differentiating features (Cranor et al., 2006). During evaluation of the Privacy Bird P3P agent, the most recognizable notification feature of the software was the use of colour (Cranor et al., 2006). Previous work has not tested individual components of the icons, or the presence of icons themselves to determine what effect they may have on user behaviour. Some work has tested complicated icons with many different features, but found that they caused negative responses from participants due to the difficulty in comprehension (Reeder, Kelley, McDonald, & Cranor, 2008). If the most recognizable components (shape, colour, size, etc) of graphical privacy indicators can be isolated, the effect of each component can be tested.

#### **2.5.4 Summary of Privacy Representation**

Icons are a common method for communicating information to users. The use of icons stems from their effectiveness at displaying complex information within a small amount of space. Simple visual cues, such as colour and shape, can communicate detailed information (Nowell, 1997). Selecting appropriate icons to represent information is difficult, and icons in software can be confusing to users in practice (Cranor et al., 2006). Interfaces built for software in stable categories should use icons familiar to users (McDougall & Reppa, 2008). For example, an icon to go "back" in a web browser should always consist of an arrow pointing left, due to this convention in existing software. Where new icons are required,

colour should be the first icon component to develop, as colour has the most significant effect on comprehension (Nowell, 1997). Testing of icons should include testing of both icon components and a non-graphical representation. Otherwise, it is impossible to determine if the icon itself is beneficial.

## **2.6 Summary of Electronic Privacy**

Users have significant difficulties in managing privacy in online environments. The distinction between physical and virtual privacy is a significant point of confusion for many users. Users can be manipulated into disclosing data unintentionally to online organizations. It is unknown how often data is submitted as part of online interactions. The use of data submission techniques needs to be surveyed to ensure that data submission is a key component of the web, and not an infrequent action on the part of users.

The communication of privacy policies and access controls among computers is described by several standards. None of them are suitable for use with websites. User agents rarely implement privacy controls, and P3P, the only standard with any real-world use is only used in very restricted situations. The tools available to generate XML-based privacy policies are hard to find and difficult to use. Writing a machine-readable XML privacy policy by hand is very difficult, unlike other languages used on the web such as XHTML, CSS, and JavaScript.

Previous work has left many unresolved questions about communicating and applying privacy policies and privacy preferences online. Privacy policies are in common use, and are often required by law (Department of Justice Canada, 2010b) (Department of Justice Canada, 2010a) (European Parliament, 1995). Due to the varying uses of private data (col-

lection, storage, and analysis), privacy policies can be quite complex, as they need to be legally sound. Communicating privacy policies and ensuring end users understand them is very difficult, and often not in the interest of the organization operating the website as it discourages users from disclosing information (McDonald & Cranor, 2009).

Gender plays a significant role in both online attitudes and privacy attitudes. Privacy and safety skills from the physical world may play a role in online privacy attitudes in both men and women. The specific effects of gender when individuals complete online privacy interactions are unknown, but must be considered when studying online privacy interactions.

Research investigating privacy online has often been conducted within an ecommerce context, without any investigation into the validity of the context. Privacy behaviour can depend on a variety of factors such as website category or physical location, and it is important to ensure that the context of any experiment is controlled for. In order to accurately measure the behaviour of users online, the context of privacy interactions must be validated and replicated.

Work on graphical representations of privacy risks has not focused explicitly on graphical representations. Previous research has often assumed that graphical representations are the best method to communicate privacy, without any experimental verification to support that claim. For example, work completed in developing the Privacy Bird interface incorporated a bird into the icon without testing other icons, or a text-based interface (Cranor et al., 2006). As well, previous work has often confused graphical representation with the type of graphical representation used. Without testing to determine if graphics are an effective communication mechanism for privacy risks, it is difficult to isolate the effect of graphics in

general from the effect of the specific graphics chosen.

If a system of representing privacy risks is to be developed, these unresolved issues need to be addressed. The remainder of this thesis will detail investigations into machine-readable privacy descriptions, user perceptions of website trust, and the effect of graphical privacy representations. By investigating these areas, it will become possible to determine the effect of graphical privacy representations within trusted contexts.

## Chapter 3

# Experiment Design

### 3.1 Introduction

As shown in Chapter 2, users of online sites are often unaware of the impact of their data disclosure on their privacy. Users will willingly violate their privacy preferences to accomplish a goal with a trusted organization. Trust in organizations can vary. For example, as discussed later in this chapter, trust in social networking websites is quite variable, while most individuals trust financial institutions. In order to evaluate the effect of graphical privacy indicators on privacy interactions, it is important that the context of the website is controlled for.

The first part of this thesis will present data about user's perceptions of trusted and untrusted website categories. This data will allow for follow-up work to control for user trust in various website categories or organizations. Otherwise, if a website is built that is trusted by some participants and untrusted by others, it will be impossible to determine the effectiveness of graphical privacy indicators.

A second component will consist of an experiment that tests the presence of graphical representations of privacy risks within a selected trusted website design. This will allow the effect of graphical privacy indicators on participant's behaviour to be tested and evaluated. It is expected that the presence of graphical privacy indicators will lead to more privacy investigations and restricted data disclosure by participants.

## **3.2 Data Submission on the Web**

Previous work does not explore how the web is used as a dynamic communications medium, and instead treats it as a publishing medium. If privacy is to be adequately protected online, it is critical to understand how often and by what methods users submit data online. The following details an investigation into data submission techniques used online.

### **3.2.1 Selecting Representative Sites**

In order to determine how common data submission is online, a sample of all websites must be chosen to analyze for the potential to submit data. Rather than using a static list of websites based on traffic or manual ranking by a third party, the results from a search engine will be used. Using Google, for example, it is possible to list the top search queries within the past hour for a given geographic location. This query can be used immediately to generate a list of 10 websites that will be receiving a significant amount of traffic at that moment. This ensures that the websites are representative of what online users actually encounter by excluding web pages from the "long tail" of the web.

These websites can be analyzed for links to other pages, and those pages can be tested for the presence of data submission elements. The recursive depth can be configured based

on the number of sites and domains to be tested. No web page will be visited more than once.

### **3.2.2 Techniques of Data Submission**

To determine if a given web page contains the ability to submit data, each method of data submission must be considered. Within current XHTML standards, there are three primary methods of submitting data:

1. Submitting data within a form tag.
2. Submitting data through a POST request generated by JavaScript, commonly using the Asynchronous JavaScript and XML (AJAX) technique.
3. Submitting data from within a plugin such as Adobe Flash.

Other methods, such as encoding data in a GET request, are possible as well. However, such methods are not considered to be a proper use of the HTTP protocol, and are difficult to separate from simple queries on a remote server. As well, the use of plugins such as Flash require JavaScript to embed properly in web browsers. In order to simplify analysis, it was decided that any page with a form tag or script tag would be considered to have the ability to submit user data. Links created with JavaScript or embedded within a plugin are ignored in a similar manner to other search engines.

### **3.2.3 Defining Data Submission**

Data is used to describe any information transmitted from an individual to a website operator. This includes information not typically considered private (such as screen resolution,



search queries, or comment submissions) as well as private data (such as addresses or credit card information).

The crawler used in this survey is built to test for the opportunities to submit data, but not for data submission itself. While forms, for example, might be present on a web page, there is no way to determine if the submitted data is actually stored on a web server. JavaScript code would have to be parsed to determine if it submits data, and would still be dependent on the assumption that the receiving web server actually stores the transmitted data.

### **3.2.4 Crawler Results**

To determine the extent to which data submission is present on the web, a crawler was run to check for web pages containing the ability to submit data. Such web pages were considered to be submittable if they contained form or script tags, as JavaScript data can be submitted using AJAX or other methods. The crawler started with the top Google search query (“shannon price”) as of May 27, 2010 at 09:59 PM EST. It recursed to a depth of three pages from the Google search results page. This depth limit was chosen because it would complete testing within a 24 hour period. In total, 53133 pages were accessed across 2600 domains. The distributions of both form and the script tags were non-normal.

Out of these pages:

- 82% of the pages accessed contained forms (M 1).
- 92% of the pages accessed contained JavaScript code (M 16).
- 89% of the pages accessed contained both forms and JavaScript code.

- The largest number of form tags on a single page was a page containing a “report abuse” form for each comment on the page.
- The largest number of script tags on a single page was 625 different script tags. Upon inspection it appeared that the web page as was being generated incorrectly by the web server as the web page contained the same content repeated several hundred times.

In this time, an interesting phenomenon occurred with the Google search engine. For a period of time, the top search result for “facebook login” pointed to a blog post about Facebook’s “Facebook Connect” service (Johnson, 2010). During this period of time, hundreds of users commented on the post complaining about “Facebook’s” new look and how they couldn’t log in. Ironically, most of these users were logged in to Facebook and the site’s comment system using Facebook Connect, but were unaware of how to get to Facebook itself. What is important to note from this incident is that many users were unable to access Facebook without submitting data to Google in the form of a search term. For these users, a search engine is likely a required tool for using the web, implying that they submit data every time they use the web.

The vast majority of web sites contain the potential for data submission. Most pages contain explicit forms users can submit, or JavaScript code that can be used to collect data automatically from users. This result validated the assumption that most online interactions can involve the submission of data by website users.

### 3.3 Simplifying Machine-Level Communications

This thesis proposes a method of representing privacy risks using graphical indicators. Privacy risks will be determined from a XML description of the privacy policy of the trusted organization. Once calculated, privacy risks will be displayed inline with the related form fields using graphical representations, prompting participants to investigate possible privacy risks.

None of the previously discussed XML privacy description languages are suitable for use as a method of representing privacy risks online. Just as previous user interface work determined that communicating a subset of privacy policies was an appropriate method to communicate to users, communicating a subset of privacy policies at the machine level will help to encourage adoption (Cranor et al., 2006).

If a new privacy description language is going to be described, it is important to understand the qualities that have made other languages successful on the web. As shown in Table 2.1, most languages with broad browser support allow editing with simple tools and are free of licensing or patent requirements. Meeting these requirements will help to encourage adoption of a privacy description language.

The SPF as defined by this thesis provides a subset of the most critical components required to communicate a privacy policy. This followed previous work in the Privacy Bird interface that implemented a vocabulary subset of the P3P specification (Cranor et al., 2006). The critical components in both P3P and Privacy Bird are modeled after the core components of the European Union's DPD. The critical subset of these privacy principles selected for the SPF includes purpose limitation, transparency, and data transfer. Data security, data quality, sensitive data protection, independent oversight, and individual redress are not

included in the SPF as they go beyond the requirements of understanding and accepting a privacy policy. If a machine-readable format can address this subset of principles, it will maintain most of the functionality of P3P while reducing the effort required for developers to understand the format.

Exposing more principles to users online will increase the time required to understand a privacy policy without a significant benefit. Principles that are not covered can be explained in a link to the full, legally binding privacy policy. The SPF format addresses the critical privacy principles as described by the DPD, to ensure that the format is focused on the most common principles while maintaining its simplicity.

The SPF is an Resource Description Framework (RDF) format allowing the description of privacy policies in a machine readable format. The need for a new language arose out of the failure of other, more complex languages to be adopted within the website development community (Beatty et al., 2007). While P3P might seem to be suited for this task, the opaqueness of the language and the other issues mentioned previously limit the ability to successfully apply it broadly.

Each keyword within the name of the framework relates to an explicit goal in the design.

**Simple** A SPF description should be readable with a basic text editor. While specialized development tools may be used, they should not be required. An *average end-user* should be able to understand a privacy policy by reading an SPF description of the associated policy. It is important that the mechanisms used to transmit and display a privacy policy are transparent to the end user. When a computing system is seen as a black box, trust is lost (Patrick, 2001). By implementing a vocabulary subset of

critical components of a privacy policy, SPF allows organizations to communicate the most critical components of the policy, while giving users the ability to link to further resources for more complicated situations. This restriction ensures that the format is as simple as possible while still maintaining its utility.

**Privacy** The SPF schema is about one task: describing privacy policies. It is not about access control, executing decisions, or managing security.

**Framework** The SPF does not directly specify what fields should exist or how they should be named. Instead, it is meant to be an extensible standard drawing from RDF, able to fit any situation involving privacy policies. This ensures that each SPF file will only contain relevant information for the entity it describes. Each SPF file should use the appropriate RDF ontologies to describe organizations, locations, and other fixed information.

For the OWL schema and an example of a SPF file, see Appendix B.

SPF provides a method for communicating privacy policies at a machine level that addresses the shortcomings of previous standards. SPF is easily viewed and modified, focuses only on describing privacy policies and their core components, and is easily extensible by incorporating other SPF RDF ontologies. By implementing the SPF, it becomes possible for website operators to expose the critical components of their privacy policies to end users in a clear and concise fashion.

In order to adequately test the design of the SPF, several other significant components would have to be implemented, including a Natural Language Processing (NLP) algorithm to adequately parse existing privacy policies. Building such a system without testing the effectiveness of graphical representations could lead to the development of a system without

utility. Given these constraints, SPF is only described as a prototype OWL schema in this thesis, and it is left to future work to implement a prototype of the SPF.

### **3.4 Evaluating the Website Trust Component of Context**

Previous work has often focused on testing privacy icons without considering how participants might trust the websites they are interacting with (Cranor et al., 2006) (Egelman et al., 2009). It is possible for one user to feel that a given website is trustworthy, while another user might consider it to be untrustworthy. Unless this effect of website context is controlled for, it is possible that the variances in user trust will confound any results.

To determine what website categories participants trust, a survey will be used to obtain trust data from the potential subject pool. Participants will be asked to list five websites that they trust and five websites that they find untrustworthy. Each participant will be given the freedom to enter specific websites or any text they choose. All results will be categorized based on a dictionary of synonyms in order to determine what websites categories participants consider to be trusted or untrusted. Participants will also be asked to indicate what data elements they would be willing to disclose to the University of Guelph Student Financial Services, as it is suspected by the author that it might be a trusted organization. Such questions will only be asked after data about trusted and untrusted organizations in general is collected. While this thesis only focuses on data collected relating to website categories, the full questionnaire used can be found in Appendix A. The results of this experiment will inform the next process of the research, allowing a trusted website to be designed and evaluated.

## **3.5 Evaluating Graphical Privacy Indicators**

In order to evaluate the effectiveness of graphical privacy indicators, an experiment will be conducted to determine the effect of graphical indicators on the behaviour of online users. Participants will be presented with a trusted website, and will be asked to indicate what information they will be willing to submit to that website. By presenting some participants with graphical privacy indicators, and other participants with a text-only indicator similar to that used by Egelman et al., it will be possible to evaluate the effectiveness of the graphical indicator.

### **3.5.1 Participant Selection**

Participants will be recruited from the University of Guelph population. Participants will be required to be able to use a computer and a web browser without any sort of assistive devices. Recruitment will be executed with email and website advertising, as well as with direct advertising to student groups. Participants are categorized based on gender and educational degree. If the participant pool is heavily biased towards a specific discipline, degrees should be categorized as either belonging to the discipline or being external to the discipline. An equal number of participants in each category should be assigned to the testing and control groups, balancing the number of participants between the groups. Full details on the participants in the study can be found in Chapter 4.

### **3.5.2 Experiment Procedure**

Participants will first be situated at a computer and presented with an online survey collecting basic demographic information. This will include age, gender, and other basic in-

formation.

Next, participants will be directed to a website modeled after a University of Guelph registration web page. This organization is chosen based on the results of the website trust evaluation survey as detailed in Chapter 4. The registration page will indicate that services will be customized for them based on their responses.

The website will present a series of questions to the participant asking if they were willing to submit that information to facilitate service customization. Participants are not required to actually submit the information, but simply indicate that they would submit it later. This method helps to preserve the privacy of our participants. For the full list of questions presented, see Appendix A.

The registration website is built to present a single form asking if the participant would be willing to submit various questions as part of a registration process at a later date to the University of Guelph. The website uses the official University of Guelph “Common Look and Feel” template. Each data item will consist of four columns as shown in Figure 3.1.






1. A checkbox indicating that the participant was willing to submit the information.
2. The question itself, such as “Your name?”
3. A longer description of the question with further details about what exactly the question would ask for.
4. A link leading to the privacy risks associated with each question. Participants in the testing group will be presented with a graphical privacy indicator and a link. Participants in the control group will be presented with the link only.





### Account pre-registration

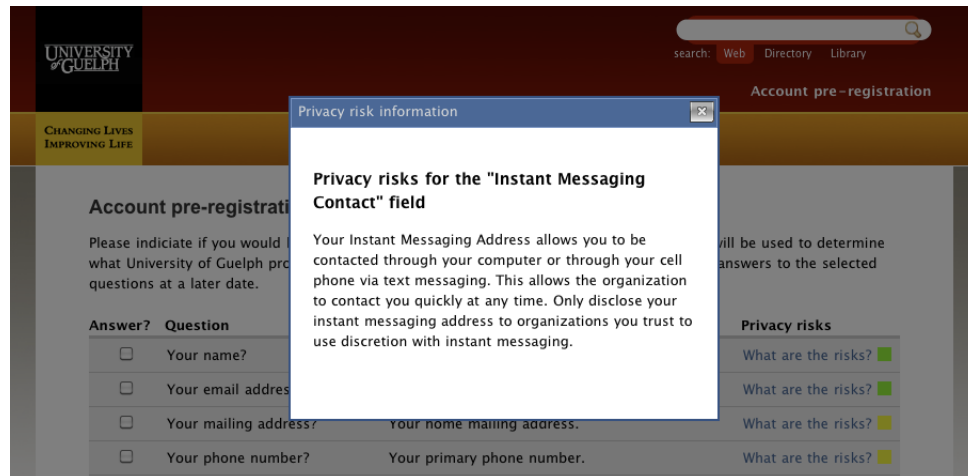
Please indicate if you would be willing to answer the following questions. This information will be used to determine what University of Guelph programs you may benefit from. You will be asked to provide the answers to the selected questions at a later date.

Answer?	Question	Question details	Privacy risks
<input type="checkbox"/>	Your name?	Your first and last names.	What are the risks? 
<input type="checkbox"/>	Your email address?	Your @uoguelph.ca email address.	What are the risks? 
<input type="checkbox"/>	Your mailing address?	Your home mailing address.	What are the risks? 
<input type="checkbox"/>	Your phone number?	Your primary phone number.	What are the risks? 
<input type="checkbox"/>	Your IM account?	Your primary instant messaging (MSN, Google Talk, Skype, etc) account.	What are the risks? 

**Figure 3.1: Account registration screen.** The account registration form shown to experiment participants.

Data requested will include basic contact information, health and demographical information, and opinions about University services. Each participant will be presented with the Safari web browser with the “Private Browsing” feature enabled. This ensures that no user data is stored within the browser between participant sessions.

The testing group will be presented with a form with graphical indicators of the privacy risks inserted beside each form item. Each representation will consist of a small green, yellow, or red square, along with the sentence “What are the risks?”. Each colour indicates a progressively higher privacy risk. Each indicator will be displayed to the right of the privacy investigation link, and will be a part of the link itself allowing the indicator to be activated to display privacy information. In both the testing and the control groups, the privacy investigation text will be displayed as “What are the risks?” in a table in the last column “Privacy risks”. This will ensure that the privacy indicators are easily visible and displayed consistently.



**Figure 3.2: Instant messaging privacy risks.** Displaying the privacy risks for disclosing an Instant Messaging address.

The graphic and the link will be combined to form a single link, which will display an inline popup or “lightbox” upon activation as shown in Figure 3.2. The lightbox will contain information about the privacy risks involved with disclosing the requested information.

The text of each inline popup will be created to represent what could be created automatically by a user agent implementing the SPF specification. No links to opt-out instructions or further details will be included in the privacy information text. Text describing similar risks will be formulated to follow the same language to simulate how an agent might present the risks. The privacy risks popup can be dismissed by clicking a “close” button.

The control group will be presented with a similar form removing the graphical indicators and only having a text link visible. Based on the previous work mentioned in Chapter 2, only colours, and not more complicated icons, will be tested to prevent possible confusion over the meaning of more detailed icons.

Data will be collected through the use of recorded screen captures, webcam video of participants, and automated JavaScript collection code. Metrics able to be captured will include

any clicks or keystrokes made by the user and the questions participants are willing to submit. Each privacy investigation will also be recorded. This data allows for the number of privacy investigations to be accurately and automatically compared against other recorded data.

After successful submission of the registration form, participants will be presented with an exit survey to capture their perceptions of the registration form. These answers, along with the captured data from the registration form, will provide the metrics for analysis of the hypothesis.

The primary metric for testing the effectiveness of graphical privacy indicators will be the number of privacy investigations. A privacy investigation is defined as one click to display the privacy risks for a given field. The number of submitted questions will also be a contributing metric, allowing it to be determined if the presence of graphical indicators causes participants to reduce the amount of data they were willing to disclose. Gender, degree, and educational background will all be tested to determine if there is a correlation between them and specific privacy interactions. The analysis of these factors will allow the effect of graphical indicators to be quantified. If a significant effect on privacy investigations is caused by the presence of graphical indicators, it will be possible to prove that graphical privacy indicators are a useful method of encouraging users to investigate privacy risks.

### **3.6 Experiment Design Summary**

When testing the effectiveness of using graphical privacy indicators, it is important to ensure that the website used for testing has a common level of trust among participants. If there is a significant variance in trust between participants, then their behaviour might vary due

to trust, and not due to the presence of graphical indicators. By gathering trusted and untrusted websites and website categories from participants, a website can be built that elicits a trustworthy response from participants.

An experiment testing the effect of graphical privacy indicators used on a trusted website will be conducted. A trusted website was chosen due to the ease of implementation, though selecting an untrusted website would also be valid. By presenting some participants with graphical indicators, and other participants with text-only indicators, it will be possible to evaluate the effectiveness of the graphical indicator. When presented with graphical indicators of privacy risks, it is expected that participants will be more likely to investigate the possible risks to their privacy. Differences in gender, age, and educational background will also be tested against the effectiveness of the graphical privacy indicator. If the use of a graphical indicator is shown to have an effect, then such indicators can be integrated into software to help protect user's privacy online.

## Chapter 4

# Experimental Results

This chapter will present the results from all experimental data gathered during the course of this thesis. Website trust will be evaluated, and trusted and untrusted categories will be determined. As well, graphical privacy indicators will be tested and evaluated against a range of factors to determine if they are an effective mechanism for communicating privacy risks.

### 4.1 Evaluating Website Trust

Participants consisted of 15 computer science students with varying levels of post-secondary education. All participants were between the ages of 18 and 25. Twelve men and 3 women participated in the survey. The survey was conducted mostly in-person with a portable laptop, though some participants participated remotely.

In order to determine trusted and untrusted website categories, participants were asked to list up to five trusted and untrusted websites. While the question asked participants

to mention websites, the opportunity was given to list general categories instead. Upon analysis, each specific website was placed into a category. All category titles in Table 4.1 were mentioned by participants, except for the “Search Engine” category, where participants always mentioned a specific company name such as Google or Yahoo!. Table 4.1 contains all of the categories and how many participants identified them as trusted or untrusted. The entire survey can be found in Appendix A.

**Table 4.1:** Trusted and Untrusted Website Categories

<b>Organization Category</b>	<b>Number Trusted</b>	<b>Number Untrusted</b>
Banks	15	0
Universities	9	0
E-commerce companies	9	3
Governments	8	3
Search engines	5	8
Social networks	3	12
Employment-related organizations	2	0
Individual preferences	0	14
Forums and blogs	0	5
Entertainment focused sites	0	3
File sharing sites	0	2
Obvious scams	0	2
Pornography sites	0	1

Banking organizations were the most likely to be trusted, followed by University and E-commerce websites. For example, specific responses treated as belonging to the “Banking” category included:

- A bank that I had an account with for a while
- Bank
- Banks
- Banks (the Royal Bank of Canada)
- CIBC

- Mastercard
- My bank
- PayPal
- President’s Choice Financial
- Scotia Bank
- TD Bank
- TD Canada Trust
- WebBanking

Due to the conclusions reached in Chapter 5, the remainder of this research used a University website as a trusted organization to test the effect of graphical privacy indicators, as it was determined that such a website would be relevant to participants and not subject to brand influence. This allowed the researchers to evaluate what privacy disclosures participants would make to a trusted organization while still violating their privacy preferences.

## 4.2 Testing Graphical Privacy Indicators

For the purpose of this analysis, a “privacy investigation” was defined as every instance of a user investigating the privacy risks of a requested item by clicking the “What are the risks?” link. Multiple clicks of the same link by a participant were summed to give the total number of privacy investigations.

A submitted question was recorded when participants submitted the registration form indicating that they would submit the associated data at a later date. Participants could

modify their selections as they saw fit and each question’s state was only recorded once the entire form was submitted.

Individual academic backgrounds were recorded as “Computer Science” or “Not Computer Science” as it was possible that there might be a significant effect related to academic background. For example, at the University of Guelph, all Computer Science students are required to take a “Social Implications of Computing” course, which might influence perceptions of privacy.

#### **4.2.1 Results Summary**

The experiment sessions were completed by a total of 29 participants. All participants were between the ages of 18 to 34. Other grouping factors include:

- 14 individuals were assigned to the testing group, while 15 were assigned to the control group.
- 62% of the participants were male, while 38% were female.
- 72% of the participants were Computer Science majors, while 28% were of a different academic background.

The time participants spent investigating privacy risks was very low, and most participants didn’t investigate privacy risks at all. Other results of note include:

- 62% of participants indicated that they either agreed or strongly agreed that they could easily determine the privacy risks of a given question. This indicated that participants were aware of the privacy risk indicators present on the page.



- The average number of privacy investigations across all participants was 0.66 investigations. This indicates that many participants never bothered to investigate privacy risks at all.
- The average participant completed the registration form in 2 minutes 17 seconds.
- The quickest registration submission was 41 seconds, while the longest submission was 19 minutes and 53 seconds.
- On average, each participant committed to answering 67% of the registration questions at a later date. This indicates that participants were willing to disclose significant information to the University.

The information participants offered to disclose was broad and detailed, and often deviated from what was actually required to successfully complete an online registration. In contrast to the information gathered during the website trust survey, account registration participants often disclosed more information. For example, 59% of participants would disclose how often they consumed illicit drugs, even though that information would not typically be required to register an account or customize services.

- 20% of website trust survey participants would disclose monthly vehicle loan payments, while 38% opted to submit that information during account registration.
- When asked about disclosing favourite restaurants on campus, only 47% of website trust survey participants indicated they would disclose the information (by selecting “Agree” or “Strongly Agree”), but 90% of account registration participants indicated they would submit that information.
- 69% of participants offered to disclose their sexual orientation.

- 59% of participants offered to disclose how often they consumed an illicit substance.
- 17% of participants offered to disclose every item asked of them.
- 60% of website trust survey participants would not disclose their Instant Messaging identities (selecting “Disagree” or “Strongly Disagree”), and 76% of account registration participants chose to protect that information. This was one of the few items that most participants across both studies chose to protect.

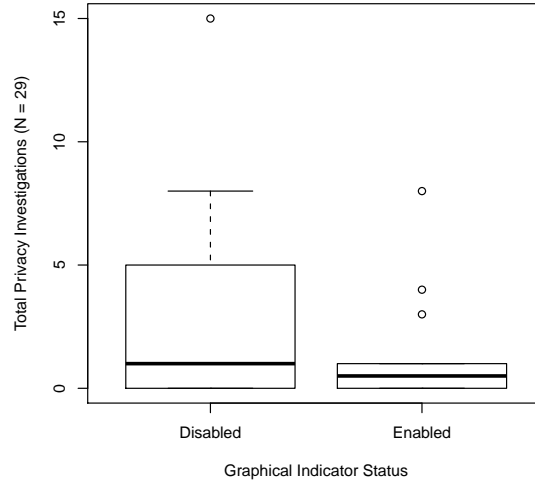
Most participants did not investigate privacy risks, and most participants were willing to disclose a significant amount of data.

Most participants (79%) felt confident in being able to ascertain the privacy risks for a given question, while 62% felt that there were enough details presented to allow for an informed decision.

#### **4.2.2 Analysis Procedure**

The results of this analysis allowed for the effect of graphical privacy icons to be tested against the number of privacy investigations and the amount of data each participant was willing to submit. Gender, degree, and age were tested to determine if they had a significant effect on how participants reacted to graphical privacy indicators.

The number of privacy investigations, the total questions chosen to be submitted, and participant’s age were all tested for normality using the Shapiro-Wilk test of normality (Field, 2005). Expertise self-rating was also tested with the Shapiro-Wilk test after converting the skill ratings to integers. Non-parametric statistics were used to match the collected data.



**Figure 4.1: Graphical Privacy Indicator Status by Total Privacy Investigations.** The presence of graphical privacy indicators did not have an effect on privacy investigations (Graphical indicators disabled  $M = 0.5$ , Graphical indicators enabled  $M = 1$ ,  $ns$ ).

Graphical icon status, gender, and degree were all tested against the number of privacy investigations using the Wilcoxon Rank Sum test. Due to the presence of ties in the data, exact significances could not be calculated. The same procedures were used to compare each grouping variable against the total number of questions willing to be submitted.

For tests involving multiple groups, the Kruskal-Wallis test was applied. This included testing any effect involving age or the number of privacy investigations. In the cases where the result was significant, the Wilcoxon Rank Sum test was applied post-hoc to the data.

### 4.2.3 Privacy Investigations Data Analysis

The number of privacy investigations were not significantly affected by the presence of graphical privacy icons as shown in Figure 4.1 ( $U = 121$ ,  $r = 0.02$ ,  $ns$ ).

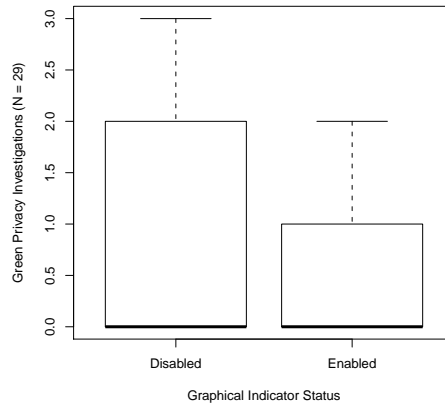


Figure 4.2: Graphical Indicator Status by Green Privacy Investigations. ( $M = 0$ ,  $ns$ ).

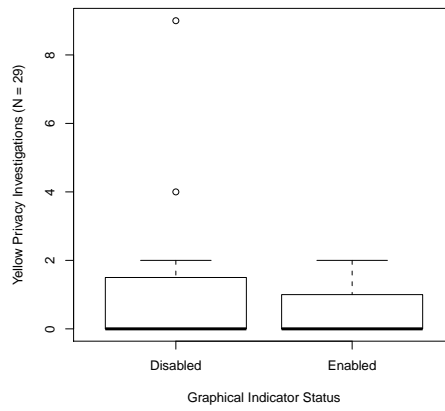


Figure 4.3: Graphical Indicator Status by Yellow Privacy Investigations. ( $M = 0$ ,  $ns$ ).

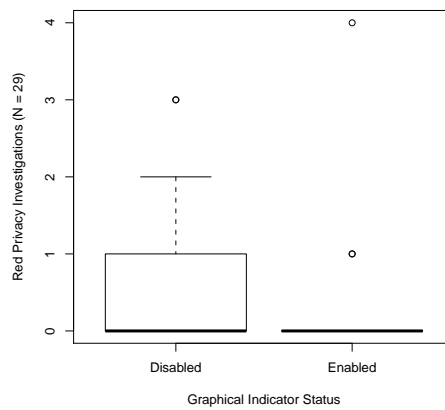


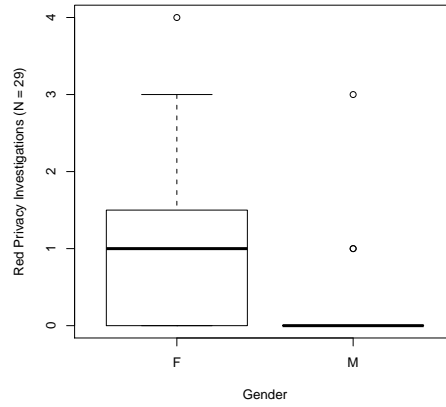
Figure 4.4: Graphical Indicator Status by Red Privacy Investigations. ( $M = 0$ ,  $ns$ ).



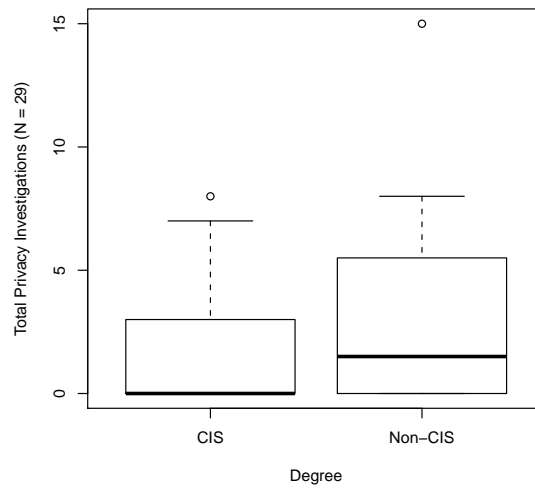
Female participants were significantly more likely to investigate privacy risks ( $M = 3$ ) than male participants as shown in Figure 4.5 ( $M = 0$ ), ( $U = 151, r = 0.41, p < 0.05$ ).

60

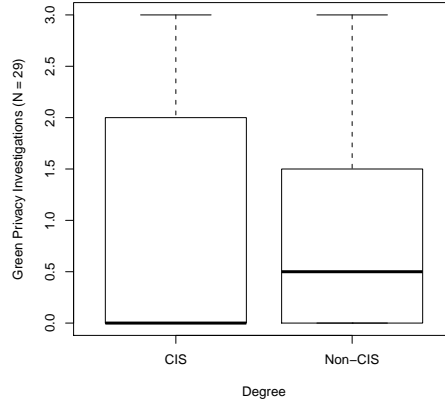




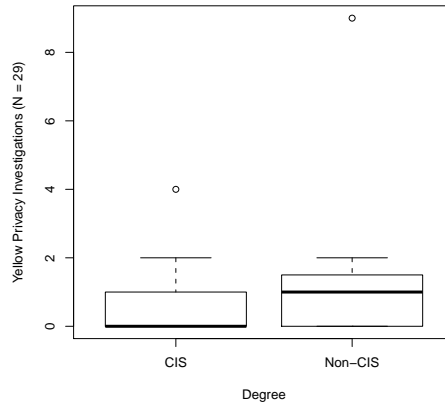
**Figure 4.8: Gender by Red Privacy Investigations.** Questions labeled red received more investigations by female participants (Female red investigations  $M = 1$ , Male red investigations  $M = 0$ ,  $p < 0.05$ ).



**Figure 4.9: Degree by Total Privacy Investigations.** The degree of a participant did not affect the number of privacy investigations (CIS investigations  $M = 0$ , Non-CIS investigations  $M = 1.5$ ,  $ns$ ).



**Figure 4.10: Degree by Green Privacy Investigations.** (CIS green investigations  $M = 0$ , Non-CIS green investigations  $M = 0.5$ , *ns*).

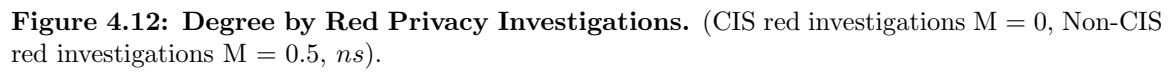


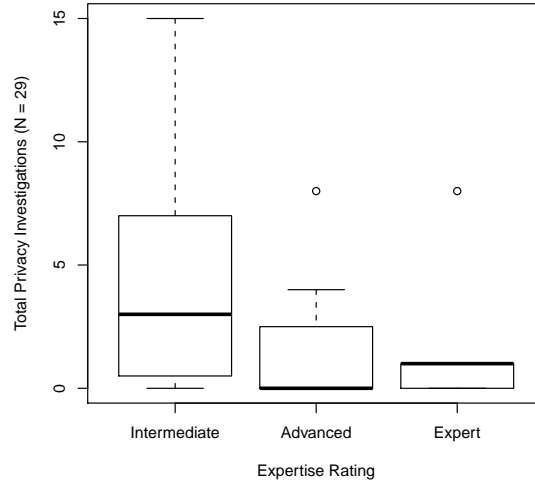
**Figure 4.11: Degree by Yellow Privacy Investigations.** (CIS yellow investigations  $M = 0$ , Non-CIS yellow investigations  $M = 1$ , *ns*).

As indicated by Figure 4.9, the number of privacy investigations was not significantly affected by the educational background of the participant ( $U = 66, r = 0.06, ns$ ).

There was no significance when privacy investigations by degree were decomposed into each colour associated with the investigation as shown in Figure 4.10 ( $U = 74, r = 0.04, ns$ ), Figure 4.11 ( $U = 59, r = 0.17, ns$ ), and Figure 4.12 ( $U = 72, r = 0.0, ns$ ).







**Figure 4.14: Internet Expertise Rating by Total Privacy Investigations.** (Intermediate investigations  $M = 3$ , Advanced investigations  $M = 0$ , Expert investigations  $M = 1$ , *ns*).

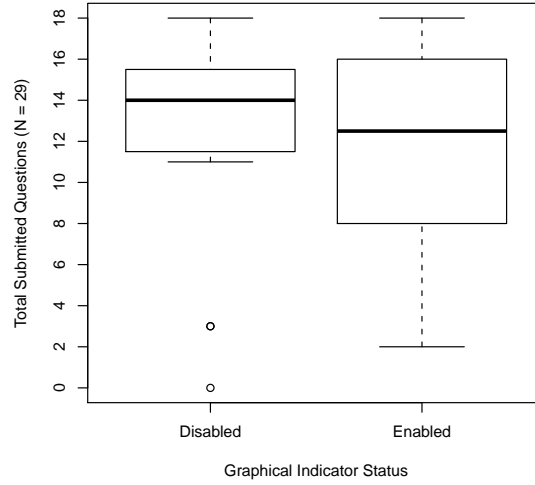
The number of privacy investigations was not significantly affected by the age of the user ( $H(7) = 5.9956, ns$ ).

The number of privacy investigations was not significantly affected by the internet self-rating of the user as shown in Figure 4.14 ( $H(2) = 2.5676, ns$ ).

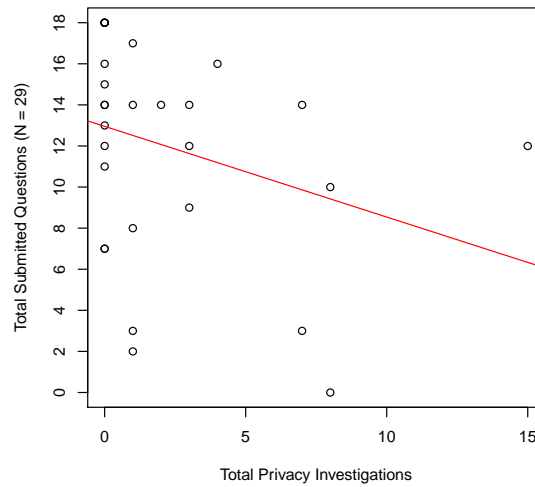
#### 4.2.4 Submitted Answers Data Analysis

The number of questions a participant was willing to submit was not significantly affected by the presence of graphical privacy icons ( $U = 116.5, r = 0.06, ns$ ) as shown in Figure 4.15.

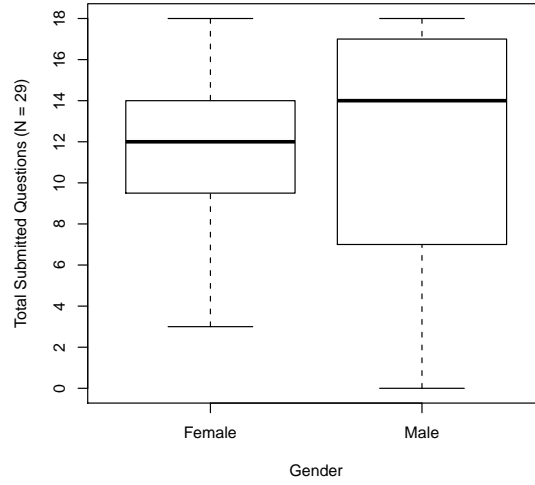
As shown in Figure 4.16, the number of questions submitted by each participant was not significantly affected by the number of privacy investigations ( $H(14) = 17.0713, ns$ ).



**Figure 4.15: Graphical Indicator Status by Total Submitted Questions.** Graphical icons did not cause participants to restrict the number of questions they would be willing to submit (Graphical indicator disabled  $M = 14$ , graphical indicator enabled  $M = 12.5$ ,  $ns$ ).



**Figure 4.16: Total Privacy Investigations by Total Submitted Questions.** Privacy investigations did not have a significant effect on the number of questions submitted by a participant as shown by a linear regression ( $dx/dy = -0.4415$ ,  $ns$ ).

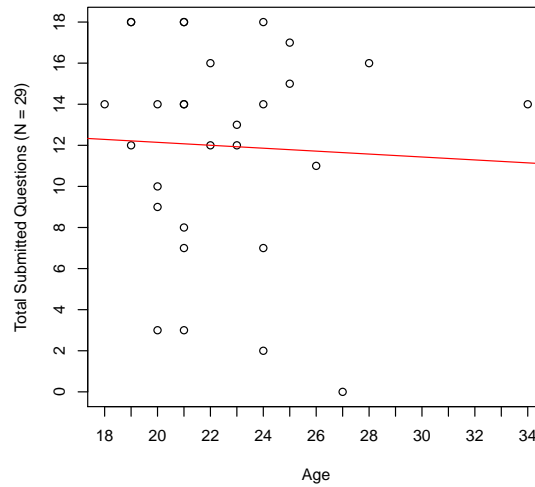


**Figure 4.17: Gender by Total Submitted Questions.** (Female submitted questions  $M = 12$ , Male submitted questions  $M = 14$ , *ns*).

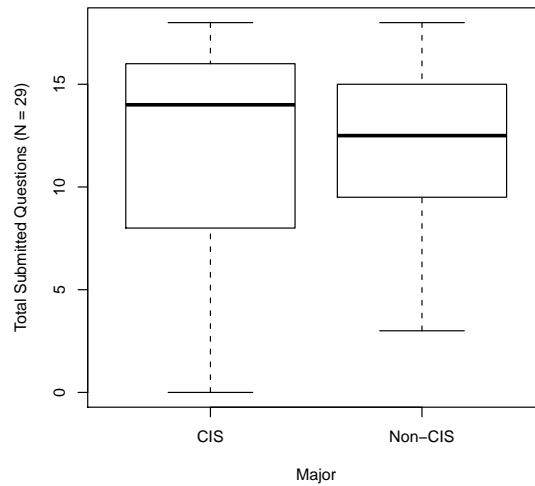
The number of questions willing to be submitted by each participant was not significantly associated with the gender of the user ( $U = 77, r = 0.08, ns$ ).

As shown in Figure 4.18, the number of questions willing to be submitted by each participant was not significantly affected by the age of the user ( $H(14) = 13.635, ns$ ).

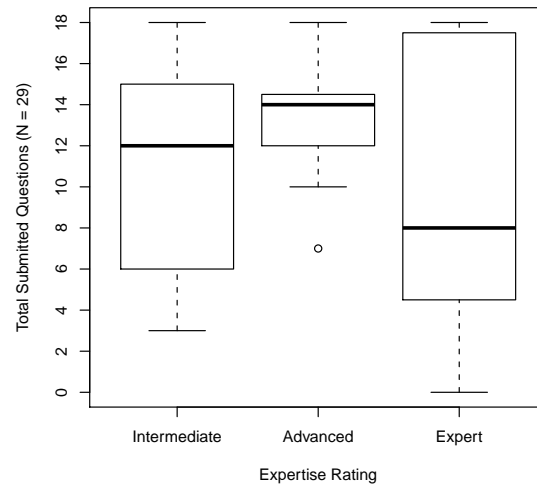
As shown in Figure 4.19, Computer Science students did not offer to disclose more or less information as compared to students with other educational backgrounds ( $U = 90.5, r = 0.14, ns$ ).



**Figure 4.18: Age by Total Submitted Questions.** Age did not have a significant effect on the number of questions willing to be submitted by each participant as shown by a linear regression ( $dx/dy = -0.07132, ns$ ).



**Figure 4.19: Major by Total Submitted Questions.** (CIS submitted questions  $M = 14$ , Non-CIS submitted questions  $M = 12.5, ns$ ).



**Figure 4.20: Expertise Rating by Total Submitted Questions.** (Intermediate submitted questions  $M = 12$ , Advanced submitted questions  $M = 14$ , Expert submitted questions  $M = 7.5$ ,  $ns$ ).

Figure 4.20 shows that the number of questions willing to be submitted by each participant was not significantly associated with the internet expertise of the participant ( $H(2) = 0.9795, ns$ ).

## Chapter 5

# Conclusions and Future Work

### 5.1 Conclusions

#### 5.1.1 Conclusions in Evaluating Website Trust

The results detailed in Chapter 4 indicate that participants had a high degree of trust for websites that offer highly secure functionality, such as managing financial accounts, registering for courses, and online purchasing.

Individual preferences, social networks, and search engines were the least trusted categories of online organizations. Banks, employment organizations, and universities were the only trusted categories not submitted by participants as being untrustworthy. For all other trustworthy categories, at least one participant indicated that the category was untrustworthy. This demonstrates that when testing privacy interfaces within existing websites, banking, university, and e-commerce website websites should be used to test privacy preferences

within trustworthy situations. Social networking and search engine websites should be used to test privacy preferences within untrustworthy situations.

### **5.1.2 Conclusions in Privacy Investigations**

As detailed in Chapter 4, the amount of data disclosed by each participant and the confidence in their ability to protect their privacy indicates that participants were aware of the tools available to check for privacy risks, but chose not to use the tools or limit their data disclosure.

This finding leads to many questions about the effectiveness of graphical privacy icons. It was expected that the presence of icons would increase the number of privacy investigations. Ignoring the significance of the result, it is clear from Figure 4.1 that if any effect exists, it is that the number of privacy investigations decreases with the presence of graphical icons.

Several factors could cause this decrease in privacy investigations. It is possible that graphical icons are actually a method of instilling a superficial level of trust in participants, just as the “padlock” icon in a web page causes users to assume the page is secure. If this is the case, then graphical icons could be used to trick participants into submitting information beyond their privacy preferences.

It is also possible that the use of a website with the official University of Guelph “Common Look and Feel” design instilled a sense of trust in participants beyond the expected level. The University design was selected as it was determined to be an organization that most participants from the sample population would have interacted with and disclosed information to in the past based on the results from the website trust survey. However, it is unlikely



that participants had ever disclosed some of the private information (such as illicit drug use) to the University. As all participants would have similar experiences with the University, any effect from the presence of graphical icons should have been significant.

Figures 4.2, 4.3, and 4.4 indicate that colour or risk may not have a significant effect on privacy investigations, which would contradict the hypothesis that an increase in the displayed risk would lead to more privacy investigations. It is also possible that increased risk leads to participants reducing their privacy investigations.

Figure 4.5 indicates an increased awareness of the social risks of disclosing private information online by women. This could be caused by personal experience, cultural biases as to how women should treat privacy, greater experience with online systems gathering their personal information, or a greater value on private information.

Figures 4.6, 4.7, and 4.8 indicate that participants were most interested in the range of risks associated with the graphical icons, as it follows that yellow risks would be roughly between green and red risks. The trust in the icons themselves is an important factor, as it becomes clear once participants are aware of a privacy icon, they are likely to trust its judgement, regardless of the trustworthiness of the agent generating the icon.

This result shown in Figure 4.9 was unexpected, as it was hypothesized that participants with computer science backgrounds would be more likely to understand and be aware of the privacy risks presented by the use of an online information database accessible by many individuals and corporations. As all participants had some level of post-secondary education, it is possible that most participants were aware of the risks, and didn't feel the need to investigate further. Or, due to the minimal difference in age, it is possible that most participants, regardless of educational background, have similar perceptions of privacy and

privacy risks.

It was expected that an increase in age would positively correlate with the number of privacy investigations, however results as shown in Figure 4.13 indicate that the opposite might be true. Due to the small range of ages in the sample population, this hypothesis was unable to be confirmed. Future work should make a concerted effort to include a broader generational population, and determine if there is any relationship between age and privacy investigations.

This result in figure 4.14 was unexpected, as it was expected that increased online experience would lead to greater privacy concerns. As no true novices were included in the sample population, future work should assess a wider sample to determine the differences in privacy perceptions between novices and experts.

The use of graphical icons within a trusted context did not have a significant effect on participants. Participants exposed to graphical icons did not execute more privacy investigations, and participants chose to ignore the ability to investigate privacy risks. Female participants were more likely to investigate privacy risks when presented with graphical representations of privacy risks. However, the presence of such icons did not cause a significant reduction in data disclosure.

### **5.1.3 Conclusions in Submitted Answers**

Figure 4.15 indicates that prominent use of graphical icons did not cause participants to become more careful about disclosing their private information. It was expected that the presence of icons would cause participants to become more aware of their privacy prefer-

ences, and thereby become better equipped to ensure that their actions online matched their privacy preferences.

Figure 4.16 indicates that even if participants had a greater knowledge of the privacy risks of a specific item, they did not change their behaviour.

Figure 4.17 is interesting when considering the significant correlation between gender and privacy investigations. While women were more likely to investigate their privacy risks, this investigation did not lead to a restriction on the information they were willing to disclose. While it is unknown if the disclosures matched the privacy preferences of each user, it is possible to conclude that female participants had increased awareness of privacy risks.

The result shown in Figure 4.18 is likely to be caused by the similar ages of the participants. It is hypothesized that participants of different generational backgrounds would have a significant difference in their responses.

Most participants chose to submit a significant amount of data when interacting within the trusted context. The presence of graphical privacy icons did not have any effect on the amount of data participants indicated they would be willing to submit. Within a trusted website, most participants appear willing to disclose much of what is asked of them even when it has little relevance to their task.

It was expected that very few participants would be willing to submit private health related information, or their sexual orientation. The number of participants who did indicate that they would submit such data changes how researchers should expect participants to behave when asked to disclose private information within a trusted website context. The speed at which participants completed the registration form indicates that most participants

have little interest in investigating privacy risks, even when personal information is being disclosed.

## **5.2 Future Work**

The results of this thesis suggest many potential future research projects. There are several unanswered questions in the fields of automated privacy agents, graphical privacy representation, data submission by internet users, and internet expertise. Future work may combine efforts from all of these fields towards solving current issues in representing and managing privacy online.

### **5.2.1 Privacy Agents**

A privacy agent is a software component that analyzes interactions involving privacy and automatically suggests or decides on actions for the user to take. The Privacy Bird software is an example of a privacy agent (Cranor et al., 2006). A privacy agent may be an appropriate method to automatically ensure that individual privacy preferences are protected online.

#### **An Automated Privacy Agent**

Enforcing privacy preferences in electronic environments is difficult for many users. An automated agent integrated with a browser would allow for a user's privacy preferences to be automatically enforced. The use of such an agent would ensure that users don't accidentally disclose information in violation of their preferences.

One advantage of the use of an agent separate from the website or application in use is that privacy preferences can be described in isolation from an interaction with an organization. This will ensure that the design or interface used to collect private information is prevented from eliciting information from the user in violation of their preferences. If a user decides to violate their stated preferences during an interaction, they can be made aware of the inconsistency and resolve the conflict as they see fit.

An automated agent would also have the ability to watch a user's behaviour and automatically modify or generate privacy preferences based on actual, and not stated behaviour. A system of this type would be able to automatically determine when a user's behaviour changes drastically, indicating a deviation from their privacy preferences. The analysis of deviation from average enforcement behaviour might be a suitable replacement for explicitly defining privacy preferences.

Finally, an automated agent would allow for configuration of various behaviours depending on the type and scope of a privacy preference violation. Some users might prefer for an agent to make subtle suggestions about privacy enforcement, while other users might prefer a complete block of a preference-violating interaction. The use of an agent allows the communication of privacy risks and potential violations to be replaceable, leading to the potential for many different methods of privacy risk communication.

### **Automatic Generation of Simple Privacy Framework Policies**

The ability to automatically generate a SPF description from a privacy policy would simplify the creation of machine-readable privacy policies. Most privacy policies are crafted by lawyers, and contain very specific language. It should be possible to automatically parse a

legal privacy policy and generate a SPF compatible file for review. Such a system would likely have greater accuracy than a system designed to parse free-form text, as it could be tuned to handle legal terms and to be stricter in its interpretation of the text.

Once it is possible to automatically generate SPF policies, such a system could be embedded into a user agent. This would allow for the user agent to be active even when a website doesn't explicitly reference an SPF policy.

### **Segregating by Privacy Perceptions**

The suitability of a privacy agent's representation of privacy risks may be significantly affected by previous privacy attitudes held by an individual. It is possible that those who have been victims of privacy attacks may find different utility in a privacy agent. It would be useful to determine the effect of pre-screening for privacy attitudes within privacy experiments.

### **Effects of Trusted and Untrusted Contexts**

The work presented in this thesis focused on testing participant behaviour in a specific trusted context, namely the University of Guelph. The results indicate that participants may disclose significant information unrelated to the task at hand. Future research should replicate the experiment within an untrusted context, such as a social networking website. It would be useful to determine if participants change their behaviour when presented with graphical privacy representations within untrusted contexts, and instead focus on previous successful interactions when dealing with trusted contexts.

### **5.2.2 Refining Privacy Risk Graphical Representation**

Simple graphical icons as presented in this thesis may not be an effective method of communicating privacy risks to users. Alternate representations of privacy risks may prove to have a greater effect on user behaviour. Determining when to display an icon may be a critical component of privacy comprehension. As well, icon positioning may be a key factor in icon effectiveness.

#### **Alternative Graphical Representations**

The use of simple, inline, coloured graphical representations of privacy risks was employed to allow for investigation into the effect of the graphical icons themselves. While the results indicate that there was no effect caused by the presence of graphical indicators, further work should investigate the difference between a simple inline icon, and a modal dialog displaying the same icon. This would allow for the effect of icon placement to be tested. If no effect is found when participants are forced to view and interact with a graphical representation of a privacy risk, then it can be concluded that there is no effect on users for any kind of graphical representation.

Another method of graphical representation would involve disabling risky form elements and replacing their contents with a graphic and text describing the privacy risks. Only after reviewing accepting the risks would the form element be enabled. Such a display would force users to evaluate privacy risks, while potentially causing frustration and avoidance of the privacy preference system entirely.

A final method of representation would be to adopt a policy of blocking access to web pages

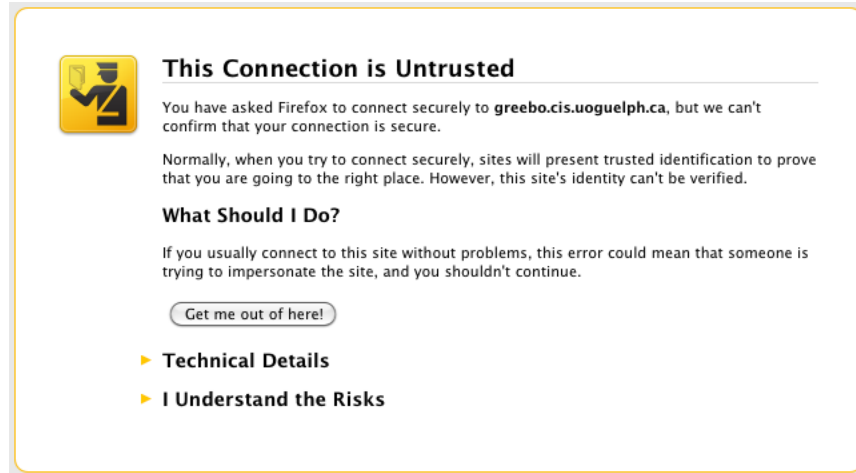


Figure 5.1: Firefox 3 SSL certificate error.

entirely if their privacy policies violate a user's preferences. Many modern web browsers have adopted an interface similar to Figure 5.1 for SSL certificate errors. A privacy agent could display a similar interface, and require users to approve the violation before allowing for any interaction with the site.

### Following the F Pattern

One failing of the registration website built for this thesis was that privacy risks and graphical privacy indicators were displayed on the right hand side of the page. If privacy indicators are to be considered an important element, they should be displayed on the left side of the page (Nielsen, 2006). As well, the size of the icon should be increased to be bigger than the text of the page to provide a bigger target to activate. A re-implementation of the registration website with these techniques in mind may significantly change how users investigate privacy risks.



### **5.2.3 Data Submission on the Web**

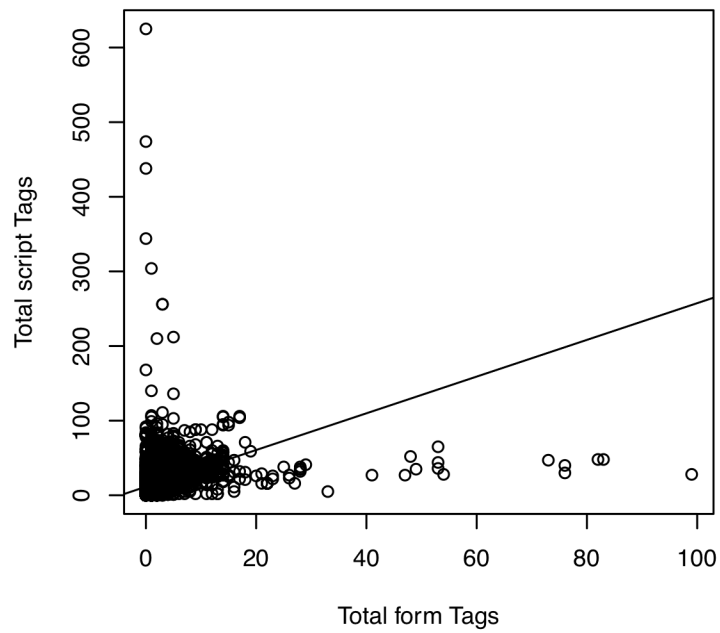
#### **Submission of Unrelated Data**

With 89% of the web allowing users to submit data, the question for researchers is no longer if users will submit data in their day-to-day use of the web. Instead, it is determining what fields are considered private enough to give serious concern to online users. It is expected that users will avoid websites which blatantly attempt to gather private information for unrelated purposes. If Google, for example, asked for a current bank account balance to search the web, it's likely that users would quickly switch to an alternative search provider. In other situations, where the expectation of privacy is not so clear, privacy preferences and their actual application by users can vary greatly.

One interesting result is the correlation between form and script tags, as shown in Figure 5.2. The use of both explicit data collection through forms and implicit data collection through JavaScript programs might significantly change as AJAX and other techniques become central to the development of web applications. For example, JavaScript can be used to submit partially completed forms, without the user's permission. This type of dynamic application can greatly increase the privacy risks of using a given website.

#### **Testing Data Submission Behaviours**

While a significant majority of the web implements technologies allowing for data to be transmitted from users to website operators, it is unknown how much data is actually submitted during an average online interaction. For explicit submission through forms, users are often given the opportunity to leave fields blank. Research to determine how



**Figure 5.2: Total form Tags by Total script Tags.** Using Kendall's Tau, a significant correlation is found between the number of form tags and script tags on a given web page ( $\tau = 0.53, p < 0.0001$ ). This indicates that pages that collect information are likely to use both static methods such as form submission and dynamic methods using JavaScript to collect data.

often users leave such fields blank would be very useful. As well, users may not consider certain actions, such as searching a site, as submitting data. It would be useful to know if online behaviour changes based on how aware a user is of data submission. Finally, data can be submitted transparently through the use of JavaScript and plugins. It is trivial to capture all keystrokes made within a browser viewport (and was in fact done with the registration website). Website operators can capture such data without the user being aware of the submission. It would be useful to determine how many websites use such subtle data collection methods.

## **Web Sites and Web Applications**

The evolution of web sites into web applications requires the understanding of how novices interpret web pages to be reviewed. It is possible that users now treat web pages as components of applications, skipping the interpretation of “web site” entirely. If users view web sites primarily as applications used to query, manipulate, and modify data, then their interpretations of user interface elements may be significantly different as compared to the past decade.

### **5.2.4 Expertise Evaluations**

Data relating to internet expertise was collected from participants in both the website trust survey and the website registration experiment. The analysis of this data points to several possibilities for future work in the field of expertise analysis.

Expertise data was determined to not be normally distributed and was tested with the Kruskal-Wallis test. The Wilcoxon Rank Sum test could then be applied as all expertise

data consisted of two nominal grouping variables. Results include the test statistic  $U$ , the effect size  $r$ , and the level of significance.

All participants agreed or strongly agreed with the statement that they were experts at using websites online. While participants could rate themselves on a five point scale, as all participants fell within two groups for the purpose of analysis it is assumed that users could only choose between “advanced” and “expert” ratings, corresponding to the “agree” and “strongly agree” assessments.

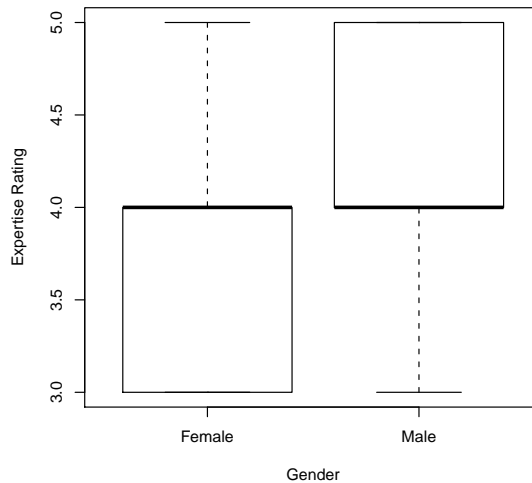
Internet expertise self-ratings showed significant differences in gender, educational background, and internet experience. The relationship between gender and educational background is especially interesting, as it would be useful to understand what factor has the most significant effect on expertise self-ratings.

### **Expertise Analysis**

Expertise and experience are often use to describe similar attributes in experiment participants. Expertise is a learned skill, while experience can only be acquired with time. It is possible to have significant experience without expertise, and vice versa. The self-reported data presented does not distinguish between experience and expertise. Future research should separate each attribute were possible.

Participant expertise in using the internet was evaluated during the demographic survey for all participants. Several trends emerged in relation to internet expertise:

- All participants ranked themselves as at least an “Intermediate” skill in using computers and the internet.

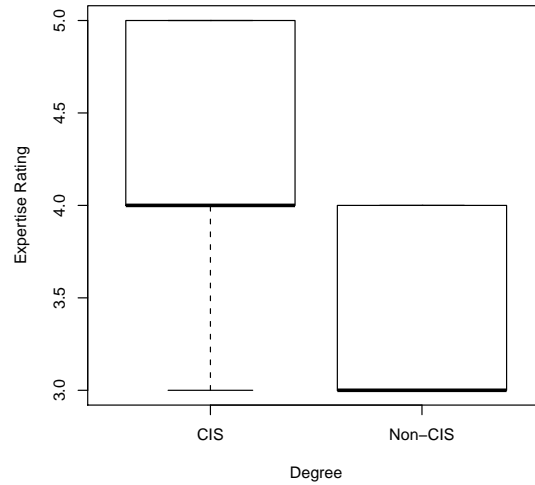


**Figure 5.3: Gender by Expertise Rating.**

- 93% of the participants had helped another person solve a problem with their computer.
- Older users had less online experience than younger users.
- There is still a correlation among younger users linking age to when they were first online; however this correlation is rapidly disappearing as internet use matures.

Male participants were significantly more likely to have a higher expertise self-rating (Mn = 4.22) than female participants as detailed in Figure 5.3 (Mn = 3.64), ( $U = 55, r = 0.34, p < 0.05$ ). When decomposed by degree, there was no significant difference in the expertise rating. This confirms previous results, and future work relying on self-ratings should determine a method to correct for this bias when inferring actual expertise from self reports.

Computer science students and graduates were significantly more likely to have a higher expertise self-rating (Mn = 4.24) than students from other disciplines as shown in Figure

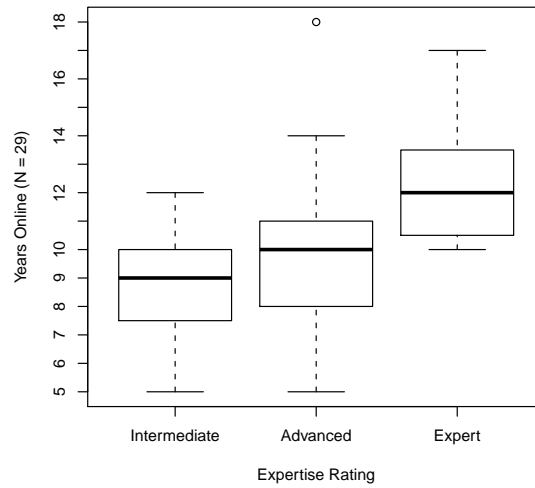


**Figure 5.4: Degree by Expertise Rating.**

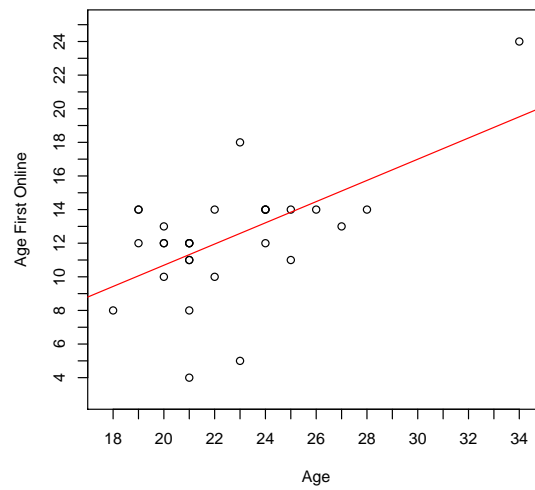
5.4 ( $Mn = 3.38$ ), ( $U = 139$ ,  $r = 0.50$ ,  $p < 0.05$ ). When decomposed by gender, there was no significant difference in the expertise rating. This finding indicates that both gender and educational background have a significant effect on the self perception of expertise. It is possible that a wider study of different generational groups may broaden this effect.

Internet expertise was significantly correlated with the number of years the participant had been online ( $H(2) = 6.2122$ ,  $p < 0.05$ ), with a longer period online corresponding with a greater likelihood of a user self-rating as an expert ( $U = 5.5$ ,  $r = 0.39$ ,  $p < 0.05$ ). Figure 5.5 indicates that internet experience is strongly correlated with internet use. Future research should determine the exact relationship between internet use and experience.

The age a at which a user first regularly used the internet was significantly affected by the age of the user ( $H(11) = 15.02$ ,  $p < 0.05$ ). Figure 5.6 confirms the status of the internet as a young medium, otherwise users would likely first be online at a similar age. It would be useful to determine if categories of web applications, such as webmail, blogging, or social



**Figure 5.5: Expertise Rating by Years Online.**



**Figure 5.6: Age by Age First Online.** For a mature medium it would be expected that a linear regression would have a slope of zero.

networking follows a similar trend.

Gender differences in internet expertise self-ratings should be compared against gender differences in self-ratings for other mediums. Comparing these results to self-ratings for other technologies, such as cell phones and digital cameras, would be useful to help determine if such differences can be applied to higher technology in general.

### **Summary of Expertise Evaluations**

During the analysis of internet expertise, a significant difference was detected between males and females, and computer science students and those from another discipline. The sample tested contained very few males from outside of computer science, or females from within computer science. This causes significant difficulty in determining if it is gender, or educational background that has the most significant affect on internet expertise self-ratings.

Future work investigating internet expertise should involve a wider range of individuals, especially those from other disciplines. This will allow for gender and degree to be isolated from each other, enabling researchers to determine which variable has the greatest effect on expertise.

Internet expertise may have an effect on the use of privacy enhancing user agents. An appropriate choice of graphical representation may be dependent on the expertise of the specific user. For example, an internet novice may be unable to distinguish between the sources of different icons on screen, and might have difficulty in ascertaining the trustworthiness of individual representations. The confusion between the viewport and the browser chrome exhibited by novices might be especially significant in terms of the placement of icons. If a



reliable test of expertise evaluation can be created, it will be possible to apply such results to not just privacy-related user interfaces, but all user interfaces in general.

### **5.2.5 User comments and Observations**

One participant described themselves as a “hacker”. This could mean that the participant saw themselves as someone who either likes to play with electronic systems, or to try and break into secure systems. While it would seem that such a user would want to investigate every facet of a system, the user didn’t investigate their privacy risks once. It is a common assumption that self-described hackers will thoroughly investigate electronic systems. It would be interesting to determine if such users are willing to investigate privacy-aware systems.

One participant commented that she had trouble with computers, but that the internet was easy to understand. However, this participant also indicated that she had helped others solve problems with their computers, so it is likely that her own perception of skills are less than what they really are. Further research investigating user’s perceptions of the boundaries between computer use and network use would allow for more accurate classification of a user’s expertise.

A participant associated advanced computer use with “games” and “downloading”. This indicates that the user associated searching for and installing new software with advanced knowledge. The user also noted that they felt comfortable with using computers for research, which could also be considered an advanced skill. Perception of what tasks are advanced would allow for greater precision in expertise self-ratings.

### 5.3 Conclusion

The development of methods to communicate privacy policies, threats, and preferences between entities contains many unsolved problems. Privacy-protecting agents are a promising technique for ensuring that individuals can accurately apply their privacy preferences online. Further work researching methods of communicating privacy threats to users will be critical to the acceptance of any privacy-protecting software agent by users. The differences in gender and educational background may have significant effects on the acceptance of graphical representations of privacy risks, and should be investigated. Expertise should be explored to determine if it has a significant effect on the interpretation of graphical representations. While this work was unable to show a significant effect of graphical privacy indicators, it did show a significant relationship exists between gender and awareness of the risks of disclosing private information online. This relationship, and its impact on privacy representation, will provide the basis for future research topics.

# References

- Akyildiz, I., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer networks*, 38(4), 393–422.
- Anderson, A. H. (2006). A comparison of two privacy policy languages: Epal and xacml. In *Sws '06: Proceedings of the 3rd acm workshop on secure web services* (p. 53-60). New York, NY, USA: ACM. Available from <http://doi.acm.org/10.1145/1180367.1180378>
- Anton, A., Bertino, E., Li, N., & Yu, T. (2007). A roadmap for comprehensive online privacy policy management. *Communications of the ACM*, 50(7), 116.
- Aviary. (2009, December). *Aviary - terms*. Retrieved December 30 2009, from <http://aviary.com/terms>
- Beatty, P., Reay, I., Dick, S., & Miller, J. (2007). P3p adoption on e-commerce web sites: a survey and analysis. *IEEE Internet Computing*, 65–71.
- Berendt, B., Günther, O., & Spiekermann, S. (2005). Privacy in e-commerce: stated preferences vs. actual behavior. *Communications of the ACM*, 48(4), 101-106. Available from <http://doi.acm.org/10.1145/1053291.1053295>
- Burke, R. R. (2002, September). Technology and the customer interface: What consumers want in the physical and virtual store. *Journal of the Academy of Marketing*

- Science*, 30(4), 411–432. Available from <http://www.springerlink.com/content/c13308q6h14k7u83/>
- Byers, S., Cranor, L. F., Kormann, D., & McDaniel, P. (2005). Searching for privacy: Design and implementation of a P3P-Enabled search engine. In *Privacy enhancing technologies* (Vol. 3424, pp. 314–328). Springer Berlin / Heidelberg. Available from <http://www.springerlink.com/content/jh4g491b17tqr26v/>
- Chellappa, R. K., & Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumer’s dilemma. *Information Technology and Management*, 181-202.
- Christofides, E., Muise, A., & Desmarais, S. (2009). Information disclosure and control on facebook: Are they two sides of the same coin or two different processes? *Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society*.
- Council of Europe. (2010, 02). *The european convention on human rights*. Retrieved February 5 2010, from <http://www.hri.org/docs/ECHR50.html>
- Cranor, L. F., Egelman, S., Sheng, S., McDonald, A. M., & Chowdhury, A. (2008). P3p deployment on websites. *Electronic Commerce Research and Applications*, 7(3), 274 - 293. Available from <http://www.sciencedirect.com/science/article/B6X4K-4SBHX4S-1/2/9dead86bf77383f97f28f97c1dab6530> (Special Section: New Research from the 2006 International Conference on Electronic Commerce - Evaluating new technological innovations for successful business on the internet, Eighth International Conference on Electronic Commerce (ICEC))
- Cranor, L. F., Guduru, P., & Arjula, M. (2006). User interfaces for privacy agents. *ACM Trans. Comput.-Hum. Interact.*, 13(2), 135–178. Avail-

able from [http://portal.acm.org/ft\\_gateway.cfm?id=1165735&type=pdf&coll=GUIDE&d1=GUIDE&CFID=75374007&CFTOKEN=73414487](http://portal.acm.org/ft_gateway.cfm?id=1165735&type=pdf&coll=GUIDE&d1=GUIDE&CFID=75374007&CFTOKEN=73414487)

Davis, C. N. (2005). Reconciling privacy and access interests in e-government. *International Journal of Public Administration*, 28(7), 567-580.

Davis, D. (2010, March). *Fedora authorization with xacml policy enforcement*. Retrieved November 29 2010, from <https://wiki.duraspace.org/display/FR22D0C/Fedora+Authorization+with+XACML+Policy+Enforcement>

Department of Justice Canada. (2010a, July). *Personal Information Protection and Electronic Documents Act*.

Department of Justice Canada. (2010b, July). *Privacy act*.

Dhamija, R., Tygar, J. D., & Hearst, M. (2006). Why phishing works. In *Chi '06: Proceedings of the sigchi conference on human factors in computing systems* (pp. 581-590). New York, NY, USA: ACM.

Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. In *Soups '06: Proceedings of the second symposium on usable privacy and security* (pp. 79-90). New York, NY, USA: ACM.

Egelman, S., Tsai, J., Cranor, L. F., & Acquisti, A. (2009). Timing is everything?: the effects of timing and placement of online privacy indicators. In *Chi '09: Proceedings of the 27th international conference on human factors in computing systems* (pp. 319-328). New York, NY, USA: ACM. Available from [http://portal.acm.org/ft\\_gateway.cfm?id=1518752&type=pdf&coll=GUIDE&d1=ACM&CFID=77737543&CFTOKEN=83139556](http://portal.acm.org/ft_gateway.cfm?id=1518752&type=pdf&coll=GUIDE&d1=ACM&CFID=77737543&CFTOKEN=83139556)

European Parliament. (1995, November). *Data protection directive*. OPOCE.

Federal Trade Commission. (2008, June). *Legal resources - statutes relating to both missions*.

- Retrieved August 6 2010, from <http://www.ftc.gov/ogc/stat1.shtm>
- Field, A. (2005). *Discovering statistics using spss* (2nd ed.). Sage Publications.
- Gittins, D. (1986). Icon-based human-computer interaction. *International Journal of Man-Machine Studies*, 24(6), 519 - 543. Available from <http://www.sciencedirect.com/science/article/B6WGS-4T4SG0S-3/2/ca68a90d820315f73c46c0a81b86cb8d>
- Hargittai, E., & Shafer, S. (2006). Differences in actual and perceived online skills: The role of gender. *Social Science Quarterly (Blackwell Publishing Limited)*, 87(2), 432 - 448. Available from <http://search.ebscohost.com.cerberus.lib.uoguelph.ca.subzero.lib.uoguelph.ca/login.aspx?direct=true&db=bth&AN=20754131&site=ehost-live&scope=site>
- Hemenway, K. (1982). Psychological issues in the use of icons in command menus. In *Proceedings of the 1982 conference on human factors in computing systems* (pp. 20–23).
- Hoffman, D., Novak, T., & Peralta, M. (1999). Building consumer trust online. *Communications of the ACM*, 42(4), 80–85.
- Jarmo, K. K., & Parkkinen, J. (2001). Signs of trust: A semiotic study of trust formation in the web. In *in the web. in.*
- Johnson, B. (2010, February). *How to confuse a facebook user*. Retrieved November 29 2010, from <http://www.guardian.co.uk/technology/blog/2010/feb/11/facebook-readwriteweb>
- Kapadia, A., Henderson, T., Fielding, J. J., & Kotz, D. (2007). Virtual walls: Protecting digital privacy in pervasive environments. In *Pervasive computing* (Vol. 4480, pp. 162–179). Springer Berlin / Heidelberg. Available from <http://www.springerlink.com/content/a651245g33k62p72/>

- Little, K. (1965). Personal space. *Journal of Experimental Social Psychology*, 1, 237–247.
- Marshall, N. (1974). Dimensions of privacy preferences. *Multivariate Behavioral Research*, 9(3), 255–271.
- McCullagh, D. (2010, February). *Mozilla weighs privacy warnings for web pages*. Retrieved November 29 2010, from [http://news.cnet.com/8301-13578\\_3-10445642-38.html](http://news.cnet.com/8301-13578_3-10445642-38.html)
- McDonald, A., & Cranor, L. (2009). The cost of reading privacy policies. *ISJLP*, 4, 543–597.
- McDougall, S., & Reppa, I. (2008). Why do I like it The relationships between icon characteristics, user performance and aesthetic appeal. In *Human factors and ergonomics society annual meeting proceedings* (Vol. 52, pp. 1257–1261).
- Merriam Webster. (2010). *privacy*. Retrieved 28 June 2010, from <http://www.merriam-webster.com/dictionary/privacy>
- Microsoft. (2010, August). *What is vbscript?* Retrieved August 8 2010, from <http://msdn.microsoft.com/en-us/library/1kw29xwf.aspx>
- Milne, G., & Culnan, M. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3), 15–29.
- Morozov, E. (2009). Iran: Downside to the “twitter revolution”. *Dissent*, 56(4), 10-14.
- Moshell, R. (2004). And then there was one: The outlook for a self-regulatory united states amidst a global trend toward comprehensive data protection. *Tex. Tech L. Rev.*, 37, 357.
- Nielsen, J. (2006, April). *F-shaped pattern for reading web content*. Retrieved November 29 2010, from [http://www.useit.com/alertbox/reading\\_pattern.html](http://www.useit.com/alertbox/reading_pattern.html)

- Nowell, L. (1997). *Graphical encoding for information visualization: using icon color, shape, and size to convey nominal and quantitative data*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University.
- Patil, S., & Kobsa, A. (2005). Uncovering privacy attitudes and practices in instant messaging. In *Group '05: Proceedings of the 2005 international acm siggroup conference on supporting group work* (p. 109-112). New York, NY, USA: ACM. Available from <http://doi.acm.org/10.1145/1099203.1099220>
- Patrick, A. (2001). Privacy, trust, agents and users: A review of human-factors issues associated with building trustworthy software agents. *Retrieved January, 18, 2006*.
- Pedersen, D. (1987). Sex differences in privacy preferences. *Perceptual and Motor Skills*, 64, 1239–1242.
- Pedersen, S., & Macafee, C. (2007). Gender differences in british blogging. *Journal of Computer-Mediated Communication*, 12(4), 1472 - 1492.
- Petronio, S. (1994). Privacy binds in family interactions: The case of parental privacy invasion. *The dark side of interpersonal communication*, 241–257.
- Privacy bird*. (2009, December). Retrieved December 30, 2009, from <http://www.privacybird.org/>
- Public Service. (2010, January). *Nhs has 33,000 health records stolen*. Retrieved August 5 2010, from [http://www.publicservice.co.uk/news\\_story.asp?id=11949](http://www.publicservice.co.uk/news_story.asp?id=11949)
- Reeder, R. W., Kelley, P. G., McDonald, A. M., & Cranor, L. F. (2008). A user study of the expandable grid applied to p3p privacy policy visualization. In *Wpes '08: Proceedings of the 7th acm workshop on privacy in the electronic society* (pp. 45–54). New York, NY, USA: ACM. Available from [http://portal.acm.org/ft\\_gateway.cfm?id=1456413&type=pdf&coll=GUIDE&dl=GUIDE&CFID=77733450&CFTOKEN=19030636](http://portal.acm.org/ft_gateway.cfm?id=1456413&type=pdf&coll=GUIDE&dl=GUIDE&CFID=77733450&CFTOKEN=19030636)



- Resnick, M. L., & Montania, R. (2003). Perceptions of customer service, information privacy, and product quality from semiotic design features in an online web store. *International Journal of Human-Computer Interaction*, 16(2), 211.
- Sun Microsystems. (2003). *A brief introduction to XACML*. Retrieved December 3 2007, from [http://www.oasis-open.org/committees/download.php/2713/Brief\\_Introduction\\_to\\_XACML.html](http://www.oasis-open.org/committees/download.php/2713/Brief_Introduction_to_XACML.html)
- The Office of the Privacy Commissioner of Canada. (2009, March). *Privacy legislation in canada*. Retrieved November 29 2010, from [http://www.priv.gc.ca/fs-fi/02.05\\_d\\_15\\_e.cfm](http://www.priv.gc.ca/fs-fi/02.05_d_15_e.cfm)
- United States Department of Justice. (1974). *The privacy act of 1974*. Retrieved March 18, 2010, from <http://www.justice.gov/archive/oip/privstat.htm>
- United States District Court, D. O. (2007, March). *CollegeNET, Inc. v. XAP CORP.*, 483 F. Supp. 2d 1058 - Dist. Court, D. Oregon 2007. Retrieved November 29 2010, from [http://scholar.google.ca/scholar\\_case?case=6046012089475502770&hl=en&as\\_sdt=2002&as\\_vis=1](http://scholar.google.ca/scholar_case?case=6046012089475502770&hl=en&as_sdt=2002&as_vis=1)
- W3C. (2007a, December). *Enterprise privacy authorization language (EPAL 1.2)*. Retrieved November 29 2010, from <http://www.w3.org/Submission/2003/SUBM-EPAL-20031110/>
- W3C. (2007b, 12). *A P3P preference exchange language 1.0 (APPEL1.0)* (Vol. 2007). Retrieved November 29 2010, from <http://www.w3.org/TR/P3P-preferences/>
- W3C. (2007c, 12). *The platform for privacy preferences 1.0 (P3P1.0) specification* (Vol. 2007). Retrieved November 29 2010, from <http://www.w3.org/TR/P3P/>
- W3C. (2007, 11). *Platform for privacy preferences (p3p) project*. Retrieved August 6 2010, from <http://www.w3.org/P3P/>

Warren, S., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, 4(5), 193.

# Appendix A

## Survey Questions

Include the questions used in the initial survey as well as the pre and post surveys in the final experiment.

### A.1 Privacy Contexts Survey Questions

#### 1. Age Range

- 18-21
- 22-24
- 25-30
- 30+

#### 2. Gender

- Female
- Male

3. Current employment status

- Undergraduate Student
- Graduate Student
- Doctorate Student
- Professor
- Other University Employee
- Employed Outside the University
- Unemployed
- Other

4. Please rate your agreement or disagreement with the following statements as Strongly Disagree, Disagree, Neutral, Agree, or Strongly Agree

- I consider myself an expert at using websites
- I routinely read websites where other users have contributed content, such as forums, blogs, or Wikis
- When interacting with companies and organizations, I prefer to do so over their website or through email
- My friends all use sites such as Facebook or MySpace to communicate with each other
- I feel comfortable registering an account on a website so I can post comments or contribute ideas

- When purchasing items online, I'd rather not create an account with the store and instead re-fill my information for each subsequent order
5. Please comment about your experience using the web
  6. Please rate your agreement or disagreement with the following statement: When registering an account online with University of Guelph Financial Services, I would feel comfortable giving the following information
    - Home Address
    - Phone Number
    - Email Address
    - Instant Messaging (MSN, Google Talk, etc)
    - Monthly vehicle loan payments
    - Favourite restaurants on campus
  7. Please comment about the information you would disclose to Financial Services
  8. Please list up to 5 organizations which you would feel comfortable disclosing private information to in an online environment
  9. Please list up to 5 organizations which you would not feel comfortable disclosing private information to in an online environment

## **A.2 Graphical Privacy Indicator Pre-Survey**

1. Please enter your age

2. What is your gender?

- Female
- Male
- Transgendered

3. What is the highest level of education you have completed?

- High School
- College
- Undergraduate Degree
- Graduate Degree
- Doctorate
- Post-Doctorate
- Other

4. What is your annual household income?

- Less than \$10,000
- \$10,000 - \$19,999
- \$20,000 - \$39,999
- \$30,000 - \$49,999
- \$40,000 - \$59,999
- \$50,000 - \$69,999

- \$60,000 - \$79,999
- \$70,000 - \$89,999
- \$80,000 - \$99,999
- \$90,000 - \$99,999
- \$100,000 - \$149,999
- \$150,000+

5. How many years have you been online? Please base your answer from when you first regularly used the internet, not when you were first exposed to the internet.
6. Please rate your skills in using computers and the internet as Expert, Advanced, Intermediate, Novice, or No knowledge. Feel free to expand with any comments you may have.
7. Have you ever helped someone else fix a problem with their computer?

### **A.3 University Account Registration**

The questions in Table A.1 were presented to participants while testing for the effects of graphical privacy indicators.

**Table A.1:** Account Registration Questions

Question	Question Details
Your name?	Your first and last names.
Your email address?	Your @uoguelph.ca email address.
Your mailing address?	Your home mailing address.
Your phone number?	Your primary phone number.
Your IM account?	Your primary instant messaging (MSN, Google Talk, Skype, etc) account.
Your height?	Your height in inches.
Your weight?	Your weight in pounds.
Your sexual orientation?	Your sexual orientation, such as heterosexual, homosexual, or bisexual.
How many alcoholic beverages do you drink per day?	The average number of alcoholic beverages you consume per day.
How many caffeinated beverages do you drink per day?	The average number of caffeinated beverages you consume per day.
Are you allergic to peanuts?	Choose if you are allergic to peanuts.
Do you suffer from hay-fever allergies?	Choose if you suffer from hay-fever allergies.
Are you allergic to cigarette smoke?	Choose if you are allergic to cigarette smoke.
How often do you smoke tobacco products?	Select how often you smoke tobacco products in a month.
How often do you smoke marijuana?	Select how often you smoke marijuana in a month.
Do you own a car or other motor vehicle?	Indicate if you own a vehicle such as a car, motorcycle, or boat.
If you make a monthly payment towards a vehicle you own, how much do you pay per month?	Your monthly loan or lease payment, if applicable.
How strongly you like or dislike the following University restaurant?	Indicate on a range from “Strongly dislike” to “Strongly like” your opinion of various on-campus restaurants such as Centre 6, Subway, or Tim Hortons.



## A.4 Graphical Privacy Indicator Post-Survey

1. What are the privacy risks of disclosing your Instant Messaging address online?
2. What are the risks of disclosing your phone number to an online organization?
3. What are the risks of disclosing your weight to an online organization?
4. What are the risks of disclosing your name to an online organization?
5. What are the risks of disclosing your sexual orientation to an online organization?
6. I felt there were enough details presented to me about my privacy risks to make an informed decision.
7. Could you tell how to find out about the privacy risks for a given question in the registration process?

## Appendix B

# Simple Privacy Framework

### B.1 Simple Privacy Framework Example

The following is an example of the Simple Privacy Framework. It consists of XML and uses RDF to categorize objects. Comments are within standard XML comment fields and explain the meaning of the example policy.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:p="http://squeaky.furrypaws.ca/privacy.owl#"
  xmlns:v="http://www.w3.org/2006/vcard/ns#">

  <!-- The organization from which the policy originated. This is a VCard
    expressed with RDF, matching existing standards. -->
  <v:VCard rdf:about="http://example.com/">
    <v:fn>Example.Com LLC</v:fn>
    <v:org>
      <!-- This describes the organization -->
      <rdf:Description>
        <v:organisation-name>Example.Com LLC</v:organisation-name>
        <v:organisation-unit>Corporate Division</v:organisation-unit>
      </rdf:Description>
    </v:org>

    <!-- This is the physical address of the organization -->
```

```

<v:adr>
  <rdf:Description>
    <v:street-address>33 Enterprise Drive</v:street-address>
    <v:locality>WonderCity</v:locality>
    <v:postal-code>5555</v:postal-code>
    <v:country-name>Australia</v:country-name>
    <rdf:type rdf:resource="http://www.w3.org/2006/vcard/ns#Work"/>
  </rdf:Description>
</v:adr>
</v:VCard>

```

```

<!-- This is where privacy statements begin -->
<p:Privacy rdf:about="http://example.com/">

```

```

  <!-- The jurisdictions to which this policy is applied to. -->
  <p:jurisdiction>
    <v:VCard>
      <v:adr>
        <rdf:Description>
          <v:country-name>Australia</v:country-name>
        </rdf:Description>
      </v:adr>
    </v:VCard>
  </p:jurisdiction>

```

```

  <!-- A series of statement sets containing each statement within
        the privacy policy. Each statement should be equal to one
        sentence, or at worst a short paragraph. -->
  <!-- Your cell phone number will be used to call and text
        message you for a period of 30 days. -->
  <!-- individualstatements are for a specific company -->
  <p:individualstatement expires="720">
    <rdf:Description>
      <p:source>Cell Phone Number</p:source>
      <p:action>call</p:action>
      <p:action>text message</p:action>
    </rdf:Description>
  </p:individualstatement>

```

```

  <!-- Your email address will be disclosed to the Acme Hammer
        Company for the purposes of emailing you. -->
  <p:individualstatement>
    <rdf:Description>
      <p:source>Email Address</p:source>
      <p:destination>
        <v:VCard rdf:about="http://acme.com/">
          <v:fn>Acme Hammer LLC</v:fn>
          <v:org>

```

```

        <rdf:Description>
            <v:organisation-name>Acme LLC</v:organisation-name>
            <v:organisation-unit>Corporate Division</v:organisation-unit>
        </rdf:Description>
    </v:org>
    <v:adr>
        <rdf:Description>
            <v:street-address>66 Enterprise Drive</v:street-address>
            <v:locality>WonderCity</v:locality>
            <v:postal-code>6666</v:postal-code>
            <v:country-name>Australia</v:country-name>
            <rdf:type rdf:resource="http://www.w3.org/2006/vcard/ns#Work"/>
        </rdf:Description>
    </v:adr>
</v:VCard>
</p:destination>
    <p:action>email</p:action>
</rdf:Description>
</p:individualstatement>

<!-- Your email address will be disclosed to Third Party
Advertisers who will store the address for profiling, and
remove it after 3 years. -->
<!-- For this, we use a categorystatement, that allows specifications
of relationships between entities so the policy doesn't have be
updated when business arrangements change. -->
<p:categorystatement expires="26280">
    <rdf:Description>
        <p:relation>Third Party Advertisers</p:relation>
        <p:action>profile</p:action>
    </rdf:Description>
</p:categorystatement>
</p:Privacy>

</rdf:RDF>

```

## B.2 Simple Privacy Framework OWL Schema

```

<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY dc "http://purl.org/dc/elements/1.1/" >
    <!ENTITY ns "http://www.w3.org/2006/vcard/ns#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >

```

```

<!ENTITY privacy "http://squeaky.furrypaws.ca/privacy.owl#" >
<!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
]>

<rdf:RDF xmlns="http://squeaky.furrypaws.ca/privacy.owl#"
  xml:base="http://squeaky.furrypaws.ca/privacy.owl"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:ns="http://www.w3.org/2006/vcard/ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:privacy="http://squeaky.furrypaws.ca/privacy.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <owl:Ontology rdf:about="http://squeaky.furrypaws.ca/privacy.owl">
    <dc:date>2010-07-11</dc:date>
    <dc:title>An Ontology for Privacy Policies</dc:title>
    <dc:creator>Andrew Berry</dc:creator>
    <dc:description>This ontology models and represents privacy policies in RDF.</dc:description>
    <owl:imports rdf:resource="http://www.w3.org/2006/vcard/ns"/>
  </owl:Ontology>

  <!--
  //////////////////////////////////////
  //
  // Annotation properties
  //
  //////////////////////////////////////
  -->

  <owl:AnnotationProperty rdf:about="&dc;creator"/>
  <owl:AnnotationProperty rdf:about="&dc;date"/>
  <owl:AnnotationProperty rdf:about="&rdfs;label"/>
  <owl:AnnotationProperty rdf:about="&dc;description"/>
  <owl:AnnotationProperty rdf:about="&rdfs;comment"/>
  <owl:AnnotationProperty rdf:about="&dc;title"/>

  <!--
  //////////////////////////////////////
  //
  // Object Properties
  //
  //////////////////////////////////////
  -->

  <!-- http://squeaky.furrypaws.ca/privacy.owl#destination -->

  <owl:ObjectProperty rdf:about="&privacy;destination">
    <rdfs:domain rdf:resource="&privacy;IndividualStatement"/>
    <rdfs:subPropertyOf rdf:resource="&owl;topObjectProperty"/>
    <rdfs:range>
      <owl:Restriction>
        <owl:onProperty rdf:resource="&privacy;destination"/>
        <owl:onClass rdf:resource="&ns;VCard"/>
        <owl:qualifiedCardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:qualifiedCardinality>
      </owl:Restriction>
    </rdfs:range>
  </owl:ObjectProperty>

```

```

<!-- http://www.w3.org/2002/07/owl#topObjectProperty -->

<owl:ObjectProperty rdf:about="&owl;topObjectProperty"/>


<!--
////////////////////////////////////
//
// Data properties
//
////////////////////////////////////
-->


<!-- http://squeaky.furrypaws.ca/privacy.owl#action -->

<owl:DatatypeProperty rdf:about="&privacy;action">
  <rdfs:label>action</rdfs:label>
  <rdfs:comment>An action specifying how the data will be used</rdfs:comment>
  <rdfs:domain rdf:resource="&privacy;Statement"/>
  <rdfs:subPropertyOf rdf:resource="&owl;topDataProperty"/>
</owl:DatatypeProperty>


<!-- http://squeaky.furrypaws.ca/privacy.owl#expires -->

<owl:DatatypeProperty rdf:about="&privacy;expires">
  <rdfs:domain rdf:resource="&privacy;Privacy"/>
  <rdfs:range rdf:resource="&xsd;dateTime"/>
  <rdfs:subPropertyOf rdf:resource="&owl;topDataProperty"/>
</owl:DatatypeProperty>


<!-- http://squeaky.furrypaws.ca/privacy.owl#relation -->

<owl:DatatypeProperty rdf:about="&privacy;relation">
  <rdfs:label>relation</rdfs:label>
  <rdfs:comment>A description of the type of destinations for this data item</rdfs:comment>
  <rdfs:domain rdf:resource="&privacy;CategoryStatement"/>
  <rdfs:subPropertyOf rdf:resource="&owl;topDataProperty"/>
</owl:DatatypeProperty>


<!-- http://squeaky.furrypaws.ca/privacy.owl#source -->

<owl:DatatypeProperty rdf:about="&privacy;source">
  <rdfs:label>source</rdfs:label>
  <rdfs:comment>The source of the data item (e.g. cell phone number)</rdfs:comment>
  <rdfs:domain rdf:resource="&privacy;Statement"/>
  <rdfs:subPropertyOf rdf:resource="&owl;topDataProperty"/>
</owl:DatatypeProperty>


<!-- http://www.w3.org/2002/07/owl#topDataProperty -->

<owl:DatatypeProperty rdf:about="&owl;topDataProperty"/>


<!-- http://www.w3.org/2006/vcard/ns#given-name -->

```

```

<owl:DatatypeProperty rdf:about="&ns;given-name">
  <rdfs:subPropertyOf rdf:resource="&owl;topDataProperty"/>
</owl:DatatypeProperty>

<!--
////////////////////////////////////
//
// Classes
//
////////////////////////////////////
-->

<!-- http://squeaky.furrypaws.ca/privacy.owl#CategoryStatement -->

<owl:Class rdf:about="&privacy;CategoryStatement">
  <rdfs:subClassOf rdf:resource="&privacy;Statement"/>
</owl:Class>

<!-- http://squeaky.furrypaws.ca/privacy.owl#IndividualStatement -->

<owl:Class rdf:about="&privacy;IndividualStatement">
  <rdfs:subClassOf rdf:resource="&privacy;Statement"/>
</owl:Class>

<!-- http://squeaky.furrypaws.ca/privacy.owl#Jurisdiction -->

<owl:Class rdf:about="&privacy;Jurisdiction">
  <rdfs:label>Jurisdiction Class</rdfs:label>
  <owl:equivalentClass rdf:resource="&ns;VCard"/>
  <rdfs:subClassOf rdf:resource="&privacy;Privacy"/>
  <rdfs:comment>Jurisdictions that this privacy policy applies to.</rdfs:comment>
</owl:Class>

<!-- http://squeaky.furrypaws.ca/privacy.owl#Privacy -->

<owl:Class rdf:about="&privacy;Privacy">
  <rdfs:label>Privacy Class</rdfs:label>
  <rdfs:comment>The privacy policy.</rdfs:comment>
</owl:Class>

<!-- http://squeaky.furrypaws.ca/privacy.owl#Statement -->

<owl:Class rdf:about="&privacy;Statement">
  <rdfs:label>Statment Class</rdfs:label>
  <rdfs:subClassOf rdf:resource="&privacy;Privacy"/>
  <rdfs:comment>A single statement within the privacy policy.</rdfs:comment>
</owl:Class>

<!-- http://www.w3.org/2002/07/owl#Thing -->

<owl:Class rdf:about="&owl;Thing"/>

```

```

<!-- http://www.w3.org/2006/vcard/ns#Intl -->

<owl:Class rdf:about="&ns;Intl">
  <rdfs:subClassOf rdf:resource="&owl;Thing"/>
</owl:Class>

<!-- http://www.w3.org/2006/vcard/ns#VCard -->

<owl:Class rdf:about="&ns;VCard"/>
</rdf:RDF>

<!-- Generated by the OWL API (version 3.0.0.1469) http://owlapi.sourceforge.net -->

```



# Acronyms

**AJAX** Asynchronous JavaScript and XML. 39, 40, 80

**API** Application Programming Interface. 18

**APPEL** A P3P Preference Exchange Language. 17, 20, 21

**CCTV** Closed Captioned Television. 2

**CRM** Customer Relationship Management. 8

**CSS** Cascading Style Sheets. 19, 22, 23, 34

**DPD** Data Protection Directive. 25, 42, 43

**EPAL** Enterprise Policy Authorization Language. 17, 18, 20, 21

**FTC** Federal Trade Commission. 24

**GPS** Global Positioning System. 2

**GUIs** Graphical User Interfaces. 30

**HTTP** Hypertext Transport Protocol. 18, 30, 39

**NLP** Natural Language Processing. 44

**P3P** Platform for Privacy Preferences. 17–22, 30, 32–34, 42, 43

**PIPEDA** Personal Information Protection and Electronic Documents Act. 25

**RDF** Resource Description Framework. 43, 44, 105

**SPF** Simple Privacy Framework. 6, 42–45, 49, 76, 77

**SSL** Secure Sockets Layer. 20, 79

**W3C** World Wide Web Consortium. 17, 18, 20, 21

**XACML** Extensible Access Control Markup Language. 17, 18, 21, 22

**XHTML** XML Hypertext Markup Language. 18, 22, 23, 34, 39

**XML** Extensible Markup Language. 6, 17, 20, 21, 34, 42, 105