MANAGING THE VISUAL PRIVACY OF INCIDENTAL INFORMATION IN WEB BROWSERS

by

Kirstie Hawkey

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
March 2007

DALHOUSIE UNIVERSITY

FACULTY OF COMPUTER SCIENCE

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "MANAGING THE VISUAL PRIVACY OF INCIDENTAL INFORMATION IN WEB BROWSERS" by Kirstie Hawkey in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: March 29, 2007

External Examiner: _____

Research Supervisor: _____

Examining Committee: _____

_____

Departmental Representative: _____

DALHOUSIE UNIVERSITY

DATE: March 29, 2007

AUTHOR: Kirstie Hawkey

TITLE: MANAGING THE VISUAL PRIVACY OF INCIDENTAL INFORMATION IN WEB BROWSERS

DEPARTMENT OR SCHOOL: Faculty of Computer Science

DEGREE: Ph.D. CONVOCATION: May YEAR: 2007

Dedication

This thesis is dedicated to my family, who offered me unconditional love and support throughout my years as a grad student. Thank you for your sacrifices so that I could follow my dreams.

My father, Dr. George T. Needler, unfortunately passed away while my research was just beginning. From childhood, he gave me an appreciation for scientific research and for thinking critically. I hope that this dissertation would make him proud.

Table of Contents

List of Tables

List of Figures

Abstract

Privacy can be a concern during informal collaboration around someone's personal display when traces of activity incidental to the current task are displayed. This dissertation examined how to help users manage their visual privacy within web browsers. A key goal was to allow users to maintain the functionality of their browser convenience features (e.g. Auto Complete, History, Favorites) while limiting the information displayed within the features to content that is appropriate for their current viewing situation.

We first needed to determine the extent of the problem, the nature of the privacy concerns, and the browsing behaviours which may impact the effectiveness of privacy management solutions. For this exploratory research, we employed a mixed methodology approach consisting of a survey (155 participants) and two, week-long field studies (35 participants total). The survey examined participants' privacy concerns for varying usage scenarios, while the field studies examined participants' application of a four-tier privacy gradient to their actual web browsing activity. Results identified several factors that impact a person's privacy comfort level in a given situation and enabled us to develop a model of visual privacy concerns.

Results also guided development of design guidelines for visual privacy management systems for web browsers. Such a system must support easy classification of new traces of browsing activity and provide mechanisms to appropriately filter those traces during collaboration. As documented in our results, the rapid bursts of activity and the magnitude of web pages visited will make it difficult for users to manually classify their activities with a privacy level. Our exploratory data allowed us to examine the feasibility of three privacy management approaches. Based on these results, PrivateBits, a proof of concept privacy enhancing web browser, was developed as an instantiation of our design requirements and leveraged usage patterns we observed in our field studies. An initial evaluation of PrivateBits showed that it was effective at allowing users with varying privacy concerns and browsing behaviours to manage the privacy of their web browsing traces.

List of Abbreviations Used

AOL	America Online
BHO	browser helper object
CHI	Computer Human Interaction
CSCW	computer supported collaborative work
GVU	Graphics, Visualization and Usability Center (Georgia Tech)
HCI	human computer interaction
ID	identification number
IE	Internet Explorer
IIP	incidental information privacy
NTD	non-technical desktop user
NTL	non-technical laptop user
PC	personal computer
PCL	privacy comfort level
PG1	privacy gradients 1
PG2	privacy gradients 2
PIM	personal information management
PIN	personal identification number
TD	technical desktop user
URL	uniform resource locator
WWW	World Wide Web

Acknowledgements

I conceived of the research idea, designed and executed the studies, analyzed and synthesized the data, and wrote the papers resulting from this research, including this document. But this was not a process which I undertook alone. Dr. Kori Inkpen, my supervisor, provided feedback and guidance throughout the process. I therefore have opted to use the term *we* throughout this dissertation.

Dr. Kori Inkpen became my supervisor rather than a member on my committee when I switched research topics mid-way through my graduate studies. I will ever be grateful that she took a chance with me and was willing to give me the attention and guidance I required as I learned how to move beyond academics into research. In addition to becoming a researcher, Kori has taught me to be a mentor and to give service to the research community. As I make the transition from student to faculty member, I will work hard to create a similar environment for my students.

Dr. Carolyn Watters and Dr. Raza Abidi, my committee members for this dissertation research, have also provided a great deal of guidance. Carolyn was an advocate for me during my entire time as a graduate student and her perspective and guidance was greatly appreciated. Raza was a more recent addition to my committee and his perspectives have strengthened this work.

Dr. Alfred Kobsa graciously agreed to be the external examiner of this thesis and made the trek to Nova Scotia to appear in person at my defence. I thank him for his excellent suggestions and thought-provoking questions. I would also like to thank him for not being grumpy despite the early hour.

I should also acknowledge those that provided technical assistance throughout this research. The survey was designed with the advice of Dr. Maryanne Fisher who provided excellent feedback based on her research experiences. I also received assistance during the software development of the data collection software and the proof of concept browser, PrivateBits. In particular, Shamus Peverill wrote the initial version of the on-line survey, which I then modified to meet the requirements of the study. I based the script which allowed me to access survey data from the database on a script written by Teresa Janz to

access data from a similarly structured database. I was fortunate to be developing logging tools for the field studies at the same time that Melanie Kellar embarked on a similar task for her doctoral research. The Browser Helper Object that performed the logging of visited pages was built in conjunction with Melanie. I personally designed and implemented the electronic diary that participant's used to annotate their data in the first field study; however, the version used in the second field study incorporated modifications made by Melanie Kellar. While I designed the user interface and functionality of PrivateBits, Ryder Ziola wrote the majority of the code. He based the browser functionality on code written by Melanie Kellar for her custom web browser and wrote the initial privacy management code. I made modifications to the interface and privacy management code prior to the user study.

I would be remiss if I did not acknowledge the guidance and support provided during my initial years as a graduate student. I began my graduate studies under the supervision of Dr. Jacob Slonim who was generous enough to work with me despite my research interests being at variance to his own. Jacob taught me many things including how to argue a point, how to think critically, and how to look at the big picture. My committee members at the time, Dr. Mike McAllister, Dr. Kori Inkpen, and Dr. Kenneth Rockwood also provided practical support during my research investigating how technology might assist persons with Alzheimer's Disease and their caregivers.

My journey has been made easier by many other people in the Faculty of Computer Science. In particular, Dr. Norm Scrimger (DVRG), Dr. Evangelos Milios (MALNIS), Dr. Vlado Keselj (DNLP), Dr. Carolyn Watters (WIFL), and Dr. Mike Shepherd (WIFL) all welcomed me into their research meetings when my interests converged with theirs and provided a sounding board for my ideas. And thanks to Menen Teferra, Angie Bolivar for making this process easier.

I also want to thank my fellow students who have become my extended family over the years, particularly those in the KALI Group and EDGE Lab. I could not have survived without you. You have provided a shoulder to cry on, been partners in mischievousness, given feedback on half-formed ideas, helped me run my studies, made me cool figures, and edited far too many rambling sentences (like this one). You set timers with me when it was tough to be motivated and made reward breaks fun. There were periods of time when my world was limited to sleeping and working frantically in the lab trying to hit some deadline. I

will always be the most productive when someone is playing Guitar Hero or Katamari in the background. You were the pumpkins and muffins of my world. You know who you are. Thank you.

My family has put up with a lot during what, at times, seemed like a never ending process. In particular, I would like to thank my husband Rob and son Mathew for living with my crankiness during the stressful times. Also, my sister Kate for being an extraordinary cheerleader, and my brothers Peter and Ian for reminding that life existed apart from my studies. And special thanks to my mother Margaret MacInnes and step-father Bev MacInnes who have provided me with a home and TLC (blueberry pancakes!) for weeks on end when I was working from Nova Scotia.

Finally, throughout this dissertation research I was fortunate to receive funding from NSERC, NECTAR, and the Faculties of Computer Science and Graduate Studies at Dalhousie University.

Chapter 1
Introduction

1.1 Problem Definition

Most people have seen worse things in private than they pretend to be shocked at in public.

- Edgar Watson Howe

As computers are used, transactions are generally logged in some manner, creating artifacts of the user's actions [119]. A great deal of incidental information (i.e., information that is incidental to the current task) about an individual's past activities on the computer may be visible with casual inspection. This incidental information includes file and application icons and names on the desktop, in the start menu, or in the file system itself (as seen in Figure 1). Traces of previous activities may also be visible within an application. This

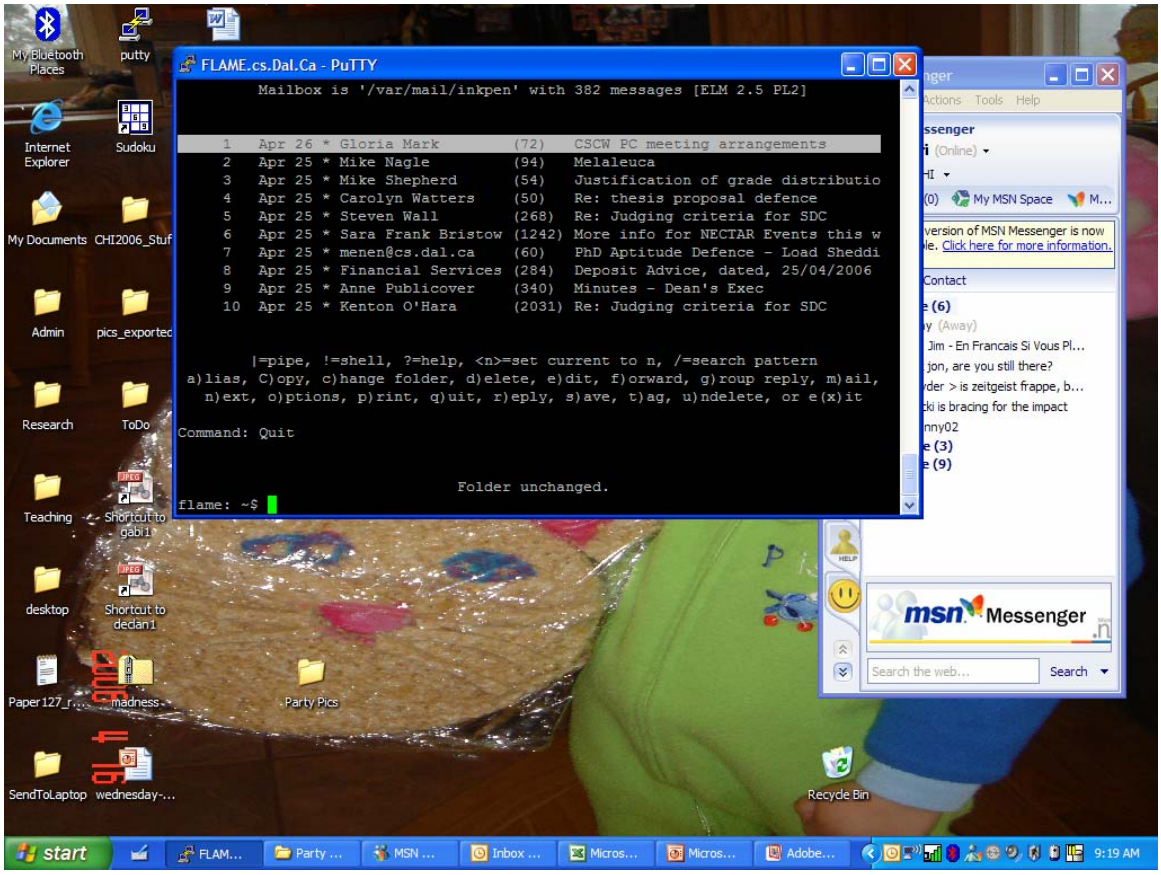


Figure 1. Example of incidental information visible on a desktop including file and application icons, personal pictures, email subjects, and contacts in Messenger.

information may or may not be appropriate for the current viewing context. For example, many a presenter has felt uneasy when a technical problem occurs during their presentation, requiring them to interact with their computer in full view of the audience. It is important to note that incidental information considered 'private' is not just that which is very sensitive (e.g., erotica, financial information, health information), it may just not be appropriate for the current viewing context (e.g., traces of personal activities viewed in a work setting, confidential business information).

Unless sharing a group machine, we generally have the notion that our computer activity is personal. The terminology used within Microsoft's Windows operating system reinforces this perception (e.g., My Computer, My Documents). Apple has chosen a less user-centric naming convention (e.g., Desktop, Documents). Ordinarily, normative privacy [107] is achieved for computer displays by physically locating the display so that others cannot easily view it [42] or by relying on the social norms that preclude others from openly staring at information on a display within someone's personal zone [42, 143]. However, there are occasions when viewing of the display is explicitly invited, such as when people gather in an *ad hoc* basis around a computer to collaborate on a project (as in Figure 2) or when a display is projected during a presentation. In these cases, normative privacy does not apply

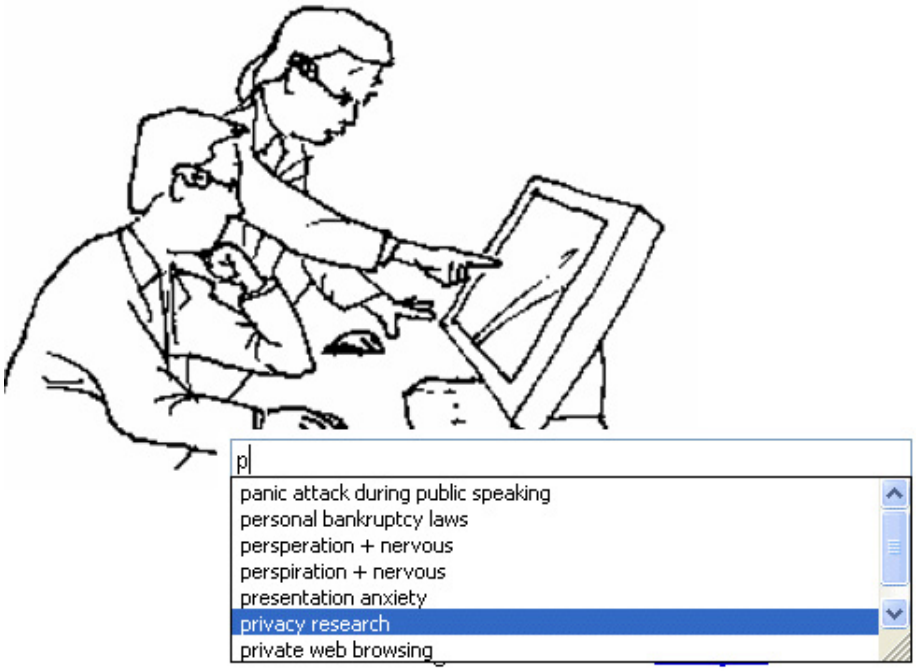


Figure 2. Example of incidental information privacy during collaboration around a personal display. Previous search terms are revealed when "privacy research" is typed in the text box.

as the display itself acts as an object in the collaboration; incidental information displayed will not only be visible, but will likely be viewed.

Web browsers were selected as the representative application for this research since they are often used during co-located collaboration to find information or share previously found web sites. In addition, web browsers are typically used for a wide variety of tasks, both personal and work related. The potentially sensitive information that may be visible within web browsers is tightly integrated with a person's actions within the web browser [91]. Web browsers have many convenience features, such as History, Auto Complete, and Favorites/Bookmarks, that are provided to assist users when browsing, but also display traces of prior activity that users may prefer to remain private. For example, if opened, the History panel will reveal previously visited web pages. Auto Complete functionality is provided both for URL completion and also for form entry. Figure 2 shows how the auto complete function will reveal search terms previously entered; during a search for "privacy research", a previous search for "personal bankruptcy laws" may be revealed.

Recently the sensitivity of search terms has been a topic in the mainstream news. In August 2006, AOL released the search terms used by 658,000 anonymous users over a three month period [102]. These search terms revealed a great deal about the interests of users and were considered to be a privacy violation. Even though only a few of the users were able to be identified by combining information found within the search terms they used, the data was soon removed from public access by AOL. What this data did highlight was the breadth of search terms with respect to sensitivity and how much the terms could reveal about the users. This insight into the extent that a person's concerns and personal activities can be revealed by the search terms they use within their web browser is illustrated in the comic shown in Figure 3. When web browsers are used during collaboration with others, privacy concerns can be magnified as the person that generated the traces of web browsing activity is not anonymous, but known to the viewer of the traces [92].

Privacy management of incidental information can be difficult for computer users. It is not always clear exactly which traces of activities are being created and stored and which can subsequently be viewed by others during normal computer usage [149]. Nor is it clear whom all the future viewers will be and the context under which material will be viewed, particularly when devices are mobile and used in both personal and business settings [119].

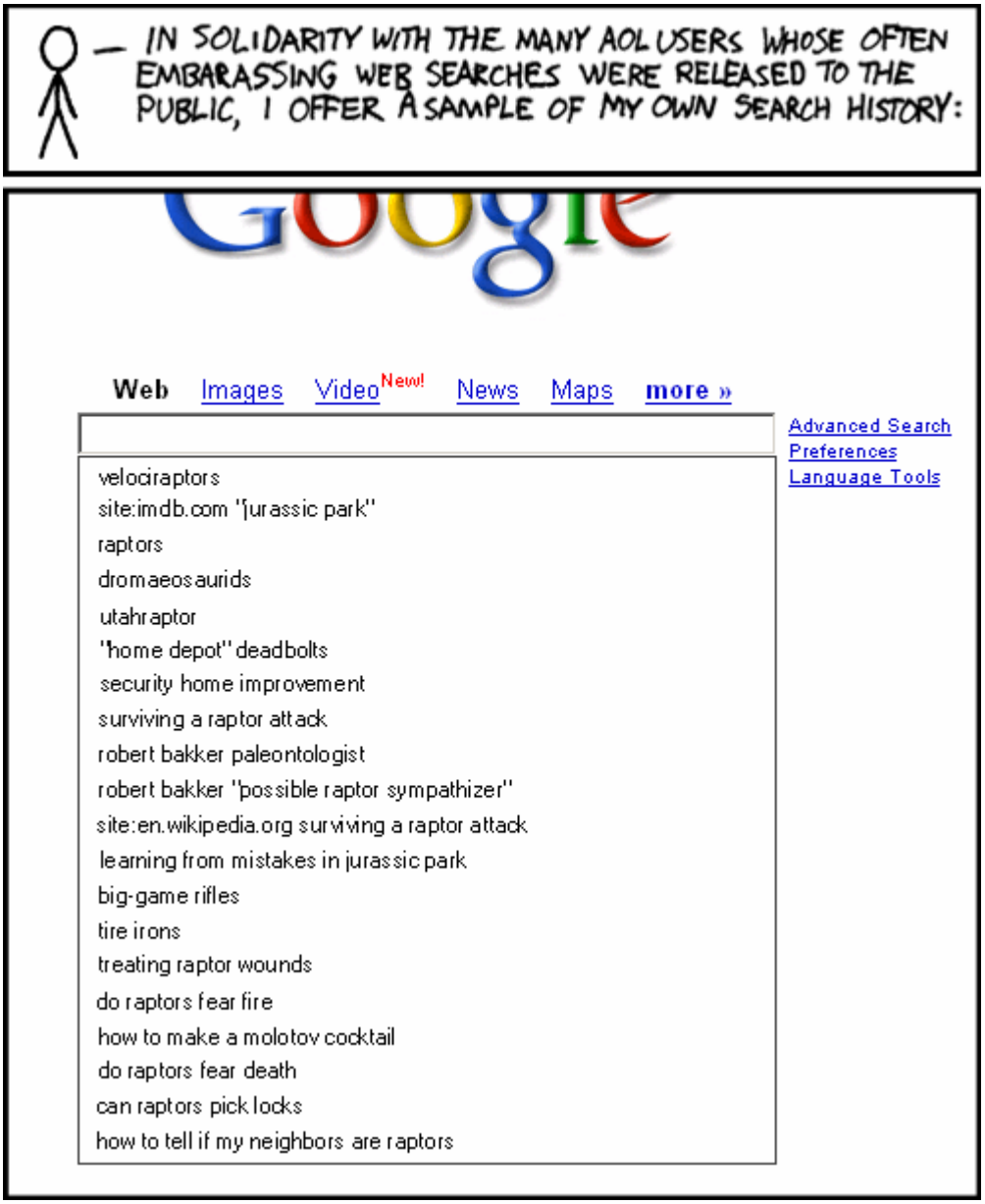


Figure 3. Comic illustrating embarrassing web searches (http://xkcd.com/c155.html).

Currently, users must make tradeoffs to manage the privacy of their incidental information. They can choose to work efficiently in a familiar environment, with access to convenience features and usual layout, but with some risk of a privacy violation if inappropriate incidental information is revealed. Alternatively, they can choose to work more awkwardly in a sterile environment. To maintain visual privacy of their previous activities within their web browsers, users must currently choose to either turn their web browser convenience features off or periodically clear the stored information with either the web

browser's tools or commercial privacy software. Commercial tools (e.g., Window Washer [4]) tend to assume that the vast majority of items are public in nature, with a small subset needing to be password protected, and that users never concurrently view sites of both types.

Research in the domain of incidental information is just beginning. Previous work in other domains has found that privacy concerns are highly nuanced and individual [8]. Tools to manage the privacy of incidental information within web browsers should allow users to only reveal information that is appropriate for their current context while maintaining the benefits of convenience features for the purpose of revisitation. Developing privacy management systems is difficult due to the diverse privacy concerns of users [8] and the many types of information that need to be protected from a variety of potential viewers [22]. There may be different levels of privacy desired depending on the relationship the individual has to potential viewers and on the type of information [107]. The amount of control that the individual retains over the disclosure of information may also impact their level of comfort [119]. The intersection of privacy management [8, 107] and personal information management [18] results in a challenging problem. The amount of control a person has over what information is displayed in their environment must be balanced with the time and effort that is necessary to provide control.

1.2 Research Objectives

The objectives of this dissertation research fall under two general areas. The first is the investigation of incidental information privacy in terms of the extent of the problem and the factors which impact privacy concerns in this domain. The second is the investigation of privacy management approaches to help users maintain the visual privacy of their incidental information within web browsers and the development a proof of concept privacy enhancing web browser.

1.2.1 Investigation of Incidental Information Privacy

Before we attempted to develop a privacy enhanced web browser to help users manage the visual privacy of their incidental information, we needed a better understanding of the privacy issues. While previous privacy research has investigated other privacy domains (e.g., online consumer privacy, information sharing privacy), we needed to conduct

foundational research to determine which aspects of privacy apply to the domain of visual privacy of incidental information.

We conducted exploratory research consisting of a survey and two field studies in order to learn more about the factors that impact privacy in this domain. Triangulating the results from these studies allowed us to develop a preliminary model of incidental information privacy. For each factor of the model, we attempted to determine the extent and variability of user behaviours and concerns. Furthermore, we examined the inter-relationship of the factors. We also learned a great deal about web browsing behaviour in general that may impact the feasibility of various privacy management solutions.

1.2.2 Investigation of Privacy Management Approaches

A second objective of this dissertation was to use our foundational research to guide development of a privacy management system to help users manage their visual incidental information privacy. Guidelines for privacy management systems were developed based upon the results of our exploratory research. We investigated how to help users with three aspects of a privacy management system: classification, filtering, and maintenance. The traces created during web browsing must be classified with privacy levels (either manually or automatically). These traces can later be filtered by the privacy management system so that only contextually appropriate content is displayed. Users must also be provided with mechanisms for maintenance of such a system. These include methods to help them inspect the privacy of their saved traces and adjust the privacy classification if necessary.

We investigated the use of content categorization of visited web pages as a mechanism to allow for automated privacy classification of traces. We also worked towards developing a predictive model of a user's privacy comfort level in a given situation that could be used to dynamically adapt which information is displayed. We have, however, left an implementation based on such automated approaches to future work. Instead, we designed and implemented a more explicit approach to privacy management. This approach leveraged privacy patterns discovered during web browsing (e.g., partitioning activities of different sensitivities between browser windows) to assist users with classifying the privacy of their information. We performed a preliminary evaluation of our proof of concept privacy enhancing web browser, PrivateBits.

1.3 Organizational Overview

We begin by presenting related work in Chapter 2. Privacy management of incidental information with respect to web browsing traces is a largely unstudied domain, but builds upon research from several areas including privacy (section 2.1), web browsing behaviours (section 2.2), personal information management (section 2.3), and research into privacy management tools (section 2.4).

Chapter 3 describes exploratory research investigating the domain of incidental information privacy within web browsers. We begin with a discussion of research methodologies for studying privacy (section 3.1) and for studying web browsing behaviour (section 3.2). We then describe our mixed methodology approach of three studies (section 3.3). The first study was the Incidental Information Privacy (IIP) survey, an on-line survey of 155 participants that explored several factors of incidental information privacy (section 3.4). The second study was the Privacy Gradients 1 (PG1) field study with 20 laptop users which examined their privacy concerns for their actual visited pages over the course of a week (section 3.5). The third study was the Privacy Gradients 2 (PG2) field study with 15 participants who were a mixture of technical and non-technical desktop and laptop users. This study was similar to the first field study (PG1), but gathered more contextual information such as the location of the browsing and the content of the visited page (section 3.6). Results from these three studies are integrated and presented throughout Chapters 4, 5, and 6.

In Chapter 4, we present results related to general web browsing behaviour. Findings from the PG1 and PG2 field studies are presented in the following areas: number of pages visited (section 4.1), browser window usage (section 4.2), speed of browsing (section 4.3) and number of browsing sessions (section 4.4). Findings from the IIP survey and the PG2 field study allowed us to examine participants' browsing activities (section 4.5) The IIP survey provided self-reports of the general types of browsing activities in which participants engaged, and the second field study (PG2) provided information about the categories of pages that participants visited and their relative frequencies.

In Chapter 5, we present results concerning the privacy of web browsing traces. Our focus in this chapter is on general privacy results, irrespective of environmental contexts such as device and location. We begin by reporting results from the IIP survey and two field

studies showing the scope of the incidental information privacy problem (section 5.1), which confirmed our motivation to conduct research in this area. We then present results concerning participants' application of privacy levels to their web browsing during the field studies (section 5.2). We next present several factors of incidental information privacy that we believe impact a person's privacy comfort level in a given situation (section 5.3). We use those factors to frame the presentation of results from the IIP survey and field studies pertaining to privacy in this domain, specifically the sensitivity of potentially visible content (section 5.4), the person's relationship to the viewer of the information (section 5.5), the level of control retained over input devices (section 5.6), and a person's inherent privacy concerns (section 5.7).

In Chapter 6, we investigate how browsing activities, web browser settings, and actions taken to preserve privacy combine to determine which content is potentially visible in web browsers. In this chapter we explore the inter-relationship of dispositional and situational variables and their impact on participants' activities and privacy concerns. Results are presented from the IIP survey and the contextual data captured during the PG2 field study. We first examine the impact of dispositional variables such as our participants' demographics and life experiences on their inherent privacy concerns (section 6.1). We then examine the impact of situational variables within the browsing environment (e.g., location, device) on inherent privacy concerns (section 6.2). We also examine the impact of this environmental context on the overall application of privacy levels by participants in the PG2 field study (section 6.3). We then break down the possible causes for the differences found. We examine how the environmental context affects browsing activities (section 6.4), browser convenience features settings (section 6.5) and the privacy preserving actions taken (section 6.6), all of which contribute to what content is potentially visible within browser convenience features. Finally, we examine whether the same types of content are perceived as having differing privacy concerns across usage contexts (section 6.7).

Chapter 7 examines the feasibility of various privacy management approaches. We begin by presenting the design requirements we developed for a visual privacy management system as a result of our exploratory analysis (section 7.1). Then, in light of those requirements, we discuss three components of such a privacy management system: classification of traces of web browsing activity, filtering of that information appropriately

during viewing situations, and maintenance (section 7.2). We then present an analysis of the feasibility of automated approaches for classification of traces (section 7.3) and for filtering of content according to the current viewing context (section 7.4). Finally, we discuss the current technological limitations to automated approaches which led us to develop a more explicit approach (section 7.5).

Our exploratory research identified design requirements and proposed an approach for semi-automatically classifying the privacy of traces of browsing activity. This approach leverages browser-window based temporal patterns observed in participants' application of privacy levels during web browsing. With this approach, the onus remains with the user to manage the classification of their browsing with system support. Chapter 8 presents the design, implementation, and evaluation of PrivateBits, an instantiation of a browser window based visual privacy management approach. We first present the design and implementation of our proof of concept privacy management system (section 8.1). We then reflect on how this design fulfills the identified design requirements (section 8.2). We next present the methodology for our preliminary evaluation of PrivateBits (section 8.3). We then present results of the evaluation and reflect on the effectiveness of the interface at meeting participants' varying privacy concerns and browsing behaviours (section 8.4) and discuss issues of trust and concealment of the privacy management system itself that are unique to privacy management systems (section 8.5).

In Chapter 9, we reflect upon the suitability of the methodological approaches taken during this research. We first discuss the suitability of the mixed methodological approach used for our exploratory studies of incidental information privacy concerns (section 9.1). We then reflect on the effectiveness of participant annotation of logged data as a method of studying rich natural behaviours in situ (section 9.2). Finally, we discuss the effectiveness of conducting a laboratory-based evaluation to initially investigate the usability and utility of PrivateBits (section 9.3).

Finally, in Chapter 10, we give a summary of this dissertation research (section 10.1) and itemize the contributions of this dissertation (section 10.2). We conclude with directions for future work (section 10.3).

Chapter 2
Related Work

Privacy management of incidental information with respect to web browsing traces is a largely unstudied domain, but builds upon research from several areas including privacy, web browsing behaviours, personal information management, and the development of privacy management tools. This chapter will present related research in each of these areas.

2.1 Privacy

There are several aspects to privacy relevant to this thesis research. We first present applicable privacy theories, particularly those that consider how privacy concerns change depending upon the context of the situation (e.g., for different viewers, in different settings). We next present relevant findings from other privacy domains such as information sharing. Finally, we present related work with respect to modeling privacy. Where relevant, we reflect on the applicability of the related research to our study of visual privacy concerns within web browsers.

2.1.1 Privacy Theory

Privacy is a fluid concept and privacy theories and definitions vary according to the domain in which the privacy issues occur. This section explores privacy theories that are most closely related to the domain of incidental information privacy.

Boyle and Greenberg [19] developed a vocabulary for interpersonal operational privacy in video media spaces. The three central modalities by which people control their privacy boundaries within a video media space are solitude (i.e., control over interpersonal interactions), confidentiality (i.e., control over access to one's personal information), and autonomy (i.e., control over one's own actions and expression of identity). For visual privacy of incidental information, solitude does not apply as the privacy violation occurs when an individual specifically invites another person to view their display. Boyle and Greenberg define privacy sensitivity as "a property of a piece of information that can be defined as a perception of how important it is to maintain control over access to it" [19].

Westin [150] defines individual privacy as "the claim of an individual to determine what information about himself or herself should be known to others". Over the past forty years, Westin has primarily dealt with consumer privacy rights (e.g., when personal information can be collected, how others can make use of the information). Visual privacy of incidental information is simpler in some respects. As there is no electronic transfer of information, issues relating to when personal information can be captured and later uses of the information are moot. Furthermore, relationships are interpersonal in nature, rather than business/consumer. Therefore, social norms can mitigate many visual privacy concerns. For example, there are social conventions as to when it is acceptable to view information on a computer display and when it is acceptable to act or disseminate such information. However, once others have been explicitly invited to view a display (e.g., during collaboration or when a display is projected), privacy concerns can arise when information that a person may not want to share with others is inadvertently revealed.

Westin [150] also discusses how individuals seek a balance between maintaining privacy and fulfilling a need for communication and disclosure. How an individual balances that privacy depends on their personal situation including their family life, education, social class, and psychological composition. Furthermore, Westin states that an individual's needs are highly contextual and continually shift depending on situational events.

The contextual nature of privacy is well established in the literature. Goffman [49] first introduced the need to project different personas or faces during social interactions. The face presented in any given situation depends not only on the current audience but also on the current conditions. The combination of audience and situation determines how much and what information will be disclosed. Furthermore, as discussed by Palen and Dourish [119], people can have many roles between which they fluidly move and can act in multiple capacities, often simultaneously. For example, one may act as an individual, a family member, and a representative of an organization. This can make a purely role-based approach to privacy management difficult. If information is conveyed that is out of character for the person's current role, the boundaries that have been maintained can collapse creating opportunities for social, bodily, emotional, and financial harm [129].

Palen and Dourish [119] describe three inter-related boundaries for privacy management: the disclosure boundary, the temporal boundary, and the identity boundary.

Boundaries between what is considered to be public or private are continuously refined depending upon the context. The disclosure boundary is the tension between privacy and publicity of information, opinions, and actions as one chooses to present a persona of oneself to the current audience. The temporal boundary is the tension between past, present and future. Not only does information tend to persist over time, but one's privacy concerns in the present are likely shaped by similar circumstances in the past. The identity boundary is defined as the boundary between self and others and is complicated by group membership, such as social or professional affiliations. This model of privacy fits incidental information privacy well. Users would like to be able to control an appropriate level of disclosure given the context of viewing (*disclosure boundary*). The temporal persistence of traces of previous activity (*temporal boundary*) makes it difficult for users to ensure that they are presenting themselves appropriately for their current role (*identity boundary*).

The impact of privacy violations depends in part on the content of what has been revealed, and the costs of a violation can be both imagined and realized [99]. Phillips [130] discusses how people vary in their perception of the utility of privacy and also in their sense of the dangers of a privacy violation. He discusses four types of privacy concerns: freedom from intrusion, constructing the public/private divide, identity management, and surveillance. Phillips' concept of identify management is similar to Palen and Dourish's [119] identity boundary. Of the remaining three types of privacy concerns, freedom from intrusion and surveillance are most relevant to visual privacy concerns.

As defined by Phillips, freedom from intrusion affords individuals the freedom to express themselves within their personal sphere without intrusion from others, either in the form of government action (e.g., searches without warrants) or through the pressure of social norms [130]. Privacy in this sense supports social interaction and "healthy functioning by providing needed opportunities to relax, to be one's self, to emotionally vent, and to cope with loss, shock, and sorrow" [99]. Increasingly the Internet has become a mechanism by which people can engage in activities to support their emotional needs (e.g., surfing the web, visiting personal support forums, blogging, investigating health concerns) [150]. Content visible within web browsers may therefore include sensitive items such as socially inappropriate activities, confidential business items, and personal activities conducted on

company time as well as more neutral items (e.g., situation-appropriate content, weather information).

Phillips focuses on surveillance as a privacy concern at the societal level, as a method by which the observations of many individuals are aggregated and used to create and manage social knowledge [130]. For our purposes, a more traditional view of surveillance is appropriate, whereby surveillance is considered at the individual level (i.e., Big Brother is watching). Privacy and surveillance are aspects of the same concept, with privacy actions serving as a nullification mechanism against surveillance [101]. Several methods of maintaining privacy in case of surveillance have been identified [101], with *self-regulating*, *blocking* and *masking* activities being particularly applicable to privacy in this domain. For example, web browsing activities may be *self-regulated* in the workplace to avoid surveillance by an employer [101], with more personal activities being conducted solely at home. A person's attitudes and perceptions about privacy, trust, and social relationships or norms (e.g., workplace rules) will influence his behaviour in a situation [95].

A common privacy preserving strategy employed with web browsers is to *block* the recording of visited sites by turning off the convenience features. This strategy is likely to be a contributing factor to the underutilization of web browser convenience features for the purpose of revisitation [12, 80, 86] (convenience feature use is discussed further in section 2.2.2). Another downside to this approach is that a complete lack of visited sites within the browser's History may be viewed as an indicator that there is an activity worth hiding. A more subtle approach would be to *mask* the activity rather than to block it completely [101]. For example, to mask browsing activities in Favorites, users can rename stored sites to conceal the nature of the page. Options to more selectively manage History and Auto Complete entries are needed.

It must be noted that guarding the visual privacy of incidental information within web browsers will not protect employee privacy if an employer is conducting workplace surveillance; many employers monitor internet activity on the web server [99]. In such cases, employees may avoid surveillance of their activities by avoiding activities which may raise a red flag (e.g., viewing pornographic sites) thereby warranting closer inspection by management [101]. Some may also opt to use a co-worker's account or shared machine so that the activities they undertake are not directly traceable back to them [101].

Lederer et al. [91] discuss personal privacy of electronic information flow within a ubiquitous computing context. They qualify personal privacy as being a "set of both deliberate and intuitive practices by which an individual exercises her claim to determine personal information disclosure and which constitutes, in part, her participation in the co-evolving technologies and expectations of everyday life". They also discuss how personal privacy allows one to "maintain compound roles in the socio-technical contexts of everyday life". This view on personal privacy relates well to the visual privacy issues we are investigating. Web browsing has become an aspect of everyday life and occurs across multiple roles and contexts.

Lederer et al. [92] discuss how activities convey the essence of a persona. Knowledge of an individual's prior activities is more sensitive when their identity is known as the activities can reveal hidden personae. With traces of incidental information, a person's actions in one area (e.g., personal browsing) may later be viewed in another area (e.g., workplace). Information that is appropriate for a friend to see may not be appropriate if viewed by an acquaintance or an authority figure with whom one would prefer to present a more formal or otherwise restricted face.

Moor [107] uses a "control/restricted access" theory of privacy. Users can fine-tune the privacy of their information by both recipient and information type via zones of privacy. However, with incidental information, not all potential viewers of the information may be apparent at the time the traces are created.

2.1.2 Research Investigating Privacy Concerns

Results from research investigating privacy concerns in other domains may not be directly applicable to the incidental information privacy, but can provide insights. We next present relevant privacy research from the domains of online privacy, information privacy, and other domains such as computer supported collaborative work (CSCW) and ubiquitous computing.

2.1.2.1 Online Privacy

The Platform for Privacy Preferences Project [3] has developed standards that facilitates user awareness of the privacy policies that govern the use of their personal information at participating websites. However, online privacy research has a different focus

from the web browser privacy issues we present here. Online privacy research generally examines issues concerning the transfer of personal data to business or governmental entities; the relationships are between consumers and corporations. This may be quite different from the privacy concerns associated with others viewing traces of previous web browsing activity in a co-located setting. Although in both cases personal information may be viewed, there are differences in the nature of the relationship to the viewer of the information. For visual privacy within web browsers, the relationship between the user and the receiver of the information is not business-consumer, but rather interpersonal in nature (within the workplace, there may also be an organizational component). Furthermore, the viewers of incidental information within web browsers are not usually anonymous, but are known to the user which may heighten privacy concerns [92]. Additionally, information is viewed but not electronically transferred.

A 1998 survey by Ackerman et al. [8] examined privacy preferences for Internet users. The authors found differing levels of sensitivity about personal data, ranging from little concern about providing such information as their favourite television show to great concern over credit card and medical information. Interestingly, 18-20% of the participants expressed concern over even the most innocuous data. The authors suggest that an individualized approach is necessary given the large variance in reactions.

2.1.2.2 Information Privacy

Recent research into information sharing has looked at privacy comfort for various types of information and recipients of that information. Cadiz and Gupta [22] found that, in general, people were open to sharing information except with strangers. Cadiz and Gupta also found that participants' privacy concerns were highly nuanced. A similar study by Olson et al. [115] investigated privacy comfort for participants sharing information with a recipient. They found that the recipients of the information could be clustered into four groups according to the level of privacy concerns associated with the recipient: public, work relationships, family, and spouse. During a preliminary phase of their study, they asked participants to give instances when they had shared something that they later regretted sharing. This information was used to inform the types of information examined in the second phase of their study. The second phase of Olson et al.'s study had participants give comfort levels for each instance of sharing 40 kinds of information with 19 types of people.

Their results suggest that the types of incidental information that may be revealed during web browsing (e.g., personal activities like viewing non-work related websites, transgressions like viewing erotic material) are considered more sensitive than other content (e.g., contact and availability information). Privacy concerns for incidental information arising from web browsing may be less clear-cut than for static information (e.g., contact information) that may be shared electronically. There are likely several levels of sensitivity of content within the traces, the amount of sensitive content may fluctuate over time, and the user may be less aware of what is actually saved.

2.1.2.3 Other Research Areas

Privacy issues have been addressed extensively in distributed CSCW research, particularly in relation to capturing and displaying awareness information in an attempt to replicate some of the benefits of co-located collaboration. However, there is a tradeoff between providing awareness information and maintaining privacy. For example, Palen [118] identified several privacy issues during a study investigating the use of groupware calendar systems. Participants had concerns about the personal privacy of their information (e.g., medical appointments), the social sensitivity of the information (e.g., internal job interview), and the security of company information (e.g., business strategies revealed by appointments with other companies). Additionally, participants were concerned that the information contained on the calendar would lead to judgments about how they managed their time.

Privacy issues also surround the use of video in awareness tools, particularly if the video could capture images of those unaware they are being viewed, as in the Notification Collage [52]. Users of that system commented that they felt more inhibited than normal and often felt the need to put on a public persona while at home by changing their dressing habits and "excusing the mess" visible in a video. Strategies for maintaining privacy in awareness systems include only storing and presenting aggregate data where possible [14] and adjusting the level of detail of information displayed depending on the size and public nature of the display [71].

Privacy issues raised in co-located CSCW research have been primarily limited to the privacy of data within an application often using specialized dedicated devices. For example, Shoemaker and Inkpen investigated the use of shuttered glasses [138] that are calibrated with the refresh rate of a monitor. The shuttered glasses are configured so that he odd frames are

viewable by one user and the even frames by the other user. Information that is private is encoded to appear on only half the frames, while public information is shown on all frames. However, this view of private information assumes that all information viewed is task-related. During collaboration around someone's computer, this may not be the case.

More recently, researchers have been investigating privacy issues that occur as a result of multiple display environments. For example, Huang and Mynatt [71] discussed privacy issues that arise when information is appears on semi-public displays within a small group environment. They found that privacy concerns can increase when displays are viewable by many people in a group and it is not clear which information is being viewed, by whom, and how often. Hutchings and Pierce [73] have investigated how (and why) people might divide an application's interface across devices in private, semi-private, and public environments. Privacy issues were a factor for all participants when choosing an appropriate division of the interface or device to use. Indeed, most participants (15/18) were concerned about others viewing sensitive information on their displays. This was particularly true in the semi-public work environment where many participants wanted to shield personal activities from their colleagues.

Privacy is also addressed in the ubiquitous computing community. Lederer et al. [93] examined the relative importance of the inquirer (spouse, employer, stranger, merchant) and the situation for the preferred accuracy of personal information disclosed (e.g., location). Participants' preferred level of accuracy was found to vary by inquirer, but not by situation (except when the inquirer was the employer). Patil and Kobsa [127] studied privacy issues related to the use of Instant Messenger. They also noted differences in privacy concerns for different viewers which suggested a more fine-grained approach to managing privacy levels for contact lists. As discussed in section 2.1.2.2, Olson et al. [115] found that viewers of information clustered into the groups spouse, family, work, and public (ordered by increasing privacy concerns for this viewing). Lederer et al. [93] found that there may be increased privacy tensions for hierarchical relationships (e.g., supervisor-employee) which they attributed to the desire for solitude (i.e., an employee may not want a supervisor to be able to contact her outside of working hours). Similarly, Patil and Kobsa [127] found no difference in terms of comfort between superiors, subordinates, and strangers.

2.1.3 Privacy Models

Researchers have developed privacy models for other domains. These models inform our overall understanding of privacy issues. We next present related work concerned with developing general models within a specific privacy domain and also models which attempt to segment users according to their privacy concerns.

2.1.3.1 General Models

Adams [11] developed an abstract model for users' privacy perceptions of multimedia information during communications. The primary factors of her model include the user's judgment as to the sensitivity of the information, their trust in the receiver of the information, and their determination of the costs and benefits of the usage of the information. Each of these factors interacts with the others and within the overall context of the situation (i.e., technology used, social groupings, national/international settings). Adams discusses how privacy is not a binary attribute; the sensitivity of information varies across many degrees. One point Adams makes that is particularly relevant to visual privacy of incidental information is that privacy concerns are often associated with the secondary information that is relayed. For example, it may not be the content of a discussion (the primary information) that is perceived to be sensitive, but the language used (e.g., abusive language), the verbal cues (e.g., tone of voice), or the visual cues (e.g., mannerisms, dress). With visual privacy of incidental information, it is not the information related to the task at hand that is sensitive, but the secondary traces of information viewed during interactions with the computer.

Malhotra et al. [98] have developed a causal model of online consumers' information privacy concerns. Their model considered the effect that Internet users' information privacy concerns have on trusting beliefs, risk beliefs, and their behavioural intention to reveal personal information. Furthermore, they incorporated the sensitivity of the information requested by marketers as a contextual variable and considered covariates such as sex, age, education, Internet experience, identity misrepresentation, past experiences with privacy invasion, and media exposure. They developed measures for new factors of privacy concerns including control (i.e., whether the user has control over the data) and awareness (i.e. whether the user is adequately informed as to use of the data) to augment existing scales for

this domain which consider collection of information (i.e., whether the exchange of personal information is equitable).

Recent research (such as [8, 77]) has been cautioning that actual behaviour with respect to privacy practices often does not follow stated privacy concerns. We therefore consider it important to not only rely on self-reported data about incidental information privacy concerns, but to also investigate the privacy concerns within the context of actual web browsing behaviour. Acquisti [9] has proposed enriching privacy models by including psychological models of personal behavior such as immediate gratification and self-control.

2.1.3.2 Segmented Models

In addition to more general models of privacy concerns, an effort has been made to model privacy for subgroups of the population. Despite individual differences in privacy concerns, individuals may be able to be grouped according to their privacy characteristics. The Westin-Harris [116] privacy segmentation model explores consumers confidence in how personal information is collected and used by companies. The model partitions consumers into three privacy categories: *privacy fundamentalists*, *privacy pragmatists*, and *privacy unconcerned*. Privacy fundamentalists feel strongly that current information practices are a threat to their privacy, while those classified as privacy unconcerned have the opposite viewpoint. Privacy pragmatists tend to weigh the risks of releasing personal information (e.g., receiving spam emails) against the potential benefits (e.g., personalization of a web site).

Spiekermann et al. [139] also studied online consumer privacy issues and further divided the pragmatists into two groups: the *identity concerned* and the *profiling averse*. Identity concerned participants were most concerned about revealing personal contact information such as their name and address to corporate sites, while the profiling averse were more concerned about information such as health status and hobbies.

Sheehan [137] conducted an email survey (889 total respondents) and examined participants' privacy concerns for the collection and use of personally identifiable information by companies. The survey inquired about privacy concerns for 15 online situations using a 7-point scale. The scores were summed across all the situations (maximum total score 105). Sheehan partitioned participants in a similar fashion to the Westin-Harris model, and found that only 3% of the total participants would be considered to be

fundamentalists (*alarmed internet users*, score < 30/105), 16% would be considered unconcerned (*unconcerned internet users*, score > 90/105), and 81% would be considered pragmatists (score 31/105 to 90/105). The pragmatists were further subdivided based on their total scores and classified as either *circumspect internet users* (38% of total respondents, score 31/105 to 60/105) or *wary internet users* (43% of total respondents, score 61/105 to 89/105).

The consumer-based privacy segmentation model has been applied to other privacy domains with limited success. Consolvo et al. [33] did not find the model to be a good predictor for disclosure of awareness information. Patil and Lai [128] used an extended questionnaire with a trust component to model their participants, but did not find correlations between the awareness information settings participants would choose and their questionnaire scores. Olson et al. [115] used nine questions from Butler's trust scale [20] and demographic data (e.g., age, gender) in an attempt to find a small number of questions that indicate privacy preferences for information sharing. Although they state that interesting patterns emerged, none were statistically significant. It is clear that privacy segmentation may vary depending on the privacy domain and that segmentation methods for a given domain must reflect the nature of privacy concerns within that domain. Consumer privacy segmentation models may not be relevant in a domain with interpersonal privacy concerns, as in the case of viewing incidental information.

Patil and Kobsa [127] found their participants to have a wide range of privacy concerns with respect to the use of Instant Messenger. However, they found three levels of privacy concern (high, medium, low) to be effective for discerning privacy attitudes. The burden of privacy management in this and other systems may be reduced through the use of templates that are appropriately set for a sub-group of users with similar concerns.

2.2 Web Browsing

Web browsing is the primary task for users, with privacy management being a secondary consideration. In order to understand how to support visual privacy in web browsers, it is important to understand users' behaviours, activities, and the features they use while browsing. Any privacy management solutions that we develop will have to be effective at managing privacy without disrupting users' desired web browsing behaviours.

Web browsing environments have been continually evolving. Appendix A gives a detailed timeline of changes in the typical web browsing environment and the typical user. This gives temporal context to the seminal web browsing research. For example, in 1994 the typical web user was a young, technical male, using a browser with limited features over a slow connection. Today, users come from all segments of the population and are often using browsers equipped with advanced navigation and search features over a high-speed connection. Web browsing increasingly occurs in mobile contexts as laptop computers and other devices accompany individuals as the move between the workplace, home, and school. Despite these contextual changes to the web browsing environment, seminal works, such as Catledge and Pitkow's [25] 1994 study, are still used as a motivation for new web navigation tools and techniques. This early research needs reevaluation against current contexts of use to see if the results are still appropriate.

2.2.1 Web Browsing Behaviours

Web browsing behaviour has been studied from a variety of perspectives. Research has considered both general web browsing behaviours (e.g., the study of web page revisitation patterns, as in [145]) and more specific areas such as information seeking behaviour (e.g., searching, as in [35]).

2.2.1.1 General Web Browsing Behaviour

One of the first studies examining users' web browsing behaviour was Catledge and Pitkow's 1994 study [25]. A modified version of XMosaic was used to log browsing activity over the course of 3 weeks. The two dominant methods of navigation by participants were hyperlinks and the back button. Tauscher and Greenberg [146] also observed user behaviour with a modified version of XMosaic and studied the revisitation patterns of users. Over a six week period in 1995, they observed that 58% of page visits were revisits. Cockburn and McKenzie [31] conducted a retrospective observational study (from October 1999 to January 2000) of History and Bookmark files retrieved from server backups. They found an average revisitation rate of 81%. Their analysis also showed that Bookmark use was highly variable. More recently, Weinreich et al. [148] reported a revisitation rate of 46% during a longer term study (avg. of 105 days captured, ranging from 52-195 days) in 2004-2005. Revisitation rates

can give us insight as to how many new pages may need to be classified with a privacy level in a privacy-enhanced web browser.

Individual differences have not received a great deal of attention in previous web research. Even in cases where individual user behaviour is distinguishable from one another, it is typically aggregated in order to develop general user models (as in [51]). However, web experience, occupation, and technical background can play a role in a user's behaviour and can contribute to large differences between users. Issues in the interpretation of study results can arise when behaviours exhibit large variability, as in [31]. In this study, participants were recruited from within the academic community, but one person was employed as a webmaster and had a much higher level of web usage. This participant was marked by the study researchers as an outlier and findings were reported both with and without his data where applicable. In other studies (e.g. [106]), researchers have not identified individuals which may skew the interpretation of overall patterns of behaviour.

Recent research by Herder and Juvina [68] has examined the impact of individual differences on participants' navigational styles. They investigated the impact of several psychological measures on web browsing activity including spatial ability, episodic memory, working memory, as well as internet expertise, affective disposition, and locus of control. They classified participants as either having a flimsy navigation style (i.e., small number of pages visited, high median view time, high rate of home page visiting) or a laborious navigation style (i.e., high number of links followed per page, high revisitation rate, high return rate). High scores on flimsy navigation were associated with low scores on Internet expertise, current mood, and working memory. High scores on laborious navigation were associated with high episodic memory scores and low spatial ability scores. The authors plan to use this information to predict which users may experience disorientation while navigating and provide adaptive navigation support that is appropriate for the navigation style of the user.

2.2.1.2 Web Browsing Activity

Task-related information seeking research is particularly relevant to this dissertation research as it gives insight to the types of web pages that people visit. During a study of knowledge workers in 2002, Sellen et al. [15] interviewed participants in front of their history lists and had them describe the web activities they had recently completed. Activities

consisted of: transactions (5%), communications (4%), housekeeping (5%) and information seeking (86%) such as fact finding, information gathering, and browsing. A more recent field study by Kellar et al. in 2005 [87] found that transactions accounted for 47% of the visited pages, with email being the most common transaction. Information seeking (fact finding, information gathering) accounted for 32% of visited pages and browsing for 20%. It is hard to compare results from these two studies directly as Sellen et al. presented their findings based on the percentage of activities participants recalled conducting and Keller et al. presented their results based on the percentage of pages participants actually visited. Kellar et al. also found that the nature of the task impacted the convenience features used [87].

Most of the research categorizing web browsing focuses on actions that people take and not on the type of content that is being viewed. For example, Byrne et al. [21] conducted a task analysis of user web behaviour in 1998. Participants were video taped in their offices as they used the Web over the course of a day. Participants spent the majority of the time on the Web reading the pages they visited and their most common navigation method was the use of hyperlinks, followed by the back button. Typically, content is examined through self-reports of the types of activities (e.g., shopping) participants engage in on the web (as in [108]). One exception is research by Curry [36] who sampled the URLs viewed by public library users and classified them by format and by subject. The author found that 39% of visits were email related. Not all pages received a subject categorization, so content analysis in terms of relative amount of activity is limited.

There are many content classification schemes in commercial use, such as the Yahoo Directory [5] which categorizes web pages using fourteen main headings and hundreds of subcategories. There are also commercial tools (e.g., [6]), both for corporate and parental use, for filtering out content that is deemed inappropriate. These tools may classify web pages into categories or use some combination of keywords and URL lists to filter inappropriate content and sites. However, web content filters suffer from both over blocking (i.e. blocking sites unnecessarily) and under blocking (i.e. not blocking sites that should be blocked) [72]. A recent examination by Consumer Reports [34] shows that although research continues to improve content filtering, commercial systems are often ineffective.

2.2.1.3 Multiple Browser Windows

A thorough literature review has revealed little direct study of user behaviour with multiple browser windows. Advancing knowledge of how users partition their browsing activities between windows may be particularly useful for the development of web browsing tools and techniques. For example, users may have different windows open for different purposes such as a literature search in one window, email in another, news in a third. The Elastic Windows browser, introduced by Kandogan and Schneiderman in 1997 [83], allowed users to not only have multiple pages open within a browser window, but also arrange them in terms of size and location. Commercial browsers (e.g., FireFox) have implemented tabbed browsers, allowing users to organize multiple open pages within the browser.

Commercial privacy management tools generally assume that sites of varying sensitivity are never viewed concurrently, allowing either a private mode or a public mode, but not both. However, experienced users often maintain several open browser windows (or tabs in the case of tabbed browsers). Aula et al. [12] conducted a survey of 236 experienced web users regarding their information seeking processes. Participants reported using multiple windows or tabs often during the search process. Multiple windows can be used as a means of in-session revisitation of web pages, to help manage the search process (e.g. one window for the query, and other windows to investigate results, and for multi-tasking [12]. Users may have multiple search goals [17] and may switch between windows and tasks, particularly when pages are slow to download [12]. As an example, Jones et al. [80] observed a participant using multiple web browser windows to represent separate search topics as well as for searches for the same topic on different databases.

Multi-tasking may also be a result of the simultaneous roles a person is performing (as suggested by [119]). For example, someone may be conducting a search for information related to a problem at work and simultaneously be searching for information related to a family activity or a personal concern. A privacy management system should support concurrent windows containing content of varying privacy sensitivity. In an examination of web browsing strategies, using data from direct observation, user surveys, and server logs, Clark et al. [28] observed multiple browser windows being opened simultaneously. The authors found that in addition to accessing an informational site as part of coursework, many

students were also using the Internet for other coursework or research in addition to surfing the web for non-academic purposes.

2.2.2 Web Browser Convenience Features

Web browser convenience features have been developed to allow users to more easily revisit content. These features work by storing traces of web pages visited or text entered. The storage may be explicit, as when a user opts to add a page to their Favorites, or may be performed automatically, as in the case of History and Auto Complete.

Browser convenience features such as Favorites/Bookmarks and History, which are designed to assist with re-visitation, are often under utilized [12, 80, 86]. Researchers have investigated different mechanisms and algorithms for displaying the traces of prior activity within browser convenience features in an effort to improve their usability (e.g., [145]). For example, Kaasten and Greenberg [81] have proposed integrating the Back Button, History, and Bookmarks into one feature. Their solution was a history list ordered temporally, with duplicates deleted, that provided users with a mechanism for explicitly marking pages they felt they were likely to revisit. Another project [94] has looked at methods to automatically organize the History into relevant topics. During a preliminary evaluation of two variations of this technique, participants reported that the topical organizations of history were more similar to their mental organizations than Internet Explorer's (IE) History. The techniques were faster when used for revisitation.

The quantity of traces saved is one barrier to convenience feature use as it can make recognizing the desired resource difficult. For example, History displays both irrelevant pages and those that are important to the user [12]. While Favorites/Bookmarks contain only those pages that were deemed to be important enough at some point to save explicitly, they also suffer from clutter and disorganization [12, 18]. Privacy management systems may be able to help reduce the clutter by allowing more control over what traces are stored. Furthermore, such systems should be designed so as to not interfere with the primary purpose of the convenience features (i.e., revisitation).

2.3 Personal Information Management

In this section relevant work from the Personal Information Management (PIM) research domain is presented, including a discussion of the relationship of incidental information privacy to PIM systems and the management styles of PIM users.

2.3.1 Relationship of IIP to PIM Systems

Personal Information Management is a growing research area. A report from the recent 2005 PIM workshop [79] defines personal information as information kept for personal use; information about a person that may be kept by and in control of others (e.g., health information retained by a doctor); and information experienced by a person, even if that information remains outside a person's control (e.g., a library book that has been read and returned). A personal information space is considered to be all the information under a person's control and the tools to manage that information [79]. Most of the traces of previous activity that appear in web browsers are not considered to be 'personal information' as traditionally defined in the PIM community with the exception of cached pages and user created bookmarks (i.e. visited web pages are explicitly excluded in the first workshop report) [79]. However, workshop organizers did express that "the personal information space should probably include the icons that applications like to leave on our computer desktops and the bookmarks and folders that are automatically created" [79]. This change was reflected in the definitions used during the 2006 PIM workshop. Whether or not all of the incidental information studied in this thesis research can be defined as personal information, much of the research in this field is pertinent.

Incidental information privacy is closely tied to personal information management systems. Essential PIM activities include storing information, finding and re-finding information, and maintaining and managing that information (including mappings between information and need) [79]. To illustrate the tie between incidental information privacy and PIM systems, we next discuss the personal information management activities that cause incidental information to be visible within web browser convenience features.

Visited web pages can be considered information items in a personal information management system (i.e., the web browser). If we want to revisit a specific page, we have an information need. The mapping between information and need can be largely internal (e.g.,

our memories) and may have an external representation (e.g., traces appearing in Favorites, Auto Complete, History), part of which can be observed and manipulated (e.g., choice of Favorites name). Some mappings are only potential and not explicit (i.e. a search function is a potential mapping until a specific search is conducted). Incidental information can be generated both through explicit user action (e.g., when information is saved for the purpose of re-visitation, when files are created) and by the PIM system itself (e.g., text stored for use in Auto Complete functions, accessed documents stored for use in the recent documents list). This information may be displayed later either statically by the system for the purpose of initiating user interactions (e.g., icons on the desktop, recent documents list) or dynamically in response to user interactions with applications (e.g., when entering a search term, Auto Complete shows other recently entered terms). It is this display of information that causes visual privacy concerns. In addition to the information pertaining to the task at hand, other information that is incidental to the current task may be displayed. This information may not be appropriate for viewing in a collaborative situation.

Many systems include advanced features to improve recognition of desired information for the end user [82]. These features can be a privacy concern as they increase the visibility of incidental information making it easier for others to see traces of previous activities with casual inspection. Examples include visualizations, such as thumbnails of web pages in history files [82], or an expanded and perhaps annotated search result (as in [37] which includes snippets of text from the retrieved information and additional annotations such as when the information was last accessed and tags applied to it).

The use of search as a method of re-finding information also introduces privacy concerns. Search often makes it easier for users to find information as there is no need to remember precisely how the information was generated or saved. However, search can make it more difficult for users to know precisely what information will appear (as opposed to when navigating through a user defined hierarchy). This problem can be exacerbated in PIM systems that incorporate results across tasks or applications. For example, if email is included in the searched documents, personal emails about difficulties working with another person on a project may be inappropriately revealed when searching for information about the project. One example is Stuff I've Seen [43] which provides a single index for all information that a person has viewed on their computer, regardless of the information type (e.g., email,

URL), and then provides rich contextual cues during the search process including thumbnails, time, and author.

2.3.2 Personal Information Management Styles

An interesting area of PIM research has been the identification of the different styles people use when managing their personal information. Whittaker and Sidner [153] described three styles of email management: no-filers, spring-cleaners, and frequent-filers. No-filers are those who don't use sub-folders, keeping most of their email in their inbox. Spring-cleaners are those who use sub-folders, but who only sporadically file their email (e.g., every 1-3 months). Frequent-filers are those who try to file new email messages into their subfolders on a daily basis.

Gwizdka [54] also studied email task management strategies. During the experiment, 24 participants completed cognitive tests and answered questions about their work habits both in general (e.g., neatness of desk) and with respect to email (e.g., when it's read, searching habits, etc.). Based on their responses, participants were clustered into two groups: the Cleaners and the Keepers. The Cleaners tended to read their email at specific times, not allowing it to interrupt their other tasks. Furthermore, Gwizdka found the Cleaners did not tend to conduct searches in their email and did not use their email to keep track of events or as a to-do lists; however, they did send themselves self-reminder email messages for action when later reading email. The Keepers tended to read email all the time, allowing it to interrupt their other tasks. Gwizdka found these participants tended to conduct searches of their email and used their email as event reminders and to-do lists. They therefore did not need to send themselves self-reminding emails. The only significant differences found between the two groups were that the Cleaners tended to have less email experience and scored low on a cognitive test measuring flexibility of closure.

These different personal information management styles may impact the suitability of visual privacy management approaches for web browsers. It will be important that any privacy management system be viable not only for those users who are willing to constantly maintain it, but also for those who will be more sporadic in their efforts.

2.4 Privacy Management Tools

In addition to understanding incidental information privacy, a key part of this dissertation research is developing tools for helping people manage their visual privacy within web browsers. We next present relevant work related to the design and development of privacy management tools. We first present research from the general field of usable privacy and security related to tool design. We begin with design principles that other researchers have offered for tool design. We then discuss privacy management tools, focusing on those most directly related to visual privacy concerns within web browsers.

2.4.1 Design Principles

Privacy management systems have unique design requirements. Early work by Bellotti and Sellen [15] attributed many of the problems with privacy in media spaces with how the technology changes natural feedback and control mechanisms for the release of information. With the introduction of technology into an environment, it is often unclear what information is being captured, conveyed to others, and how that information may be used. The authors propose that systems must be explicitly designed to provide the feedback and control mechanisms that are lost when not dealing with others on a face to face basis. Similarly, Lau et al. [90] state that privacy interfaces should make it easy to create, inspect, modify, and monitor privacy policies and that the policies should be applied proactively to objects as they are encountered.

De Paula et al. [39] discuss three design principles for enhancing the usability of systems with a security and privacy component (e.g., peer to peer file sharing on a local network, web browsers): providing visualization mechanisms, developing event-based architecture, and integrating configuration of the system with users' actions during normal system use. Visualization mechanisms are important as they allow users to see and understand the consequences of their actions. An event-based architecture affords the visualization of underlying system activities. The integration of configuration of the system and actions during normal system use (e.g., not having a separate control panel for preference setting) brings together users' expression of their privacy preferences and the environment in which those preferences are invoked. These principles are intended to create conditions whereby users can not only recognize privacy and security issues as they arise, but

also understand the issues well enough to make informed decisions and take appropriate actions.

Dourish et al. [42] examined the everyday security concerns of twenty participants through interviews. They found that decisions about security were often a practical problem to be overcome before a primary task could be accomplished. They conclude that it can be difficult for users to specify security needs ahead of time as needs are contextualized by the specifics of the usage situation. This context includes the people, information, activities, and other aspects such as physical, social, and organizational considerations.

One key problem discussed by De Paula et al. [39] is that the traditional goal of reducing complexity in interfaces by hiding system complexity can lead to users being unaware of the privacy and security implications of their actions. Additionally, there is often a disconnect between configuration of the system and the interface where information is shared. Web browsers were used as a test-bed to demonstrate how visualizing network activity could provide users with an understanding of security concerns such as the use of off-site images to maintain records of visitor activity. A similar approach may be useful to help people understand the traces of activity that are stored within web browser convenience features.

Lederer et al. [91] discuss how users should be able to maintain personal privacy through *understanding* and *action*. Understanding is required so that users are aware of potential privacy violations. Opportunities for action are required so that users can appropriately manage their privacy when necessary. The authors identified five pitfalls for designers of systems with personal privacy implications. Four of these pitfalls are applicable to visual privacy in web browsers: obscuring potential information flow, emphasizing configuration over action, lacking the option for coarse grained control, and inhibiting existing practice. The fifth pitfall, obscuring actual information flow, is not an issue as incidental information is transferred visually so the information flow is apparent. The authors make the point that unless the first pitfall is avoided (i.e., users can readily determine the nature and extent of potential information disclosure), users will not be able to fully understand the privacy implications as a result of system use. For the visual privacy of information within web browsers, the information which may be disclosed is limited to recent page visits and data entry in forms. Which traces of prior activity may be disclosed

depend on the convenience feature settings and any preventative actions a user may take when they know their display will be viewed.

Lederer et al.'s [91] remaining three pitfalls relate to privacy preserving actions. Users should not have to extensively configure a system a priori in order to maintain privacy, but rather should be able to manage privacy within their normal interaction with the system. Additionally, their normal interaction with the system should not be hampered by the actions they must take to preserve privacy, nor should their normal mechanisms of preserving privacy (e.g. taking advantage of plausible deniability) be hampered by the technology. Furthermore, users should be able to quickly stop the release of information (i.e. have mechanisms of coarse-grained control) so that they can respond to unanticipated or quickly changing situations of use. For incidental information within web browsers, beyond stopping the release of information (i.e. filtering the content appropriately), it is also important to allow users to easily limit which content is recorded.

2.4.2 Tools for Managing Privacy

The pitfalls that Lederer et al. [91] discuss arose from their evaluation of Faces, a privacy management tool for specifying privacy preferences in a ubiquitous computing environment. Faces allowed users to assign preferences for the granularity of the information disclosed (identity, location, activity, nearby people) by specifying faces (i.e. a persona they wanted to maintain) for specific inquirers given a specific situation (e.g. location, activity, time, nearby people). Wildcards were used to specify a face for an unknown inquirer or when the user's conditions did not meet a specified situation. The granularity of the information was specified at one of four levels: undisclosed, vague, approximate, and precise. For example, a user could specify an "anonymous" face to be used when an inquirer was not known. Results in an evaluation with five participants found that the faces that participants specified a priori were often different from their disclosure preferences given a contextualized scenario. Despite having conducted contextual studies into the privacy preferences for location disclosure (as discussed in [93]), Lederer et al.'s solution (which depended on configuration outside the context of use) was not found to be viable. Participants had difficulty with the indirection the system required (i.e., specifying a face outside the context of the situation in which it applied).

Berry et al. [16] presented an approach for managing visual privacy during presentations. Rather than visual privacy on a single display, they investigated the case where there was a public view which was projected to an audience and a private view that a presenter could see. The authors took a role-based approach to enable privacy in shared views of applications such as Internet Explorer (IE) and to allow protection of objects within documents. For example, in the public view of an IE window, the Auto Complete options for URLs could be masked, while the presenter retained full functionality of this feature in the private view.

Tarasewich et al. [144] developed web browser privacy blinds for use when browsing is conducted on displays that may be visible to others. Rather than intentional sharing of a display, they focused on those occasions when a personal display could be viewed in a public area. Their privacy blinds occlude selected data items (e.g. monetary amounts, email addresses, user-specified phrases). This approach provides visual privacy of select content within a web page, but does not protect the privacy of traces of previous activity at the browser level. As the mask is visible to both the user and viewers of the display (unlike in [16]), using such a mask would preclude use of the convenience feature for navigation.

COLLABCLIO [90] is a research system developed to support automated electronic sharing of web browsing histories in a company setting. While this is different than preserving visual privacy in a co-located setting, the techniques examined are relevant to our work. COLLABCLIO provides users with a binary classification scheme (public/private) that allows them to indicate which visited URLs should be shared with others. The users of this system expressed a wish for finer-grained classification to reflect differing privacy needs for sub-groups of people.

While there are commercial products that allow the erasure of traces of browsing activities, those traces are often valuable for future transactions and may decrease productivity if removed entirely. As an example, WebRoot Software's Window Washer [4] allows a user to delete artifacts such as auto completions, histories, and recent documents. However, with the exception of the ability to save selected cookies, the decision to erase a class of traces erases all instances indiscriminately.

2.5 Summary

This chapter has presented related privacy research with respect to prior privacy theory (section 2.1.1), research investigating privacy concerns for other domains (section 2.1.2), and research developing models of privacy (section 2.1.3). Prior research has found that privacy concerns are highly individual and contextual. Much of this research gave us insight into how we may expect users visual privacy concerns to vary depending on the situations of viewing. Table 1 summarizes several factors of incidental information privacy that we believe may directly impact a user's *privacy comfort level* in a given viewing situation: 1) their *inherent privacy concerns*, 2) their *level of control* retained, 3) their *relationship to the viewer* of the display, and 4) the *sensitivity of potentially visible content*.

Table 1. Prior literature incorporated into our identification of the primary factors of visual privacy for the incidental information found within web browsers.

	Sensitivity of Potentially Visible Content	**Relationship to the Viewer**	**Level of Control Retained**	**Inherent Privacy Concerns**
Section 2.1.1 Privacy Theory	Margulis [99] Phillips [129]	Goffman [49] Palen & Dourish [119]	Boyle & Greenberg [19] Westin [150] Lederer et al. [91]	Phillips [130]
Section 2.1.2 Research Investigating Privacy Concerns	Ackerman et al. [8] Hutchings & Pierce [73] Olson et al. [115] Palen [118]	Cadiz & Gupta [22] Greenberg [52] Huang & Mynatt [71] Hutchings & Pierce [73] Patil & Kobsa [127] Lederer et al. [93] Olson et al. [115]		
Section 2.1.3 Privacy Models	Adams [11] Malhotra et al. [98] Sheehan [137] Spiekermann et al. [139]	Adams [11]	Malhotra et al. [98] Sheehan [137]	Malhotra et al. [98] P&AB [116] Patil & Kobsa [127] Sheehan [137] Spiekermann et al. [139]

The concepts of sensitivity of potentially visible content, relationship to the viewer, and inherent privacy concerns are likely similar between incidental information privacy and other privacy domains. However, while prior research has investigated the level of control retained over the transmission, use, or retention of data, there is no similar component of visual privacy. People may, however, attempt to control which information becomes visible

during collaboration. Therefore, when we refer to level of control in this research, it is with respect to control over input devices such as the keyboard and mouse.

None of the prior literature emphasized all of these privacy factors, but we hypothesize that all of them may be pertinent to privacy concerns during the viewing of incidental information. While prior research lends insight into the factors that we may expect to impact privacy concerns, given the highly contextual nature of privacy, it is unclear exactly how privacy concerns for the visual privacy of incidental information may vary from privacy concerns identified in other domains such as on-line privacy or electronic information sharing. Within each privacy domain that has been investigated there is a specific set of situations that generate information and a specific set of circumstances under which information is viewed or received. Furthermore, within each privacy domain, the specific nature of the visible information, the viewer relationships, the amount of control over the information, and the impact of privacy violations may vary. For example, a common concept is that individuals have inherent privacy concerns, but privacy segmentation models developed in one domain (e.g., the Westin-Harris privacy segmentation model [116]) have not been found to generalize well across domains. It was clear that study of users' specific privacy concerns within the domain of visual privacy within web browsers was required before we could begin to develop privacy management systems.

Furthermore, it is important to consider privacy management within the context of the primary task of users, browsing the web. The related work presented from the areas of web browsing behaviour and personal information management gave us some perspective as to the issues that must be considered when managing privacy in the web browser. In a web browser, the specific content that may be visible depends upon recent *browsing activity*, *browser settings*, and any *preventative actions* taken. Additionally, the context (i.e. location, device) of the browsing activities and viewing opportunities may impact web browsing behaviours and privacy concerns.

Our exploratory research, presented next in Chapter 3, was designed to give us an understanding of the specific visual privacy issues within web browsers and the web browsing behaviours which will constrain the design space of potential solutions.

Chapter 3
Exploratory Studies

In this chapter, we present the methodologies used for our exploratory research investigating the visual privacy of incidental information. All research methodologies have inherent flaws and benefits in terms of the ability to generalize results, measure behaviours and attitudes precisely, control confounding factors, and conduct the research within a realistic context [103]. This chapter begins with a discussion of the research methodologies suitable for studying privacy issues and web browsing behaviours. We then present our chosen mixed methodology approach of a survey and two field studies, giving details of our participants, procedures, data collection, and analysis techniques. A reflection on the suitability of our methodological choices is given in Chapter 9.

3.1 Research Methodologies for Studying Privacy

Privacy is a challenging area to study as privacy concerns vary on an individual basis and can be difficult to invoke in a controlled environment. Recently. workshops such as the Privacy and HCI: Methodologies for Studying Privacy Issues workshop at CHI 2006 and the Security User Studies workshop at SOUPS 2006 have focused on these challenges. We next discuss the suitability of various research methodologies for studying privacy.

3.1.1 Surveys

Survey research is popular as surveys are relatively easy to develop, administer, and analyze. While a carefully sampled survey may increase ability to generalize results, a survey is limited to measurement of self-reported attitudes and behaviours. This can be particularly troublesome with the sensitive nature of privacy research as the attitudes and behaviours reported by participants may be skewed due to participants' tendency to give socially desirable responses [103]. Attitudes may also be impacted by situational and cultural relativities [29]; for example, recent events (e.g., a privacy violation) can temporarily heighten sensitivity.

There is often a difference between responses on attitudinal surveys and the actual privacy preserving behaviours observed [10]. Attitudinal surveys may measure an ideal

privacy standard; however, in practice privacy issues are not as straight forward. Users must weigh the costs and risks of releasing information with the potential benefits (e.g., personalized interactions). It is important to determine under which contexts idealized privacy concerns may be altered. Surveys may be best suited to evaluate attitudes (e.g., privacy concerns) and can be used as a baseline with which to compare actual behaviour [77].

3.1.2 Laboratory Studies

Laboratory studies allow researchers to observe privacy practices in action in a controlled fashion; however, it is difficult to provide a sufficiently realistic experimental setup that will compel participants to engage in normal behaviours. This is particularly challenging in privacy and security research due to the highly personal nature of the data at stake. It can be difficult to motivate participants to make the effort and take the same actions with study data as they would normally take if the data was their own [128, 152]. For instance, three participants in a study of privacy preferences for an awareness application indicated that they set preferences at the team level instead of the group level because it would allow them to finish the study more quickly [128]. Similarly, in a study of the cues that participants view to evaluate the security of a web site, real participant data (e.g., credit card numbers) could not be used and participants had difficulty treating the dummy credit card number with the same care as their own [152].

3.1.3 Field Studies

Field research theoretically allows the study of actual behaviours in a realistic environment. However, the act of observing or recording participants' personal interactions may cause them to alter those behaviours. For example, behaviours deemed to be socially inappropriate may be avoided during the period of the study. This is particularly challenging when studying privacy as those behaviours that invoke privacy concerns may be the behaviours participants are most likely to avoid. As well, participants may be unwilling to have logging software installed that may record personal interactions, particularly if that software logs data across applications (e.g., a keystroke logger may capture passwords). Observational studies with researchers in the field may be well suited to capture high-level information (e.g., task) over short periods of time; however, logged data is necessary to

capture finer-grained details (e.g., speed, frequency, and actions) throughout participants' interactions with technology.

3.2 Studying Web Browsing Behaviour

The study of user behaviour on the Web is also complex and well suited to study in a field environment. Behaviours can be influenced by a number of factors, such as task [87], motivation [96], and individual differences [148] such as domain expertise [70]. Web behavioural studies in a field setting can often provide a more realistic picture of behaviours than can be evoked in a controlled laboratory setting, as the tasks are more likely to be motivated by the users themselves. Furthermore, in the field, participants have access to their usual web tools, browsers, and physical environments.

One common method of studying user behaviour in a field environment is through the collection of logged data. This method can be unobtrusive to the user and provides researchers with details of the user's actions. However, logged data by itself does not provide a full understanding of users' activities, goals, attitudes, and processes. Contextual information plays an important role in how we understand and interpret people's everyday behaviour. Information that provides additional details about people, such as their location or task, can help us better understand and interpret their actions. In a web environment, contextual information can be used to determine the activity in which a user is engaging, their motivations for engaging in that activity, as well as perceptions about the current tool or the information being viewed.

It can be difficult to capture natural web browsing behaviour that is also rich in detail without altering the browsing environment of the participant. The browsing environment includes many factors such as the user's physical location and their usual browser application including all normal settings and features (e.g., user-installed toolbars). There are some logging tools (e.g., browser helper objects), which can work within the participants' normal browsing environment and log data unobtrusively; however, these tools can only record limited types of data (i.e., interactions at the web document level). In order to record richer interactions with the web browser itself, a custom web browser must be used (unless researchers have access to the source code of a commercial browser). Developing a custom

web browser that fully mimics the appearance and functionality of participants' commercial browser applications is challenging.

It is important during studies of natural browsing behaviours that we record specific aspects of context that may be influencing behaviours at the time, and capture those behaviours across all normal usage contexts. Web usage can vary across different locations (e.g., home, work) and devices (laptop, desktop). Additionally, different web browsers or web browser settings may be used in these environments and browsing may be conducted for different purposes (e.g., personal, work-related). Chapter 6 will present results which support these claims.

There are tradeoffs between the ability to capture rich data about browsing activities across all contexts of use, the ability to maintain the participants' normal web browsing environment, and the implementation costs inherent to each data collection methodology (see [67] for a discussion of the costs and benefits of various logging methods). These tradeoffs were carefully examined for each of our field studies. Sections 3.5.3 and 3.6.2 describe the requirements that shaped our choice of logging method.

3.3 Mixed Methodology Approach

We chose to employ a mixed methodology approach of a survey and two field studies to reduce the bias inherent within each approach and to allow triangulation of our results. Our survey was designed to examine privacy concerns related to the incidental viewing of web browsing traces. As the survey can only represent users' self-reported perceptions of their concerns, it was important to build a more complete picture by integrating the results from the survey with results grounded in actual behaviours, as revealed through the field studies. For example, the survey allowed us to present scenarios of web browsing activity and to examine participants' stated privacy comfort levels for varying levels of control and relationships to viewers. In the survey, the potentially visible content presented was limited to scenarios sampled from the breadth of privacy sensitivities (e.g., a scenario of web browsing for information about genital shingles was selected to represent browsing that is very sensitive in nature). In contrast, the field studies allowed us to examine how participants felt in terms of privacy about specific instances of visible content (the web pages they had visited that day) and to examine patterns in the application of privacy levels

to that content. We next present the methodologies employed in the survey and two field studies in more detail, including the participants, the procedures, and the types of data collected.

3.4 Study 1 – Incidental Information Privacy (IIP) Survey

The contextual nature of privacy is well established in the literature (as presented in section 2.1). However, as there is little prior work directly addressing visual privacy concerns, it was unclear exactly which usage contexts would have an impact on visual privacy concerns in web browsers, and the extent and interrelationship of the contextual factors. The IIP survey was designed to explore several factors of incidental information privacy that arise when web browsers are used during co-located collaboration or are used by multiple people without separate logins. The three main objectives of the survey were to 1) determine the scope of the problem, 2) gain an understanding of the type of web browsing activities that are conducted and the physical context of those activities, and 3) measure privacy comfort levels for different contexts of browsing. This survey was available on-line from June 2004 to March 2005.

One limitation of survey research is that participants must reflect upon their attitudes and experiences while not in the context of those experiences. However, in the incidental information domain, current privacy management is largely a matter of speculation: What traces of my past activities will be visible on my monitor? Who will be able to view it? Should I clear my history files? Additionally, people have to speculate about how others would regard these traces of activity that they have conducted in the past. In this regard, a survey was a good choice to explore attitudes and get self-reported data about typical web browsing behaviour and current privacy management practices.

Depending on the privacy domain under study, there can be a huge volume of information items to be considered and many contexts in which the information may be viewed. We elected to use general cases in our survey (e.g., viewer categories such as 'close friend') so as not to burden participants with too many questions, but there is also a need to look at specific instances in order to increase the realism of the scenario. Some researchers (e.g., Olson et al. [115]) have had participants instantiate an attribute (e.g., give the name of a close friend and use that in the questions). However, even an instantiated attribute may not

reflect the spectrum of possible situations. For instance, a participant may consider several people to be close friends, but may not share information with them all equally. Even for a specific person, privacy concerns may fluctuate (e.g., after a disagreement).

The survey was designed with the advice of Maryanne Fisher, a psychology researcher with experience teaching research methodology and statistics. Care was taken when crafting the survey questions to reduce biasing the responses through the use of suggested question formats as presented in survey design literature (e.g., [41, 47]). The survey was refined through several iterations of pilot testing and critiques by researchers in the DVRG, EDGE, and WIFL research groups at Dalhousie University, as well as a fourth year class of Human-Computer Interaction students. Approximately 65 people gave feedback on the survey before the study began. Appendix B contains the final version of the survey.

3.4.1 Participants

Participants (155, 57% male) were recruited from businesses, the university community, and the public through email lists and hand-distribution of notices. As participants were not randomly sampled from the Canadian population of web users, survey participants may not be representative of all web users. Our study population is characterized by a high level of education (median Bachelor's degree) and computer experience (avg. 12 years, 2-35). Most participants were frequent computer users (median 29-35 hours per week) and web users (median 15-21 hours per week). Participants were diverse with respect to age (avg. 31.5, 17-59). A 2005 Statistics Canada report [142] indicates that higher percentages of individuals in younger age groups are internet users (e.g., 88.9% of those aged 18-34 are web users, contrasted with 75.0% of those aged 35-54, 53.8% of those aged 55-64, and 23.8% of those aged 65 years and over) and higher percentages of individuals at higher levels of education are internet users (e.g., 89.4% of those with a university degree are web users, contrasted with 72.0% with a high school or college degree, and 31.2% of those with less than a high school degree). Our study participants may therefore be similar to the general web user population.

While occupations ranged from homemakers to professionals, students were over-represented at 42.6% of the participants. It is unclear whether this over-representation will affect the generalizability of our results. Prior research by Metzger et al. [105] investigating

college web use has found that Internet usage, ability to access the Internet, and familiarity with Internet information were not significantly different between students and non-students despite differences in age, years of education, and income. Furthermore, Flanagin and Metzer [46] investigated the perceived credibility of web-based information and contrasted results by their two sub-groups of participants (those randomly sampled from registered voters in the United States, students in an undergraduate communications course); few differences were found in results between the sub-groups.

3.4.2 Procedure

The on-line survey took about 20 minutes to complete and participants received no compensation. Access was controlled through unique personal identification numbers which were distributed to participants with an information letter which contained the URL for the survey. Submission of the completed survey was taken as an indication that participants had read the explanation about the study and had consented to take part in it. The on-line nature of the survey was suggested by Dalhousie's Research Ethics Board as a way to ensure that participants recruited through businesses would be able to participate in the study while away from the workplace without fear of employers learning of their privacy concerns. Mode effects (e.g., elevated responses on ratings scales) between paper- and web-based surveys are generally minimal; however, responses to questions on web-based surveys that deal with computing and information technology can be more positive [24]. Due to the survey's technology focus, we did not want these effects to impact our results; therefore, it was only made available on-line. A benefit of having the survey available on-line was that it allowed interested participants to complete the survey on their own time, in a place of their choosing, and may therefore have promoted more honest responses for questions of a sensitive nature [147].

3.4.3 Data Collection

The survey was written in Perl and CGI. Responses were stored in a password-protected MySQL database, located on a server managed by the Faculty of Computer Science at Dalhousie University. Results were retrieved with a web-based script written in Perl and CGI.

After entering their PIN, survey participants were asked to specify their primary location of web browsing (work, school, home) and the primary computer that they use in this location. We hoped to use this information to determine whether or not participants with different primary usage environments had different privacy concerns. We then asked a series of demographic questions as well as questions exploring general web browsing behaviours. Questions probed web browsing and computer usage at home and away from home, as well as the types of browsing activities in which the participants engage, where those activities take place, and on which types of computers.

The next series of questions examined the general scope of privacy issues participants have related to the incidental information that may be visible in web browsers. These included the *frequency* with which ten different types of people (both interpersonal and business/school relationships) might *view* or *use* a participant's computer. Participants were asked to think about who can clearly *see* the contents of their screen as they use it and approximately how often they may be in that situation. Similarly we asked who might subsequently *use* their computer and the frequency of that use. A five point scale for frequency was used (ordered as daily, weekly, monthly, rarely, never).

The next section of the IIP survey was designed to investigate how specific contexts (e.g., sensitivity of content, type of viewer, level of control retained) affect privacy concerns. Rather than examining privacy comfort for all types and sensitivities of traces, privacy comfort was examined for three different levels of content sensitivity through scenarios. The three scenarios were explicit descriptions of hypothetical web browsing activities and their order of presentation was counter-balanced. The scenarios were designed to discover the range of comfort a participant had for information of varying sensitivity. All scenarios discussed a situation that led to information seeking behaviour on a web browser and described a set of search topics and web page visits that might be revealed during a future web browsing episode. The scenarios were contrived to be universally 1) *embarrassing* (genital shingles), 2) *neutral* (buying a car), and 3) *positive* (winning a trip). The embarrassing scenario (Table 2) was designed to be extremely sensitive in content, but with no judgment on the morality of the activity.

Table 2. The embarrassing web browsing scenario.

You have been experiencing itching and pain in your groin area. You go see the doctor who unfortunately diagnosed you with shingles on the genitals. Shingles can occur in people who have previously had chicken pox. It is a very painful disease. You have been experiencing uncomfortable symptoms and have been looking for relief. You use your web browser to search for such topics as "burning genitals" and "itching groin" and have visited such web pages as www.yoursexualhealth.com/stoptheburning.html and www.genitalhealthcare.com/topics/infectiousdiseases (which you add to your favorites for future reference).

After reading each scenario, participants were asked to think about "*how comfortable* [a situation] *makes you feel in terms of privacy.*" Participants rated their privacy comfort level using a seven point scale (ranging from extremely uncomfortable (1), to neutral (4), to extremely comfortable (7)). For each of the three scenarios, participants were asked to give a privacy comfort level for five types of potential viewers (close friend, supervisor, parent, spouse/significant other, colleague/fellow student) for each of three levels of control over input devices. The levels of control specified were as follows: if the participant was the one in control of the web browser (you), if the viewer was in control of the web browser with the participant sitting right there (other), or if the viewer was in control of the web browser and the participant left the room (away). Therefore, a total of fifteen privacy comfort levels (i.e., one for each of the 5 viewer types x 3 levels of control) were recorded for each of the three scenarios. After answering questions about their privacy comfort levels according to viewer and amount of control for each of the three scenarios describing hypothetical browsing activities, participants were asked to do the same exercise in a scenario which had them reflect on their *usual web browsing behaviour*.

Participants were also asked to reflect upon how they currently handle the tradeoff between convenience and privacy in their web browsers. They indicated their current settings for their History, Auto Complete, and Favorites. They also reported the actions they would take if given advanced warning that somebody else would be working closely with them as they used their web browser and could see their display. Finally, they were asked to give an optional example of a situation where incidental information privacy was a concern.

3.5 Study 2 – Privacy Gradients 1 (PG1)

A field study was conducted in August 2004 to examine how individuals perceive the privacy of their web browsing activity if others can view traces of it later. The study was

conducted over the course of a week to capture normal web browsing behavior as much as possible. We selected a one week period in order to capture the full cycle of participants' normal web browsing behaviour (e.g., including a weekend).

Privacy is a complex issue with both privacy concerns and willingness to maintain a management scheme varying on an individual basis. However, we believed that people would be willing to organize their information across a small number of privacy levels or gradients. The privacy comfort levels of participants in the IIP survey were measured using a 7-point scale to allow participants to report the nuanced changes in their privacy comfort given different contexts of viewing. However, we felt that a similarly highly nuanced decision process may be overly complex for participants in the field study who would be required to evaluate the privacy level for each of the sites visited during their web browsing. A four-tier privacy scheme was proposed to see if that level of granularity was appropriate to allow participants to effectively express their privacy concerns for their web browsing activity while not requiring a great deal of mental effort to distinguish between the different levels.

3.5.1 Privacy Gradients

To facilitate classification of visited websites, a common terminology was required. The four-tier privacy gradient scheme used was *public*, *semi-public*, *private*, or *don't save* (see Figure 4). If a site needs to be accessed again, traces of it should appear in the browser convenience features; and these traces should be stored with some associated privacy level. *Public* sites are those someone is comfortable with anybody and everybody viewing, including the Queen of England (hence the crown in Figure 4). *Private* sites are those a person would be comfortable with only themselves and possibly a couple of close confidants viewing. *Semi-public* sites fall somewhere in between: depending on the context of the viewing, pages may or may not be appropriate. Web sites classified as *don't save* primarily fall into one of two categories: ones that are irrelevant (i.e. the first 17 pages of a search before finding a page) or ones that are so private it is preferred that there is no record of them at all.

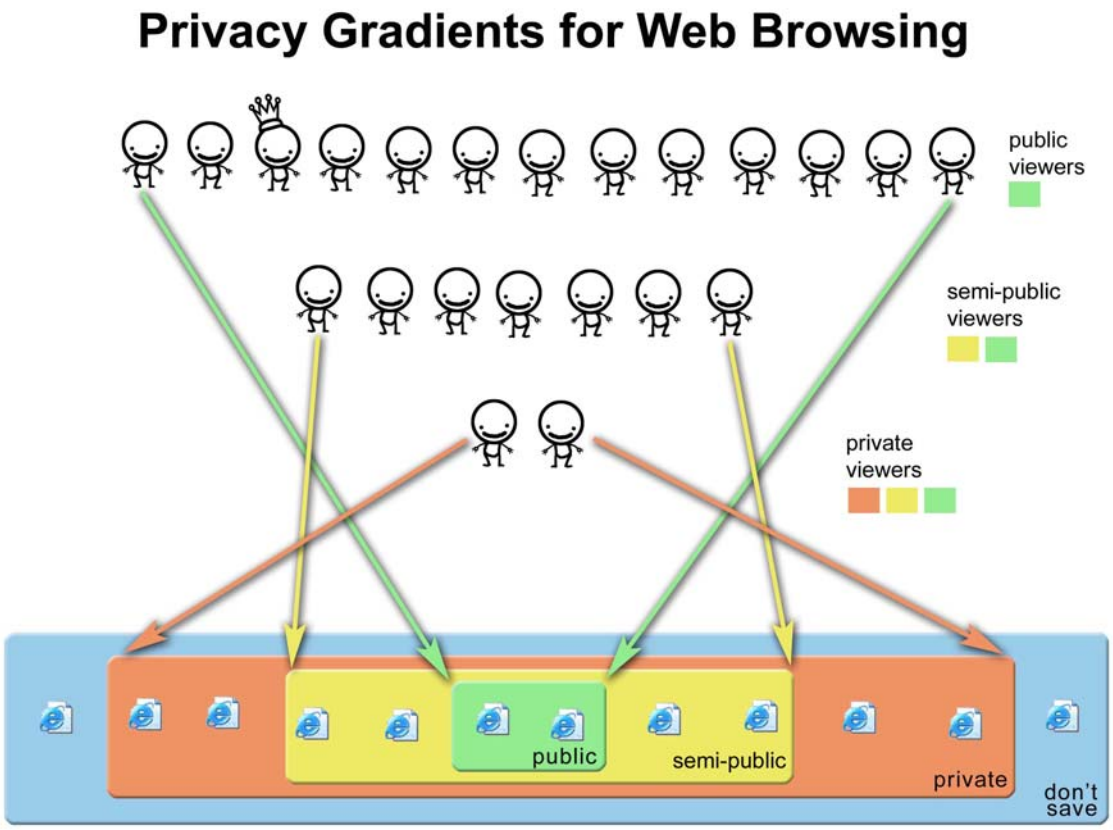


Figure 4. Diagram conveying the four-tier privacy level scheme, used by participants when classifying categories of web sites during the field studies.

3.5.2 Participants

Participants were recruited from the general university community. To qualify for inclusion, participants needed to be IE users and perform the majority of their web browsing on a laptop computer so that we could capture the majority of their personal and work/school related web browsing as they moved between physical locations. Participants also needed to have had occasions in the past where their web browser window was visible by others, so that the concept of privacy in this situation had some relevance. Participants had to be willing to have a logging program installed on their laptop to record their web browsing for the period of one week and to complete daily diaries recording the privacy levels of the web browsing done that day. The also had to agree to come to the Faculty of Computer Science and complete pre-study questionnaires for approximately 30 minutes as the logging program was installed and, at the end of the week, come back to complete the post-study questionnaires and have the data transferred and logging program uninstalled.

Twenty participants, age 19-47, took part in the study (16 males, 4 females). Participants were highly educated, with 65% having completed at least an undergraduate degree in primarily technical fields (14 Computer Science, 4 Science). There were eighteen students, one professor, and an Information Technology professional. Participants were generally experienced computer users (median 10 years) and spent a considerable amount of time each week using their computer (median 29-35 hrs/wk) and web browsers (median 22-28 hrs/wk). On average, they reported usually spending 48% of their time browsing for personal reasons, 16% for work reasons, and 35% for educational reasons.

Participants in the study represented a fairly homogenous group: highly educated, predominantly male, laptop users. This sample is similar in construct to those used in earlier related research, so comparisons with previous web browsing behaviour results may be valid. However, this group is not representative of the overall web browsing population; therefore, the external validity of these results is limited.

3.5.3 Data Collection

3.5.3.1 Challenges

We wanted to collect both quantitative and qualitative measures of web browsing behaviour. The quantitative data we wanted to capture consisted of a record of the web page visits, including the date/time stamp, page title, and URL. In order to investigate patterns that may occur on a per window basis, the browser window in which the page visit occurred was also required. The qualitative data consisted of participants' perceived privacy of their web usage. Standard logging tools did not support our data collection requirements. Although several research and commercial logging tools record visited page data, none include the browser window ID. We therefore had to develop two client-side data collection tools: one to log users' web activities and the other to allow participants to annotate their web activity with a privacy rating.

The design of the data collection tools presented several challenges. First, we needed to explore normal web browsing activities to see if privacy patterns existed. Therefore, it was important that the experimental software not interrupt the flow of participants' web browsing [26]. Second, we wanted to maintain the participant's normal web browsing environment (i.e. their usual web browser with all convenience features and settings intact).

Finally, we were also concerned about participants' privacy; we did not want the recording of the sites visited to impact their normal web browsing activity (i.e., we wanted participants to visit websites as they normally would, regardless of the social desirability of the content).

3.5.3.2 Solutions

The ability to maintain participant privacy (recording data locally) and to gather rich information about user activity on a per-window basis led us to a client-side solution. To record the browsing activity of participants, a browser helper object (BHO) was developed to work with IE. A BHO is a .dll file that loads every time IE loads. As each IE window opens, the BHO loads and logs all web sites visited until the window closes. For this study, the visited web page (URL and page title), time stamp, and ID number of the browser window were recorded. All pages viewed in the browsing process were logged, even if navigation continued before the document fully loaded. Individual frames or images loaded within a web document were not logged, just the complete document. An advantage of the BHO was that the users' browsing environment did not change; they continued using IE with their normal settings intact.

An electronic diary was designed and developed to allow participants to assign privacy gradients to their web browsing on a daily basis (see Figure 5). The diary displayed all the logged data and required participants to indicate how they would classify the privacy level of each web page they visited if others were able to view the history of this activity later. Participants could annotate individual entries with a privacy level or select multiple entries for annotation. The entries could be sorted by any field (time, URL, page title), allowing participants to easily classify groups of page visits (e.g., repeated visits to the same site). Participants could modify a previous privacy annotation by re-selecting the entry and selecting a new privacy level. We chose this intermittent approach to classification as we did not want to impact the flow of participants' browsing as it occurred. Retrospective reflection on the appropriateness of our methodological choices is provided in Chapter 9.

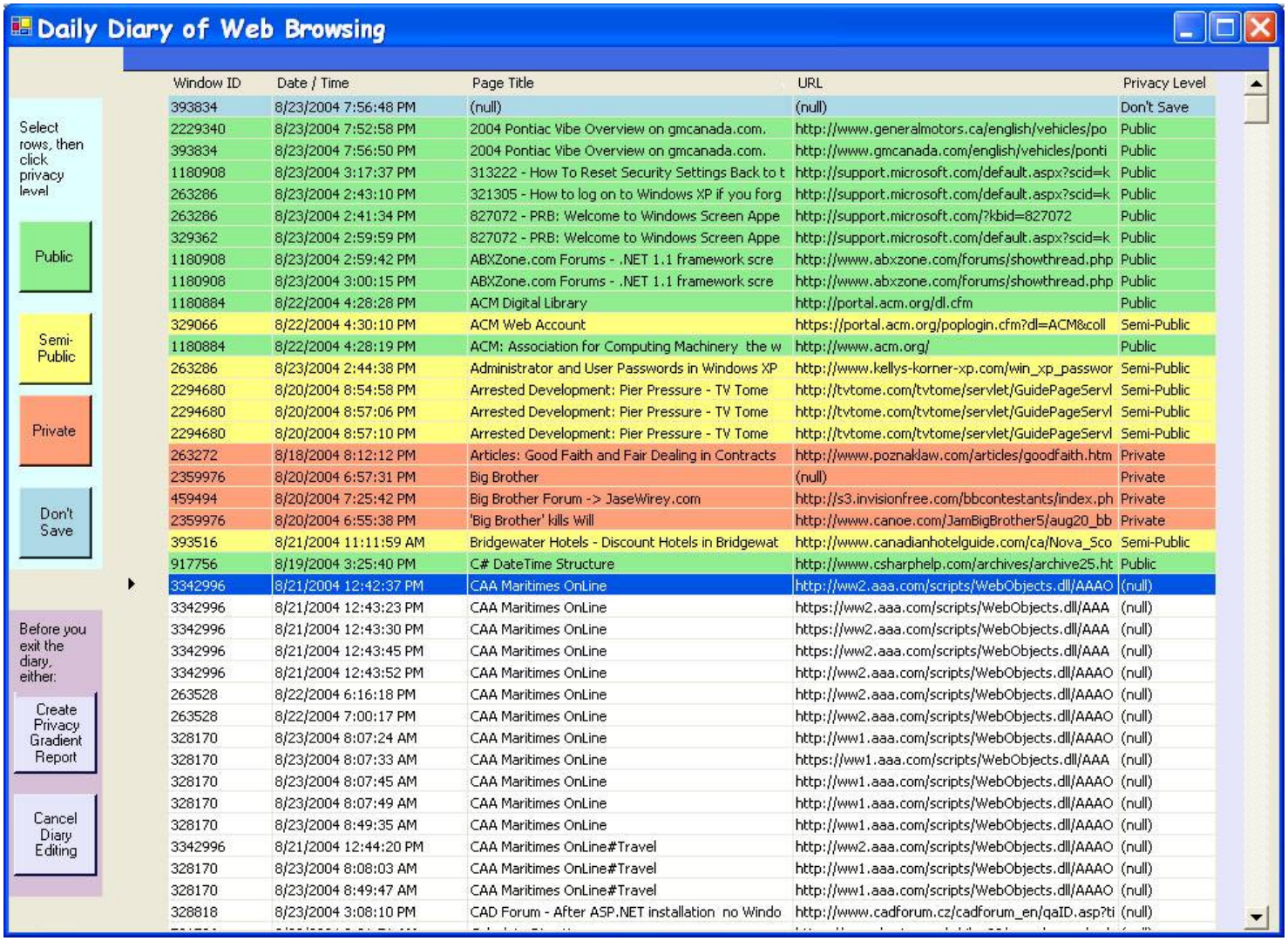


Figure 5. Screenshot of the electronic diary participants used in PG1 to annotate their web browsing with a privacy level (mock data).

After classifying their browsing activity with a privacy level in the electronic diary, participants generated a report to inspect and email to the researchers (Figure 6). In this report, the viewing history was sanitized so that the URL and page title were eliminated. While it was hoped that this approach to maintaining privacy would contribute to participants' willingness to engage in their usual browsing activities, the lack of URL information meant that the number of unique web sites visited or the extent of site re-visitation is unknown. Although the data being sent was visible for inspection by the participant, they were unable to edit the generated report.

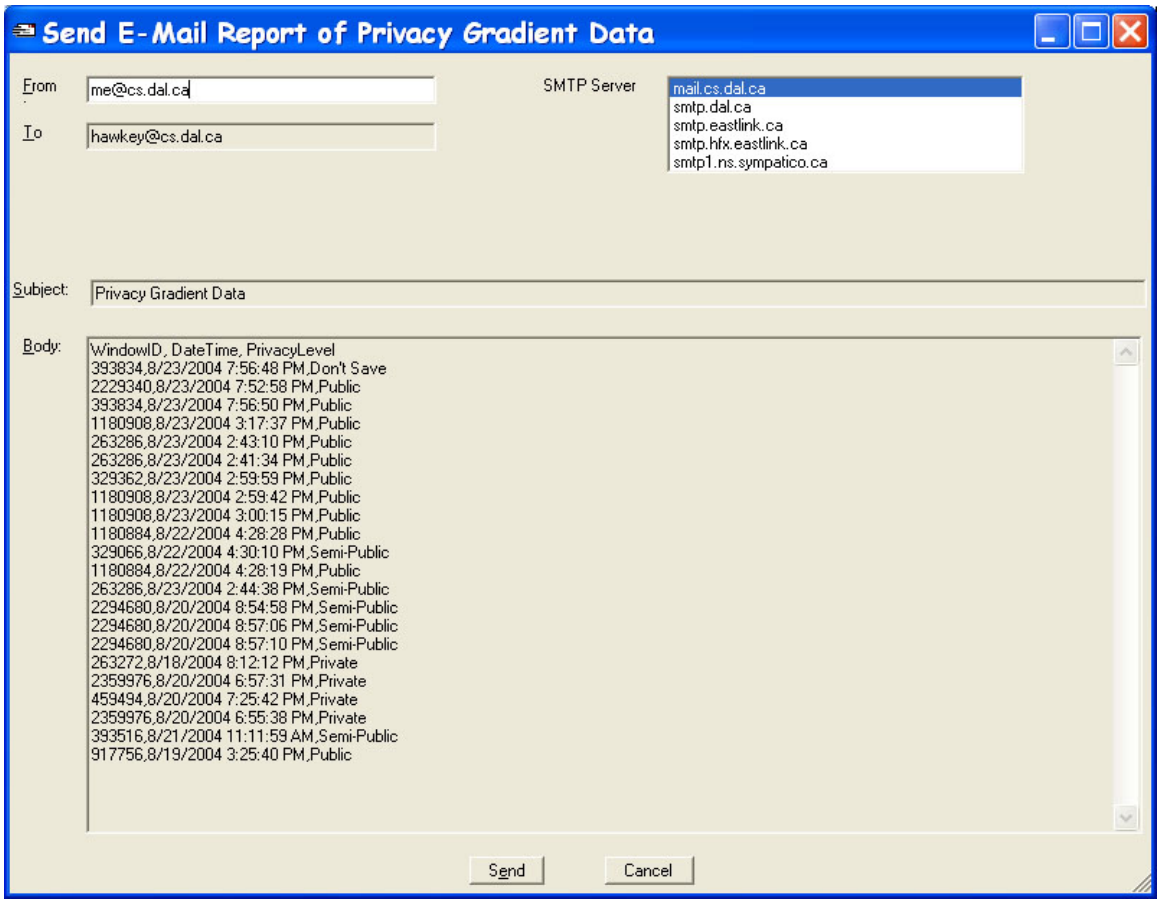


Figure 6. Screenshot of email generated by the electronic diary, showing sanitized data sent to researchers (mock data).

The sanitized report received from participants consisted of a browser window ID, date/time stamp, and privacy level. The browser window ID allowed us to examine browsing activities on a per-browser window basis, while the date/time stamp allowed us to investigate temporal patterns in the data. Based on this information, general web browsing

behaviour was examined, including the number of web pages visited, and the number of browser windows utilized (sections 4.1-4.2). This data was further analyzed to find temporal browsing patterns including bursts of activities, sessions, and transitions between browser windows (sections 4.2-4.4). The privacy levels assigned were analyzed to find patterns in participants' application of privacy levels to their visited pages, including patterns within a browser window such as streaks of two or more pages at a privacy level (section 5.2).

3.5.4 Procedure

Participants completed three study components: the install session, the field study, and the uninstall session. At the end of the study, participants were given a $50 honorarium. After participants had successfully been screened for inclusion in the study, the install session was scheduled at the EDGE Lab in the Computer Science Building at Dalhousie.

The install session took approximately 30 minutes. Participants first read and signed the informed consent form. Participants were then given a description of the study and introduced to the privacy gradients scheme (public, semi-public, private, don't save; see section 3.5.1). As the logging application and electronic diary were installed and tested on their laptops, participants completed a subset of the questionnaires used in the IIP survey, including demographic and background information, the frequency of various types of viewers/users of their laptop, their general privacy comfort level when this viewing occurred, their current browser convenience settings, and their privacy management strategies. Additionally, two theoretical privacy classification tasks were given to participants. The first task asked them to classify the privacy of categories of 55 websites (based on content) into the four levels (public, semi-public, private, and don't save). The web-site categories (e.g., online games, news/media) and their descriptions were based upon those used in commercial products to filter and block internet content [1]. The second task asked them to classify categories of viewers at one of three levels: allowed to only view pages classified as public, allowed to view pages classified as both public and semi-public, and allowed to view all visited pages. Appendix C includes all the questionnaires administered to participants during the install session. Participants were then shown how to use the electronic diary, and an uninstall session was scheduled for the following week.

As described in section 3.5.3, during the field study component, the logging application automatically tracked all the web pages the participants visited (URL and title), the time they visited them, and which browser window the pages appeared in. Participants were asked to fill out the electronic diary generated from the logs of their browsing activity on a daily basis. For each web site, they were asked to classify it as being at one of four privacy levels. They then generated a report that removed the web site URLs and titles from the collected data and were given an opportunity to inspect this report before sending it via email to the researcher. The researcher followed up with any participants that had not sent in a report for two days to make sure that the delay was not due to problems with the software.

At the end of the week, participants returned for the uninstall session, which took approximately 30 minutes to complete. Prior to this visit, we checked that the install session questionnaires were filled out correctly and that all browsing data had been received. Participants were asked (if necessary) to clarify questionnaire responses and classify any remaining data that remained in the electronic diary. As the software was being uninstalled, participants again completed the two classification tasks (i.e., web page classification, viewer classification) as well as the privacy background questionnaire. We were interested in whether participants' responses changed after reflecting on their incidental information privacy during the course of the field study. They were also given a questionnaire about the four-level privacy scheme used during this study. Appendix C includes all questionnaires.

3.6 Study 3 – Privacy Gradients 2 (PG2)

A second field study, PG2, was conducted in March 2005 to extend our understanding of visual privacy concerns within the context web browsing activity. In PG2, we gathered additional contextual information about regular web browsing activity such as the page title, URL, and location of the browsing. This data enabled examination of the relationship between the context of the browsing activity (location, page content) and the privacy comfort levels that participants applied to their web browsing.

3.6.1 Participants

Participants in the PG1 field study consisted solely of laptop users; post hoc analysis of their demographics revealed that they were primarily male with a technical background. The second field study was designed to include participants with varying technical experience

and computers in use. Three different classes of participants were recruited: technical desktop users, non-technical desktop users, and non-technical laptop users. A screening process assessed participants' technical background and identified computers on which they conducted their web browsing. Participants were classified as technical if they had formal training in computer technology or were employed in a technical capacity (e.g. web master). Given our small sample size, no statistical comparisons between subjects will be made; privacy is a domain known for individual variability and participants within each group were not balanced by dispositional factors such as age, sex, or computer experience.

Participants were recruited from the general university community. Fifteen people, age 18-44 (avg. 27.8), took part in the study (5 males, 10 females) (see Table 3 for the demographic breakdown of recruited groups of participants). Participants were highly educated. Eleven participants were students and four were office or administrative staff. Participants were generally experienced computer users (avg. 9.7 years, 6-20) and spent a considerable amount of time each week using their computer (median 29-35 hrs/wk) and web browsers (median 15-21 hrs/wk). On average, they reported spending 37% of their time browsing for personal reasons, 18% for work reasons, and 45% for educational reasons.

Table 3. Demographic breakdown of recruited groups of participants in PG2.

	Overall	Non-technical desktop	Non-technical laptop	Technical desktop
Age	27.8 (18-44)	27.8 (18-40)	22.8 (18-30)	31.2 (25-44)
Sex	5 M, 10 F	1 M, 4 F	1 M, 4 F	3 M, 2 F
Occupation	11 students 4 office staff	3 students 2 office staff	5 students	3 students 2 office staff
Computer Experience	9.7 yrs. avg. (6-20)	8.0 yrs. avg. (6-10)	11.2 yrs. avg. (6-15)	10.0 yrs. avg. (6-20)
Usual reasons for browsing	37% personal 18% work 45% school	31% personal 30% work 39% school	39% personal 3% work 58% school	42% personal 22% work 36% school

As discussed in section 3.4.1, while our participants were more highly educated than the general public and many were students, these are characteristics of web users in general; our results may therefore not be as limited in terms of generalizability as if we were attempting to represent the overall population. However, given that participants were recruited from an educational domain, browsing activities may include more educational and reference sites than if participants were from another domain.

3.6.2 Data Collection

3.6.2.1 Challenges

For the PG2 study, we had increased concerns about the data collection changing participants' normal browsing activities. We needed to receive additional data so that we could examine the impact of context (location, visited page) on privacy concerns. We therefore needed to not only collect the URL and page title for use by participants within the electronic diary, but to also receive that information as part of the generated report. As we did not want our receipt of this additional information to impact participants' willingness to visit sensitive sites, we decided to provide participants with the ability to selectively blind any sensitive data contained in the URL and page title.

3.6.2.2 Solutions

Quantitative data collected consisted of date/time stamp, page title and URL of visited pages the browser window ID, and location of browsing. The BHO used in PG1 (see 3.5.3.2 for details) was modified to record the additional location information. Participants' location was hard coded into the BHO installed on desktop computers. Laptop users indicated their current location with a radio button that appeared in a form as the browser window closed; options were home, work, school, and other (a text box was provided for entry of the specific location). Additionally, the BHO was modified to record window events (focus, open, close) so that we could determine when participants moved between windows, not just when they moved between windows for the purpose of navigating to a new page.

The electronic diary was modified to allow participants to sanitize entries in the diary by removing the page title and URL after applying a privacy level (see Figure 7). Participants were asked to give a general reason for the sanitized browsing (e.g., "looking for medical information"); the default label was "no reason given". After classification, participants generated a report to email to the researchers. The report was similar to the one for PG1 (Figure 6), but also included the page title and URL information for each visited web page. It was hoped that the privacy afforded by participants' ability to selectively sanitize their browsing record would contribute to their willingness to engage in normal web activities while still providing us with context for most visited pages.

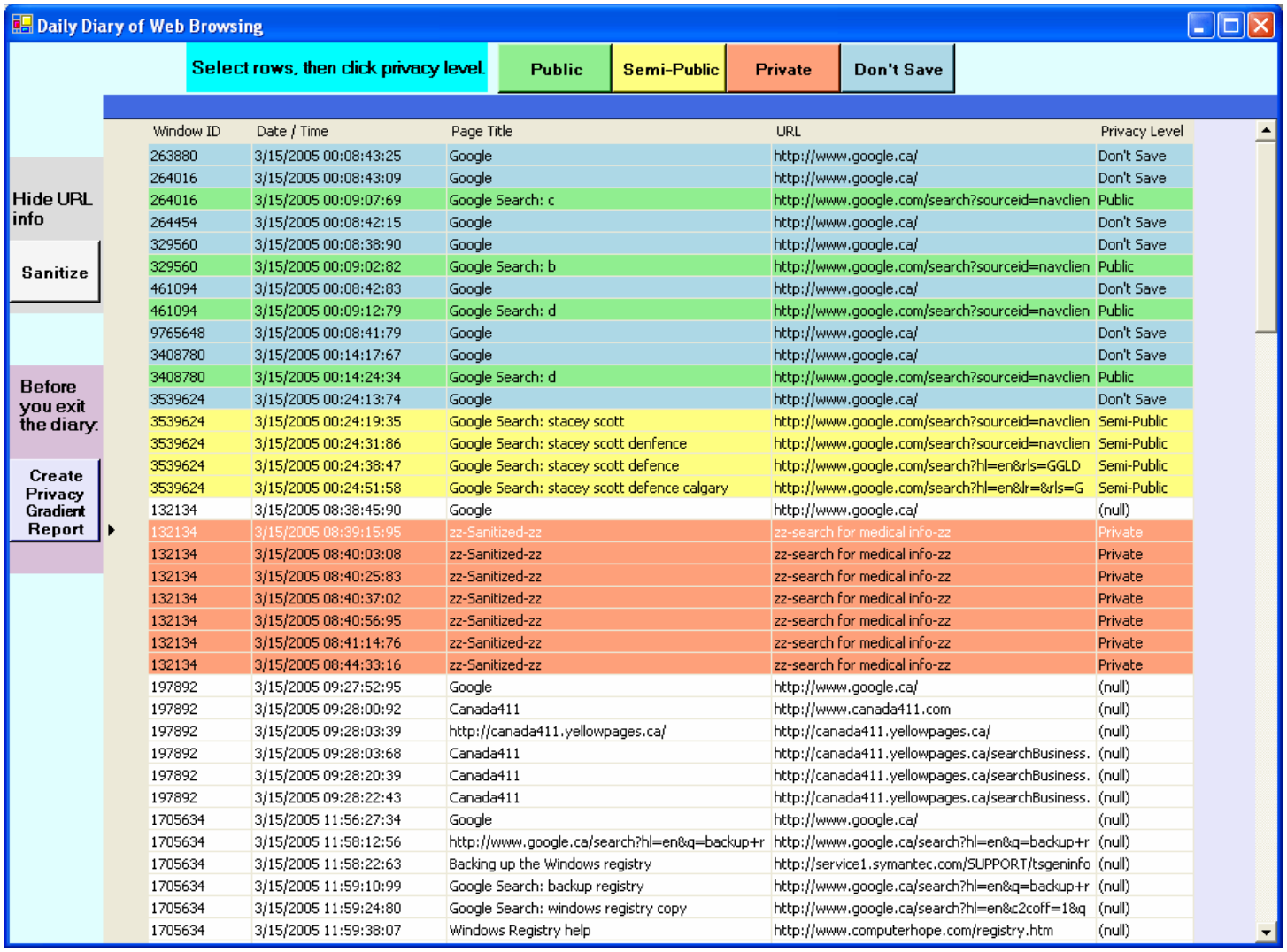


Figure 7. Screenshot of electronic diary used in PG2 for participant annotation of web browsing with a privacy level.

3.6.3 Procedure

The procedure mirrored that in PG1 with a few exceptions. For the laptop participants, the installation session was virtually identical. The only difference was that rather than solely relying on participants' self-reports of their browser settings through the privacy background questionnaire, we also made note of their actual settings as the installation was completed. For the desktop participants, the software was installed on their desktop computers located in their normal browsing environments (e.g. home, work, school). For desktop participants with multiple computers, the informed consent, install session questionnaires, and demonstration of study software were completed during the installation of software on the first computer. An appointment was also made to install the software on the secondary computer, but no questionnaires were completed by participants at this time. A version of the privacy background questionnaire was created to reflect desktop PC use rather than laptop use (see Appendix C for all questionnaires used in PG2).

During the PG1 study, we had participants complete the privacy background questionnaire, the viewer classification task, and the web site classification task at both the install and the uninstall session. We were interested in whether participants' responses changed after explicitly reflecting on their incidental information privacy concerns during the course of the week. Analysis revealed minimal differences. We therefore elected to administer the privacy background questionnaire only during the install session and the two classification tasks only during the uninstall session to reduce the burden on participants. When presenting results, we will use a subset of the questionnaires completed by PG1 participants, matching the timing of administered questionnaires in PG2.

During the uninstall session, we also verbally asked the participants to reflect on how representative their browsing was that was captured during the week and what percentage of browsing they thought had been captured (i.e. were other computers used that did not have the logging software installed).

3.6.4 Content Categorization

The same set of web site categories (from [1]) that participants used in the theoretical web site classification task (Appendix C) was used to classify all of the browser activity conducted by participants over the course of the week. The parental control feature of Zone

Labs Security Suite [6] was enabled and all 34 categories offered (a subset of the classification task categories) were blocked. All browsing was sorted by URL and the URLs were then pasted into the address bar of a browser window. If the web site was blocked, its category was given as a reason. If the site was not blocked (approx. 50% of the time), it was manually classified by Kirstie Hawkey according to all 55 category descriptions and examples used in the theoretical task. Making use of the automated classifications where possible allowed us to ground the categorization in an actual commercial system and this exercise provided training for which types of sites were considered to be in each category. In a few instances, the automated classification seemed unreasonable; the site was then manually classified and feedback was provided to the commercial provider using the channel provided.

While classifying the visited web sites, we created two additional categories to cover the collected data. Pages were classified as *web content management* when it was clear that participants were using a content management tool within their browser rather than actually visiting a web page. Entries were classified as an *empty window,* if there was a log entry with no accompanying URL. These entries occurred when an image (e.g., a web advertisement) was loaded into an empty pop-up window, when no home page was set in the browser, or as a result of scripting on a page.

3.7 Summary of Mixed Methodology Approach

Our mixed methodology approach allowed us to examine the privacy of incidental information both in terms of general attitudes and also based on actual behaviours. Generalization of our results will be limited by the small sample sizes, particularly for the field studies where our focus was on capturing rich data from each participant. Furthermore participants in all three studies tended to be highly educated and students were over-represented. Additionally, participants in the PG1 field study were laptop users and were primarily technical males. Participants in the PG2 field study, however, were selected to explore the generalizability of results across device (laptop/desktop), technical background (non-technical/technical), and gender (more females). We believe triangulating the results from all three studies strengthens the overall validity of our exploratory results.

The results from these three exploratory studies will be presented next. Chapter 4 presents those results pertaining to general web browsing behaviour, Chapter 5 presents those results pertaining to visual privacy of traces of prior web browsing in general, while Chapter 6 presents the impact of context, including location and device, on web browsing behaviours and privacy concerns. We reflect upon the suitability of our chosen research methodologies in Chapter 9.

Chapter 4
Results: General Web Browsing Behaviours

This chapter presents results pertaining to general web browsing behaviour from our exploratory studies. Although web browsing behaviour was studied in detail in the mid-to-late 1990s (e.g. [25, 145]), few recent results have been reported. The nature of web browsing has changed extensively since these early studies, both in the profile of the typical web user and in the context of their browsing (e.g., location, connection speed, web browser features) (see Appendix A).

Privacy management of incidental information will be the secondary task; in order to build an effective privacy management tool, we must support privacy within the context of users' primary task of browsing the Web. Our goal in this research area is to explore those aspects of web browsing behaviour that will impact the design of a privacy management system. Additionally, it is important to understand web usage patterns as web browsing is such a frequent activity in many people's lives. Whittaker et al. [154] include the need to research daily activities and gain an understanding of users' tasks and behaviours as part of their reference task agenda for HCI.

Direct comparison of our results with earlier studies is difficult due to methodological differences. These include the participants' environment, task, and the location of the logging software location (client-side, proxy, or server-side). When comparing quantitative information (e.g., the number of pages visited, session length) it is crucial to understand the context of the prior studies [67], particularly given the continually evolving web browsing environment (as shown in Appendix A). Was all the browsing of the user captured or just that in a certain environment? When was the study conducted? Did the pages visited include cached pages, all pages navigated to, all pages fully loaded, frames, or other page elements such as images? How were sessions discriminated? For much of the related work in this area, it was difficult to determine pertinent methodological details so that we could relate our results to those obtained previously. We do, however, provide comparisons with previous research where appropriate.

This chapter presents findings from the PG1 and PG2 field studies in the following areas: number of pages visited, browser window usage, speed of browsing, and sessions. Findings also come from the IIP survey which provided self-reports of the general types of activities in which participants engage. The second field study (PG2) also provided information about the categories of pages that participants visited and their relative frequencies.

4.1 Number of Pages Visited

The number of pages visited impacts the feasibility of different approaches to classifying web browsing activity. If users conduct little browsing on average, manual privacy classification of each visited page may be feasible; however, if users visit many pages during the course of the day, a per-page approach may be overly burdensome.

Table 4 gives a summary of pages visited and browser window usage for participants in the PG1 and PG2 field studies. On average, each participant in the PG1 study (20 participants, August 2004) visited 1808 pages during the seven days (~258/day). However, the volume of page visits was highly variable; the total page visits by each participant ranged from 422 (~60/day) to 5127 pages (~732/day), with a standard deviation of 1252.7. This is a dramatic increase from earlier reports: 42 page visits/day (1999/2000) [31], 21 visits/day (1995) [146], and 14 visits/day (1994) [25]. Participants in the PG2 field study (15 participants, March 2005) averaged 2077 pages during the seven days (~297 pages per day). Again, this was highly variable; the total page visits by each participant ranges from 699 (~100/day) to 4966 (~709/day), with a standard deviation of 1328.7. A t-test found no significant difference between the mean numbers of page visits recorded during our two field studies ($t(33)=-.612$, $p=.545$).

Table 4. Quartile and mean values for number of pages visited by each participant and their browser window usage over the course of the week during the PG1 and PG2 field studies.

	Pages Visited		***Browser Windows***		***Pages per window***					
					Mean		***Mode***		***Max***	
Quartile	**PG1**	**PG2**	**PG1**	**PG2**	**PG1**	**PG2**	**PG1**	**PG2**	**PG1**	**PG2**
0%	422	699	47	64	3	5	1	1	27	51
25%	1064	1043	134	107	5	6	2	1	55	85
50%	1508	1338	246	205	7	8	2	2	92	119
75%	2133	3124	441	431	9	11	2	2	170	264
100%	5127	4966	799	516	20	15	2	2	255	355
Mean	1808	2077	289	260	8	9	1.85	1.67	108	166

There are several factors that may account for the increases we found in the number of page visits from that reported in prior research. During the earlier research studies [25, 31, 146], browsing was only captured in a single location, not accounting for all browsing that users may have done during the day. In contrast, PG1 participants conducted the majority of their browsing on their laptop computers and our PG2 participants indicated that we had captured almost all of their web visits across locations (avg. 98%, ranging from 80% to 100%). Timing of data collection may also have impacted page visit rates. Cockburn and Mackenzie [31] collected data via history files captured on university backup files. The data collection period included the holidays, which may account for the lower traffic levels. High-speed internet access is also far more prevalent than it was during earlier studies. Previous statistics have shown that users with high-speed internet view more web pages and surf the web more often than those with dial-up connections [125]. The popularity of web-mail, news sites, and the prevalence of pop-up windows may also account for increases.

It is difficult to contrast our results directly with those from a more recent long-term web usage study conducted by Weinreich et al. [148] (2004-2005). They captured data through a proxy, and augmented this with client-side data for a subset of the participants. It is unclear if the logging software was installed on all computers that their participants regularly used. Their method of determining a page visit was also different from our approach of recording page visits at the web document level. The authors counted individual html requests as page visits, and performed data processing in an effort to consolidate related frames into pages and to remove non-participant generated requests (i.e. auto reloading pages, advertising pop-ups). Furthermore, the authors do not report an overall daily average for page visits; however they do report that the browsing style and activity of participants varied widely, with participants averaging between 19.5 and 204.8 page visits per 'active' day. The authors define an 'active' day as a day where some logging was recorded.

4.2 Browser Window Usage

Browser window usage is another aspect of browsing behaviour that may impact the feasibility of different privacy management approaches. Managing privacy on a per-window basis might be an appropriate strategy. For example, one approach might be to have users classify all visited pages within a browser window as it closes rather than having users interrupt their flow of browsing by classifying pages as they are encountered.

Overall, participants in the PG1 study opened an average of 289 different browser windows during the seven days. This decreased somewhat for participants in the PG2 study, who opened an average of 260 browser windows. Again, this result was highly variable, as shown in Table 4. Across participants, the number of different browser windows opened ranged from 47 to 799 for PG1 and 64 to 516 for PG2. Figure 8 shows the actual per-window pattern of browsing for the first participant for the first hour of the PG1 study. This participant opened 7 windows and visited 78 pages during the hour.

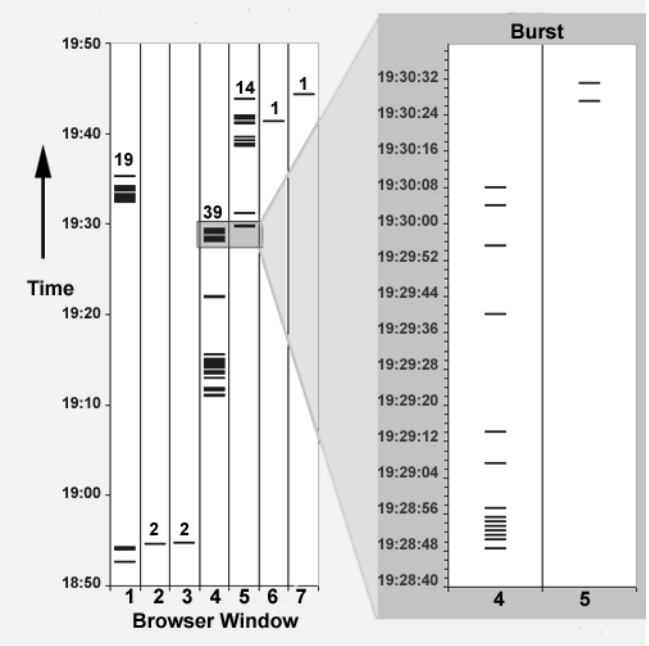


Figure 8. Example of temporal patterns of web browsing on a per window basis. A burst of activity is shown on the right.

The number of pages loaded in a browser window varied widely within users. In most cases, only one or two pages were viewed within each window, as can be seen by mode number of pages per window (Table 4). This relatively low number is not surprising given the number of windows automatically spawned while browsing. However, as the values for maximum number of pages viewed illustrate (Table 4), there were also several instances where large numbers of page views occurred within a browser window. The browsing patterns shown in Figure 8 are fairly typical: three browser windows of 14, 19, and 39 pages and four windows of 1-2 pages.

People frequently moved between open browser windows. Results from PG1 found an average of 158 (from 22 to 430) browser window revisitations. For example, Figure 8 shows three browser window revisitations (browser window #1 once. window #5 twice). Because we did not capture the browser window closing in PG1 and only logged the navigation to (and loading of) web pages, our analysis is limited to browser window revisitation for the purpose of navigation. We captured more extensive window focus events in PG2 so that we could perform an analysis of window revisitation for viewing as well as navigation purposes. However, the window focus events did not consistently appear in the log file, particularly during periods of rapid browsing. We were therefore unable to determine the rate of window revisitation for the purpose of viewing the page again.

We attempted to gain a sense of the number of concurrent browser windows that participants had open so that we could begin to analyze the extent to which participants had multiple browser windows open, containing pages of varying content sensitivity. Unfortunately, we found that the window close events were also not always captured, most likely as a result of a parent browser window automatically closing its child windows. In an effort to gain a conservative measure of multiple window usage, we tried inserting a close event after the event was navigation event was logged for that window. Figure 9 shows the number of browser windows that participant NTD1 (a non-technical, desk top user) had open at any given point during the course of the study. As can be seen, this participant has up to 6 windows open at a time (average of 1.8) and has periods with several windows open, but also times when single windows are being opened and closed. However, the analysis we can do through scripting is based purely on a sequential basis, and does not reflect the temporal nature of windows opening and closing (i.e. opening or closing multiple windows

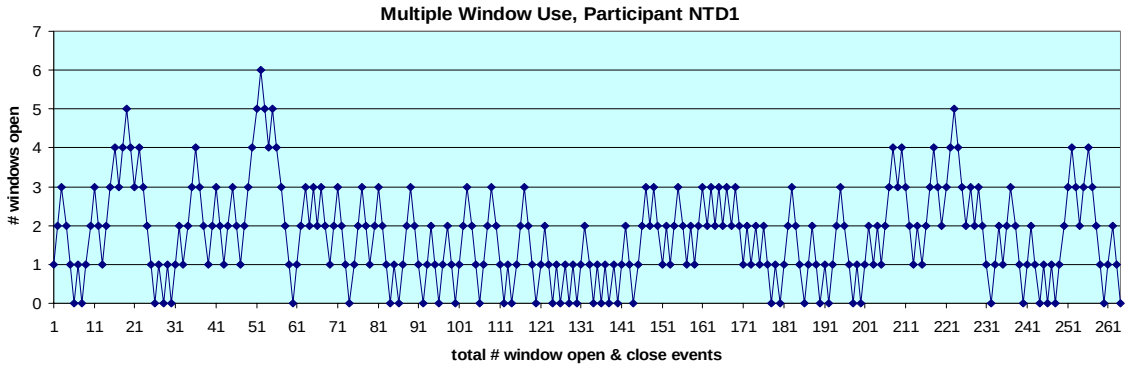


Figure 9. Number of concurrent browser windows open for participant NTD1 during the course of the week.

quickly). Furthermore, we cannot be certain which windows were intentionally spawned by the user and which were spawned automatically by advertisements or by the action of clicking on a link. Due to this uncertainty as to the interpretation of the findings a, we have elected to not pursue this analysis further.

As presented in section 2.2.1.3, the reasons for multiple browser window use have been described in studies investigating information seeking behaviour [12, 17, 28] and personal information management [80]. However, none of these studies provide metrics of multiple browser windows use. The only field study we found which quantifies some aspect of multiple browser window use is the 2004-2005 study conducted by Weinreich et al. [148]. They report that 10.5% of their participants' navigation actions were to open a new window and conclude that multiple window use has increased from earlier studies [25, 145] which found a new window rate of less than 1%. It should be noted that browser windows do not have to be opened from within the browser (i.e. multiple web browsers can be opened through short cuts or application icons); so there may not be a direct correlation between the number of new window navigations and the total number of browser windows opened. Furthermore, they did not examine how many windows may be open concurrently.

Tabbed browsers, such as Mozilla Firefox, now allow users to organize multiple open pages within the browser. Weinreich et al. [148] report that one participant with a tabbed browser explained that new tabs were used for closely related tasks while new windows were opened for the purpose of multi-tasking. Further study will be important to learn how tabbed browsers have impacted general web browsing behaviours. Per-window behaviours may be useful when incorporating the concept of task into web tools.

4.3 Speed of Browsing

The speed at which browsing occurs may impact the feasibility of some approaches to privacy management. Manual, real-time annotation of browsing would only be feasible if it did not interfere with rapid browsing. Participants in both field studies frequently exhibited rapid bursts of browsing with several pages loaded per minute. We define a burst to be a rapid sequence of web visits with less than one minute's elapsed time between web pages loading. Several examples of bursts can be seen in Figure 8, including one that runs from 19:28:48-19:30:31 with 16 pages opened in 104 seconds (6.5 seconds/page).

The average number of bursts of rapid browsing for each PG1 participant was 258 (~37 bursts per day). Overall, the average duration of a burst was 82 seconds, although the longest burst of rapid browsing was over 36 minutes. The average length of a burst was 7 pages, with bursts of up to 172 pages loaded quickly. The average speed was 12 seconds per page. See Table 5 for the quartile and mean values for bursts of rapid browsing during PG1, including the number of episodes per week, and the speed, duration, and length of bursts.

Table 5. Quartile and mean values for the number of episodes, speed, duration, and length of bursts over the course of the week (PG1).

	Burst (1 minute cutoff)					
	# Bursts per week	Speed (seconds/page)	Duration (minutes)		Length (pages)	
Quartile	***Total***	***Mean.***	***Mean***	***Max.***	***Mean***	***Max.***
0%	43	8	0.8	5	4	22
25%	156	11	1.2	7	6	38
50%	264	12	1.4	11	7	51
75%	324	13	1.6	13	8	81
100%	510	17	2.1	36	17	172
Mean	258	12	1.4	12	7	63

Prior research has also described rapid bursts of browsing. In a study of the Netscape History files of 17 users over 4 months (1999-2000) [31], rapid browsing was noted with few gaps longer than 10 seconds per page loading. However, this picture is incomplete as the History files included frames loading within a page. This overestimation of speed is mitigated somewhat as pages were sorted daily by the last time of access; if a page was revisited multiple times throughout the day, only one visit would be recorded and used during the calculations. The authors speculate that the rapid browsing may not only occur within a single window, but may occur across multiple windows. For example, if a separate window is opened to investigate search results, a participant may rapidly scan the page and then return to the results list to navigate to the next interesting result. Our data from PG1 confirms that rapid bursts do indeed occur across windows.

More recently, Weinreich et al. [148] found that 25% of their participants' web documents were displayed for less than 4 seconds before the next navigation event and 52% were displayed for less then 10 seconds. Only 10% of documents were displayed for more than 2 minutes. They authors note that they do not know if documents were being actively viewed, only that the next navigation event had not yet occurred.

4.4 Sessions

Privacy management on a per-session basis may be a viable approach when managing visual privacy within web browsers. As our participants were not required to specify the end of a browsing session, we used periods of inactivity to demarcate a session. We calculated sessions in the same manner as bursts, but with 10 and 30 minutes breaks between page loads delimiting sessions. With the 10 minute cut-off (see Table 6) participants in PG1 averaged 66 sessions per week (9.4 per day). Each session had an average duration of 13 minutes and length of 28 pages. Using a 30 minute cut-off (see Table 7), the number of sessions dropped to 38 per week (5.4 per day). Each session had an average duration of 33 minutes and length of 46 pages. Again, there is a large variability in per-session behaviour that may impact any per-session solutions to privacy management in web browsers.

Table 6. Quartile and mean values for the number of episodes, speed, duration, and length of sessions (10 minute cut-off) over the course of the week (PG1).

	Session (10 minute cut-off)					
	# Sessions per week	Speed (seconds/page)	Duration (minutes)		Length (pages)	
Quartile	***Total***	***Mean.***	***Mean***	***Max.***	***Mean***	***Max.***
0%	14	15	7	38	12	58
25%	48	29	9	59	18	105
50%	66	33	12	75	25	143
75%	82	38	14	88	32	237
100%	123	59	23	177	57	394
Mean	66	35	13	79	28	174

Table 7. Quartile and mean values for the number of episodes, speed, duration, and length of sessions (30 minute cut-off) over the course of the week (PG1).

	Session (30 minute cut-off)					
	# Sessions per week	Speed (seconds/page)	Duration (minutes)		Length (pages)	
Quartile	***Total***	***Mean.***	***Mean***	***Max.***	***Mean***	***Max.***
0%	9	24	17	60	16	72
25%	26	46	21	97	27	143
50%	40	53	34	174	38	226
75%	48	65	43	223	57	313
100%	62	88	54	295	102	805
Mean	38	56	33	171	46	258

In a 1994 client-side study [25], 25.5 minutes was used as the session delimiter; participants averaged 9.4 sessions over 3 weeks (~1 session every 2 days). During a longitudinal field study of home internet use in low-income families (circa 2001-2002) [75], participants logged in an average of 0.6 sessions per day. A study of laptop use by university students (circa 2000) [51] found an average of 3 sessions per day with a 10 minute cutoff. Our results demonstrate that number of sessions have changed greatly over the years.

4.5 Types of Browsing Activity

Another aspect of web browsing behaviour that will impact the development of a privacy management system is the types of browsing activities in which people engage. A person that only conducts browsing activities of a nature acceptable to their browsing environment and to their typical viewers (e.g., an employee who only conducts work-related, non-confidential, activities will at work) will have little need for a privacy management system. A person that has very limited activities of a sensitive nature may be able to manage their privacy more simply than someone who multi-tasks between sensitive and non-sensitive browsing tasks. In this section we examine the types of browsing activity that participant in the IIP survey reported and that we were able to observe in the PG2 field study.

4.5.1 General Activities

Almost all participants in the IIP survey reported that they used their web browsers for email (99.4%) and for accessing entertainment information (94.2%). Banking (82.5%), viewing medical information (81.3%), accessing technical support forums (78.9%), shopping (75.5%), and playing games (57.9%) were also popular activities. Fewer participants reported using their web browsers to view erotic material (43.0%) or visit personal improvement forums (37.7%).

These activities were reported at a higher rate than in a randomly sampled 2003 Stats Canada survey [140, 141]. This survey revealed that, of the 64% of households that had Internet access, 81% reported using it for email, 48% for banking, 56% for medical information, 29% for shopping, and 44% for games. The higher activity rates for our IIP survey participants may therefore indicate that they are more frequent and experienced Internet users than typical Canadians.

4.5.2 Categories of Web Pages Visited

During the PG2 field study, participants visited sites from 41 of the 55 possible web categories used in the theoretical classification task. These categories were taken from a commercial web filtering product (see [1] for full list of categories). Each participant visited a subset of those categories (15-29, avg. 21). Only 21 categories included page visits by at least half the participants.

Table 8 gives per-category descriptive statistics including overall page totals and the number of participants with page visits in each category. It is important to note that participants had very different usage patterns within a category. For example, News/Media appears to be a very popular category with 14 participants visiting a total of 1320 pages; however, a single participant accounted for 1032 of those pages and only 7 participants visited 10 or more pages in this category. Categories with less than 40 total cases each were grouped into *other*, including chat/instant messaging, cult/occult, gambling, gay/lesbian, hacking/proxy avoidance, military, sex education, and vehicles.

It is interesting to note that 2115 of the pages were categorized as Empty Window, likely resulting from scripting, blank home pages, or pop-up windows generated for advertisements. A further 158 pages were classified as web advertisements. Web advertisements could account for up to 7.2% of the total visited pages, despite the fact that in 15/20 of the computers that participants used during the study had pop-up blockers installed in their browsers (12 instances of IE blocking, 4 of Google, 1 of Yahoo). Weinreich et al. [148] found that for their eight participants who did not use pop-up blockers, over 28% of html requests were likely to have been generated by advertisements. This highlights the extent to which irrelevant pages may be included in convenience features intended for revisitation.

Only six participants sanitized some of their web page visits before submitting their data to us, accounting for 433 pages total. Of these, 107 did not have sufficiently detailed explanations to assign the page to a web browsing category. A further 14 pages could not be classified as the page was no longer accessible at the time of coding and did not have sufficiently descriptive URLs or page titles

Table 8. Per category descriptive statistics including overall pages and number of participants with page visits (total, 10+ pages).

Category	Overall page total	# participants	
		Total	10+ pages
Search Engines/Portals	6310	15	15
Education	3315	15	14
Email	5082	14	14
Reference	2055	14	13
News/Media	1320	14	7
Shopping	770	14	10
Arts/Entertainment	665	14	12
Society/Lifestyle	1136	13	8
Web Advertisement	158	12	3
Computers/Internet	146	12	5
Financial Services	510	11	10
Government/Legal	385	11	5
Web Communication	660	10	6
Sports/Recreation/Hobbies	431	10	5
Travel	366	10	7
Software Downloads	236	10	6
Health	165	10	6
News Group	1303	9	3
Job Search/Career	449	9	4
Business/Economy	178	8	4
Religion	127	8	2
Online Games	520	7	5
Streaming Media/MP3	148	7	4
Web Content Management	598	6	4
Political/Activism/Advocacy	57	6	2
Dating/Personals	600	5	4
Internet Auction	101	5	3
Humor/Jokes	77	5	1
Restaurants/Dining/Food	279	4	3
Pornography	258	4	2
Web Hosting	60	4	2
Real Estate	147	3	1
Brokerage/Trading	110	3	1
Intimate Apparel/Swimsuit	94	2	1
Other	229	13	
Empty Window	2115	15	
Total	31160	15	

4.6 Summary

In this Chapter we have presented results pertaining to the general web browsing behaviours exhibited by participants in our two field studies (PG1 and PG2) and as reported by participants in our IIP survey. Table 9 and Table 10 give a summary of chapter findings, including the implications of the results on the design of a visual privacy management system. Design guidelines will be synthesized from all results chapters and presented more formally in Section 7.1.

Table 9 summarizes results from the field studies that demonstrate how users' web browsing behaviours will complicate the development of any tool or technique for web browsing. The sheer number of pages that people visit while browsing means that manual tools, that operate on a per-page level, will be overly arduous and therefore impractical. Beyond the number of pages visited, the speed with which users browsed was at times staggering. The high volume of web sites visited and the rapid browsing indicate the need for seamless interactions between users and their web browser tools. There are also indications that participants may be multi-tasking at times, moving between multiple browser windows that are open.

Another important theme to our general web browsing behaviour results was the individual variation in web browsing behaviours (as summarized in Table 10). Participants' behaviours varied considerably in terms of the number of pages visited, number of separate windows in use, the session length and speed of browsing, as well as the content of visited pages. This variability makes it difficult to arrive at standard solutions for web browsing tools and techniques. Furthermore, there is great variability both across users and within the browsing of a single user. Any privacy management approach must be sensitive to the changing needs and behaviours of users and allow users flexibility.

Next, in Chapter 5 we present results pertaining to general incidental information privacy concerns during web browsing.

Table 9. Summary of chapter findings, including design implications for a visual privacy management system.

Concept	Sec.	Our Findings		Design Implications for a Visual Privacy Management System
		Study	Result	
Page visits, windows opened & browsing sessions	4.1	PG1, PG2	~275 page visits per day across studies. This is a dramatic increase from earlier studies [25, 31, 146].	Manual classification of web browsing activity on a per-page level would be difficult for users. Post hoc management on a per-window basis or per-session basis may be more feasible, but might be difficult to maintain.
	4.2	PG1, PG2	~275 browser windows opened per week Average 8-9 pages per window opened	
	4.4	PG1	~9.4 sessions per day (10 minute cut off) ~5.4 sessions per day (30 minute cut off) Increase from earlier studies [25, 51, 75]	
Rapid bursts of browsing	4.3	PG1	~258 bursts per day (PG1). Rapid browsing also noted in [31, 148] Bursts observed to continue across browser windows.	A privacy management system should not interrupt users' rapid browsing behaviours.
Concurrent use of multiple browser windows	4.2	PG1	~158 browser window revisits for the purposes of navigation	Privacy management system must support users switching between multiple windows, some of which may be opened for the purpose of multi-tasking (i.e. varying privacy sensitivities).
	4.2	PG2	Indication of multiple browser windows opened (e.g., NTD1 had up to 6 concurrent windows), confirming previous anecdotal observations [12, 17, 28, 80, 148].	

Table 10. Summary of web browsing behaviour results demonstrating the range of individual variability.

Concept	Sec.	Study	Our Findings	Design Implications
Each user exhibits variability within their own browsing behaviour	4.2	PG1, PG2	Mode of 2 pages loaded per browser window opened, but average max. ~130	For each user, privacy management approaches must be viable across the range of their web browsing behaviours.
	4.3	PG1	Avg. burst duration is 82 seconds, but avg. max. duration is ~12 minutes. Avg. burst length is 7 pages, but avg. max. length is ~63 pages	
	4.4	PG1	Avg. session duration is 13 minutes, but avg. max. duration is 79 minutes (10 min. cutoff) (33/171 for 30 min. cutoff) Avg. session length is 28 pages, but avg. max. length is 174 pages (10 min. cutoff) (46/258 for 30 min. cutoff)	
Web browsing behaviour is highly individual and varies between users	4.2	PG1, PG2	Number of page visits per day ranged from 60/day to 732/day.	A privacy management system must be customizable to an individual or flexible enough to work for users with varying behavioural patterns.
	4.2	PG1	Number of browser window revisits for navigation ranged from 22/week to 430/week	
	4.3	PG1	Number of bursts ranged from 43/week to 510/week.	
	4.4	PG1	Number of sessions ranges from 14/week to 123/week (10 minute cutoff) and 9/week to 62/week (30 minute cutoff),	
	4.5.1	IIP Survey	The percentage of participants reporting each activity varied; fairly high occurrence rates for some of the more sensitive activities (e.g., medical information, 81.3%; erotica 43.0%).	
	4.5.2	PG2	Overall wide range of activities (41/55 categories), but each individual visited a subset (avg. 21).	

Chapter 5
Results: Incidental Information Privacy in Web Browsers

This chapter presents the results from our exploratory studies which pertain to incidental information privacy in web browsers. We begin by reporting results showing the scope of the incidental information privacy problem, which confirmed our motivation to conduct research in this area. We then present results concerning participants' application of privacy levels to their web browsing during the field studies. Finally, we present several factors of incidental information privacy that we identified and use those factors to frame the presentation of results. Our focus in this chapter is on the general privacy results, irrespective of environmental contexts such as device and location. Chapter 6 will examine the impact environmental contexts had on both web browsing activities and on privacy concerns.

5.1 Scope of the Incidental Information Privacy Problem

We were interested in determining how often participants are in situations where they are working closely together so that others could view their displays and how frequently others actually use their computers. We also asked participants to indicate what actions they take to protect their privacy in these situations if they were given advanced notice that somebody would be viewing their display. While incidental information privacy is certainly not a concern for everybody, our results show that it is a concern for many which validates our motivation for conducting research in this area.

5.1.1 Frequency of Viewers and Users

In each of our studies, we asked participants to tell us the frequency with which ten different categories of people (e.g., spouse, colleague) could view their display and use their computer. We present the results from the IIP survey here as the survey population is most representative of the general population. Participants in the field studies were required to have prior incidental information privacy concerns as a prerequisite for inclusion so may have more opportunities where others can view their displays. The original scale upon which

participants indicated the frequency of potential *viewers* and *users* of their computers was daily, weekly, monthly, rarely, and never. For the purpose of this analysis, we collapsed the frequency responses into the categories 'regularly' (daily, weekly), 'occasionally' (monthly, rarely) and 'never'.

All 155 IIP survey participants reported at least one category of viewer that could sometimes *view* their display and 93.5% (145/155) reported at least one category of potential *user*. The viewing frequency and usage frequency (see Table 11) both varied depending on the category of the viewer/user. The most regular viewers were colleagues, spouse/significant other, and supervisors. Close friends, acquaintances, and technical support were more likely to be occasional viewers. Audiences at presentations, employees, parents, and clients were least likely to have been reported as potential viewers. As can be seen in Table 11, participants reported that others *used* their computers with a lower frequency than for others *viewed* their display. Spouses were reported to be the most regular users of participants' computers. Over half of the participants reported that spouse/significant other, close friends, colleagues, and technical support staff were at least occasional users of their computers.

Table 11. The percentage of participants at each frequency (regularly, occasionally, never) for each category of potential viewers and users. The most common responses for each viewer type are highlighted.

	Frequency of viewing (%)			Frequency of using (%)		
Viewers	**Regularly**	**Occasionally**	**Never**	**Regularly**	**Occasionally**	**Never**
Close friends	36.5	49.6	13.9	16.4	41.8	41.8
Colleagues	56.0	29.9	14.1	16.7	35.6	47.7
Acquaintances	20.6	58.1	21.3	2.2	34.3	63.4
Spouse/Significant other	49.6	20.7	29.6	38.1	24.6	37.3
Technical support	9.7	59.7	30.6	7.4	48.9	43.7
Supervisor	37.1	28.1	34.1	5.4	22.3	72.3
Audience	3.0	47.0	50.0	--	--	--
Employees	22.9	19.1	58.0	4.6	18.5	76.9
Parents	11.7	24.1	64.2	3.6	18.2	78.1
Clients	9.8	19.6	70.7	0.8	6.1	93.2

5.1.2 Actions Taken to Preserve Privacy

During both of the field studies as well as the IIP survey, participants were asked to reflect on what actions they might take to conceal potentially sensitive information if given advanced warning that somebody else would be working closely with them (see Table 12 for

question specifics). Participants responded to this question for each *applicable situation of use*: laptop, home computer, work/school computer. In this section, we focus on the actions taken averaged across all applicable situations of use. In Chapter 6, we will further break down this analysis to examine the impact of device and location on the results.

Table 12. Question investigating privacy preserving actions prior to collaboration.

Question: If you had advance warning that somebody else would be working closely with you as you used your web browser and could see all areas of your screen, what actions would you take to conceal potentially sensitive information? (check all that apply) ***Answer choices:*** *1) No actions; 2) Retain control of the keyboard/mouse and limit functionality; 3) Check Favorites/Bookmarks and remove any inappropriate web pages; 4) Check Favorites/Bookmarks and rename any inappropriate web pages; 5) Check History and clear if any inappropriate entries; 6) Check Auto-completions and clear if any inappropriate entries; 7) Erase all Favorites/Bookmarks; 8) Erase all History records; 9) Erase all passwords in Auto complete; 10) Erase all forms in Auto complete*

The majority of participants in the survey (64.3%) reported that they would take some action if given advanced warning that someone could view their display. Furthermore, in the PG1 field study, 95% (19/20) of participants indicated that they would take some actions; all of these participants used laptops for the majority of their browsing. In the PG2 field study, participants indicated they would take some actions in 91.7% of applicable situations of use. The higher rate for actions found in the two field studies were likely due to the fact that one of the inclusion requirements for the field study participants was that they have occasions where others can view their browser window, while the survey participants had no such requirement for inclusion. The field study participants may therefore have a heightened level of privacy awareness. Across all three studies, participants reported taking some actions in 67% of applicable situations of use.

We next discuss the different actions that participants reported they take to preserve their privacy when given advance notification that somebody will be working closely with them and will be able to look closely at their display. Figure 10 shows the percentage of applicable situations of use (i.e. laptop, home computer, work/school computer) for which participants across the three studies indicated they would take each action.

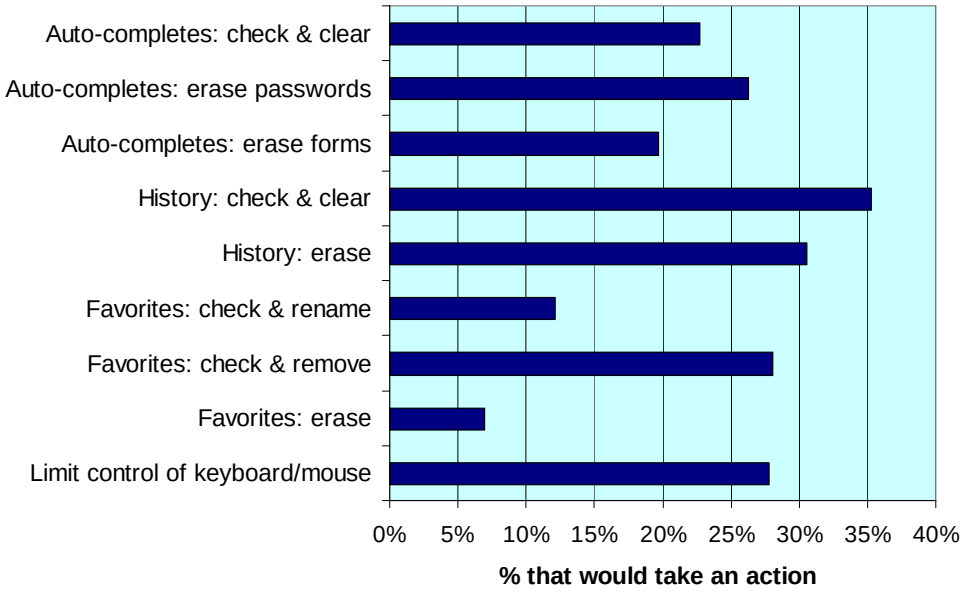


Figure 10. Percentage of applicable situations of use for which participants (across all three exploratory studies) indicated they would take each action.

One of the actions participants (27.8%) reported taking was physically limiting their collaborator's control of the keyboard or mouse during the collaboration. Most participants (59.5%) also reported taking actions involving the data stored within their browser convenience features (i.e., History, Auto Complete, Favorites).

Overall, 56.4% of participants reported taking at least one action with either their Auto Complete or History data (Figure 10) and there was a great deal of overlap between the two. At the time of these studies, it was not easy to clear the text that would appear in the Auto Complete functions in IE. In the Auto Complete menu, users could only clear the data used for the form and username/password text; users had to clear their History to prevent web addresses from appearing as Auto Complete options in the URL field. The latest version of IE follows the lead of other browsers and provides a single location where users can specify what traces of browsing activity they would like to clear. Because of the dual use of History records (to populate both the History feature and the Auto Complete selections for URLs) it is difficult to know how many participants took actions to check and/or clear their History because of concerns with their Auto Complete or concerns with the History.

Fewer participants indicated they would take one or more actions involving their Favorites (35.3% total) than would take one or more actions involving their Auto Complete (40.9% total) or their History (49.1% total) features (Figure 10). Given that users must

explicitly store web pages within their Favorites, they may be more selective about which items they store and have less concern about what information may be visible.

5.1.3 Summary

The results we have presented in this section show that privacy of incidental information is indeed a concern for many. Our findings clearly show that not only did participants have regular occasions when others could view their display; the majority would also take some action if given warning that this would happen. These findings support our motivation for investigating visual incidental information privacy concerns and developing privacy management approaches.

5.2 Patterns in Privacy Level Application

As our results from Chapter 4 illustrated, management of incidental information may be difficult due to the large volume of information. One of the main issues when managing the privacy of traces of incidental information within web browsers is classifying web pages and other artifacts with an appropriate privacy level. We examined the actual web browsing activity during the field studies in an effort to find patterns in the application of privacy that may support a semi-automated approach to privacy management. We first present patterns in the application of privacy levels depending on the different content categories of visited pages. We then discuss temporal patterns related to browser window usage.

5.2.1 Per Content-Category Utilization of Privacy Gradients

Results from participants' privacy classifications (using the 4-level privacy gradient scheme) of their actual browsing during the PG2 field study give insight about the sensitivity of various categories of web pages. For the PG2 field study, we determined the content category of each visited page using the Cerberian content categories [1] used in commercial web filtering applications. We would expect that page visits classified in the Financial and Health categories would be considered sensitive [8, 150] as would page visits from categories that might be considered by some to be a social transgression (e.g., Pornography, Gambling) [115]. Furthermore, web sites in categories that might reveal personal activities (e.g., Religion, Travel, Sports/Recreation/Hobbies) might also be considered sensitive, particularly for those browsing in a workplace environment [115].

Not surprisingly, participants classified different categories of browsing with varying privacy sensitivities. We chose to use k-means cluster analyses to determine whether the categories could be grouped into clusters based on the relative proportions of pages that were classified at each privacy level. K-means is an iterative distance based clustering method which uses a Euclidean distance function to determine in which cluster to place each instance [155]. The statistical package used, SPSS, selects initial cluster centers to represent k well-spaced cases across the data [48]. Using a data vector consisting of category name, % of pages classified as public, % of pages classified as semi-public, % of pages classified as private, and % of pages classified as don't save, we performed an iterative k-means cluster analyses of the 33 most common categories. We assessed values for k ranging from three to six; the best fit to the data in terms of cohesion and comprehension was found when k was equal to five. Table 13 shows the cluster means, number of categories in each cluster and the percentage of total page visits attributed to categories in the cluster. Examination of the cluster centers reveals the predominant privacy levels that characterize each cluster: C1: *public/don't save,* C2: *public,* C3: *semi-public,* C4: *mixture*, and *C5: private.*

Table 13. Results of cluster analysis of web page categories by applied privacy levels. Highlights indicate the privacy levels that characterize each cluster.

	Clusters	C1	C2	C3	C4	C5
Privacy Level	Overall	Final Cluster Centers				
Public	40.0%	48%	84%	23%	51%	3%
Semi-Public	19.6%	10%	8%	72%	22%	10%
Private	25.3%	3%	2%	3%	16%	81%
Don't Save	15.1%	39%	6%	2%	11%	6%
Number of Categories		5	8	5	10	5
% of Total Page Visits		9.2%	9.8%	6.4%	44.1%	21.0%

Cluster C1 (public/don't save) accounted for 9.2% of all pages visited and included the categories Arts/Entertainment, Shopping, Society/Lifestyle, Web Advertisements, and Streaming Media/MP3 (see Figure 11). These categories are fairly general and may contain pages with content of varying sensitivities. Participants labeled most (80-95%, avg. 87%) of the pages in each category as being either *public* or *don't save.* Still 5-15% of pages were classified as private or semi-public (i.e. potentially private) depending on the viewing context. Given the high amount of public browsing, for these categories, the *don't save* label most

likely means a page is irrelevant, rather than being extremely private, with the possible exception of the Streaming Media/MP3 category which exhibited a lower percentage of public pages.

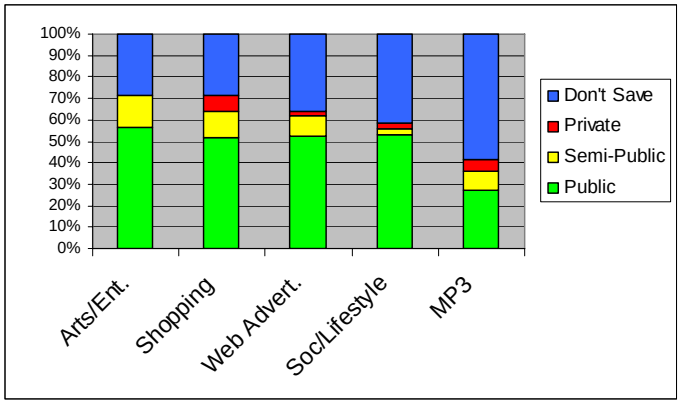


Figure 11. Relative privacy levels of categories in C1 (public/ don't save).

Cluster C2 (public) accounted for 9.8% of all pages visited and included the categories Real Estate, News/Media, Brokerage/Trading, Government/Legal, Political/ Activist/Advocacy, Restaurants/Dining/Food, Online Games, and Software Downloads (see Figure 12). The majority (75-100%, avg. 84%) of the pages in each category were labeled as *public*. However, there were still some potentially sensitive pages within these categories (i.e. 11-20% of visited pages were labeled as either private or semi-public for 5/8 categories).

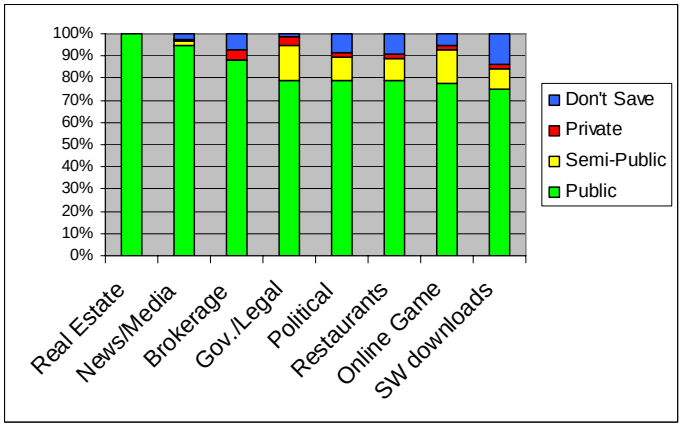


Figure 12. Relative privacy levels of categories in C2 (public).

Cluster C3 (semi-public) accounted for 6.4% of all pages visited and included the categories News Group, Job Search/Careers, Humor, Web Hosting, and Internet Auction (see Figure 13). Participants classified the majority (64-78%, avg. 74%) of pages in each category as *semi-public*, indicating that the pages may be public or private depending on the

viewing context. Interestingly, with the exception Job Search/Careers, these categories had very few pages (in 3 cases, none) indicated as being private.

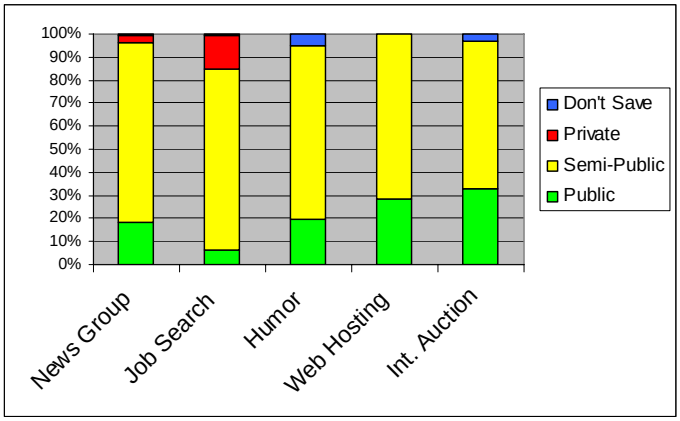


Figure 13. Relative privacy levels of C3 categories (semi-public).

Cluster C4 (mixture) accounted for 44.1% of all pages visited and included the categories Education, Web Communication, Sports/Recreation/Hobbies, Business/ Economy, Computers/Internet, Reference, Search Engines/Portals, Religion, Travel, and Health (see Figure 14). These categories were frequently visited, both in terms of number of pages (165-6310 pages per category) and in number of participants (8-15 participants per category). Categories in this cluster were characterized as having a more even spread across privacy levels than in other clusters (*public*: 30-64%, avg. 51%; *semi-public*: 14-36%, avg. 22%; *private*: 1-37%, avg. 16%; *don't save*: 0-24%, avg. 11%).

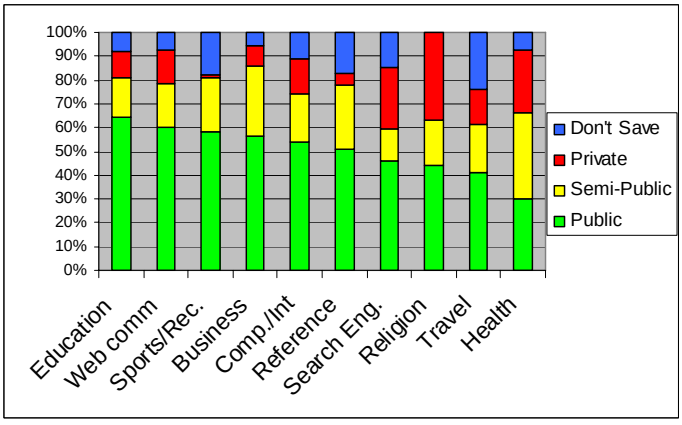


Figure 14. Relative privacy levels of categories in C4 (mixture)

Cluster C5 (private) accounted for 21.0% of all pages visited and included the categories Intimate Apparel/Swimsuit, Dating/Personals, Pornography, Financial Services, and Email (see Figure 15). Categories in this cluster are characterized as being *private* (58-

94%, avg. 81%) or potentially private depending on the viewing context (total private/semi-public: 85-97%, avg. 91%). For these categories, it is likely that those pages classified as *don't save* include some that are due to the pages being extremely private rather than irrelevant.

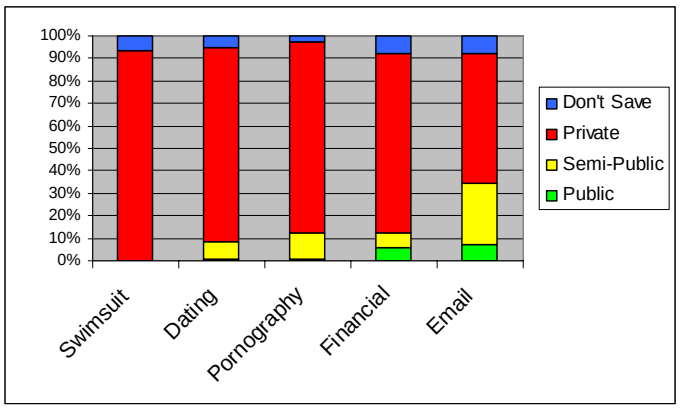


Figure 15. Relative privacy levels of categories in C5 (private)

5.2.1.1 Limitations

As was discussed throughout these results, the dual nature of the privacy classification *don't save* has complicated our analyses of the privacy perceptions for the different categories of web browsing. The privacy classification of other pages within the category may provide some indication of whether don't save was used as irrelevant or extremely private. For example, if most other pages were classified as public (e.g., cluster C1), classifications of don't save may indicate pages that were irrelevant. Similarly, if most other pages were classified as private (e.g., cluster C5), then the use of don't save may be more likely to indicate pages considered to be extremely private.

While we can not be sure of the privacy sensitivity of pages classified as don't save, it is clear that our participants found this category useful. A privacy enhanced web browser should provide mechanisms to allow users to easily remove unwanted traces of activity from their convenience features, whatever the underlying reason for not wanting to save a record of the activity. Such mechanisms could prevent the storage of the traces as the time of browsing or allow easy deletion of selected traces after the fact.

5.2.2 Temporal Patterns of Privacy Application

We examined the data from both field studies to identify patterns in the application of privacy levels on a per window basis. We defined a streak to be two or more consecutive

web pages of a given privacy gradient within a browser window. For example, in Figure 16, which represents one participants browsing during the course of one hour, 4 streaks occurred in browser window #4: there was a single *semi-public* page, followed by a streak of 3 *public* pages, a streak of 8 *semi-public* pages, a streak of 20 *don't save* pages, a single *public* page, and, finally, a streak of 5 *semi-public* pages. Detailed analyses of the PG1 field study revealed that 85% of all page visits occurred within a streak and the average streak length was 6.5 pages (maximum 166 pages). For the PG2 field study, 87% of all page visits occurred within a streak and the average streak length was 7.5 pages (maximum 355 pages).

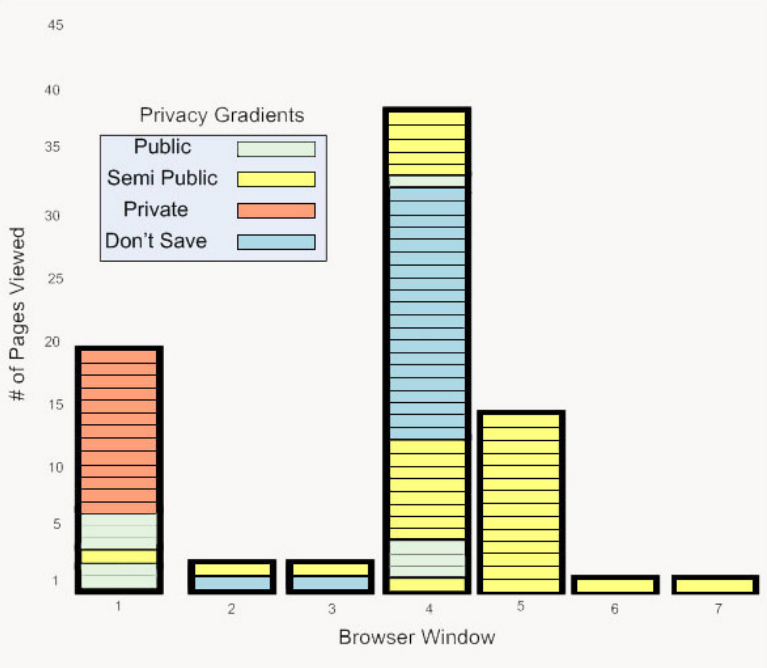


Figure 16. Hand crafted visualization of one participant's browsing during one hour showing example of sequential patterns of privacy application in browser windows.

A transition is defined to be a switch between privacy levels within a browser window. For example, in Figure 16, there are five transitions in window #4. In PG1, 56% of browser windows contained no transitions, and on average, participants had 0.9 transitions per window. In PG2, 57% of browser windows contained no transitions, and on average, participants had 1.1 transitions per window. Strictly looking at the number of transitions in a browser window may be misleading. For example, 5 transitions over 11 pages would indicate that the user transitioned between privacy gradients very frequently; however, 5 transitions over 50 pages are more reasonable. Transitions were normalized (# transitions ÷ # pages in window), resulting in a numerical score between 0 and 1 where high values indicate rapid

transitions. On average, participants in PG1 had a transition score of 0.14 (from 0.03 to 0.31). For PG2, participants had an average transition score of 0.13 (from 0.0 to 0.25).

5.2.3 Summary

Both the per-content category patterns and the temporal patterns in privacy level application that were evident in the data collected from both field studies (PG1 and PG2) have potential as mechanisms to support an automated or semi-automated approach to privacy management. In Chapter 7, we examine the feasibility of content categories for automating privacy level classification. In Chapter 8, we present PrivateBits, a proof of concept visual privacy enhancing web browser that makes use of the temporal privacy patterns inherent in web browsing.

5.3 Factors Impacting Incidental Information Privacy

A key objective of our research was to define the domain of incidental information privacy with respect to traces of web browsing activities. As presented in Chapter 2, prior privacy theory (section 2.1.1), research investigating privacy concerns for other domains (section 2.1.2), and research developing models of privacy (section 2.1.3) have found that privacy is highly individual and contextual. Through an examination of the related work (summarized in Table 1), we identified several factors of incidental information privacy that we believe directly impact a user's *privacy comfort level* in a given viewing situation: 1) their *inherent privacy concerns*, 2) their *level of control* retained over input devices, 3) their *relationship to the viewer* of the display, and 4) the *sensitivity of potentially visible content*. Furthermore, in a web browser, the specific content that may be visible depends upon recent *browsing activity*, *browser settings*, and any *preventative actions* taken. Additionally, the context (i.e. location, device) of the browsing activities and viewing opportunities may impact web browsing behaviours and privacy concerns. While Figure 17 shows what we believe to be the major influences on privacy comfort levels, these factors are often inter-related. For example, advance knowledge of a specific viewer may trigger preventative actions to limit the visible content.

The factors shown in Figure 17 are specific to traces of web browsing activity; however, while the nature of the visible content will change for other types of incidental information, the impact of sensitivity of the potentially visible content, level of control, viewer, and inherent privacy concerns will likely apply to other personal information

management systems. For example, a desktop search PIM system will generate different types of potentially visible information and have different settings and filtering mechanisms for results. However, the sensitivity of the information which may be visible, the level of control retained over what is displayed (e.g. avoiding specific searches), the relationship to the viewer of the incidental information, and the inherent privacy concerns of the user will likely impact the privacy concerns for a given situation.

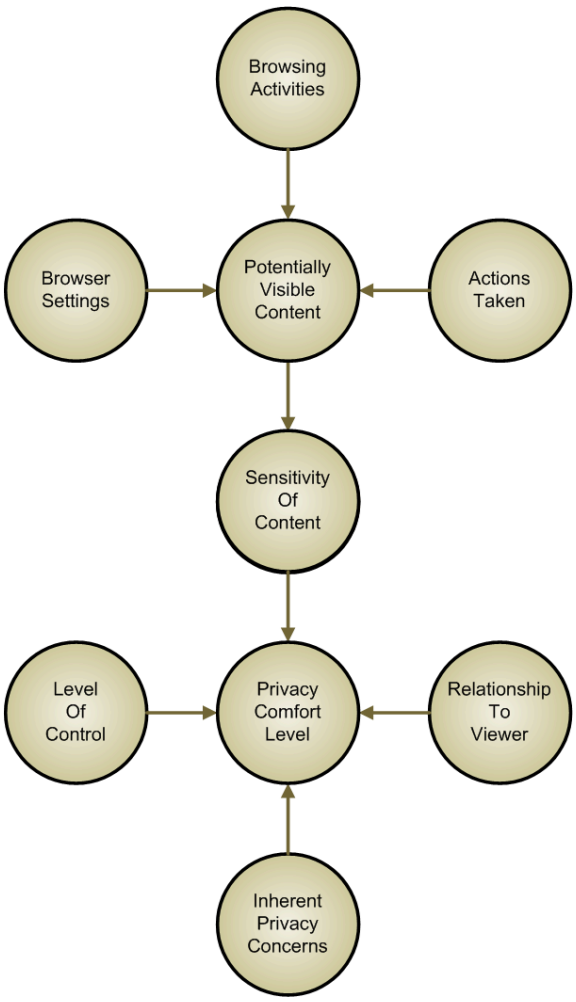


Figure 17. Factors that affect the comfort level of users during incidental viewing traces of prior web activity.

For each factor of incidental information privacy, we wanted to examine the extent and variability of user behaviour and concerns. If behaviour and concerns are consistent across users, we can use a standard approach in a privacy management solution. If participants cluster into groups, we can try to determine best management practice for those instances. However, we will also need methods of determining to which group an individual

belongs so that the appropriate automated approach to privacy management is taken. Individualized privacy management systems may be able to simplify privacy preference configuration by only presenting options along those aspects of privacy pertinent to the individual.

The four primary factors of incidental information privacy will be used to frame the discussion of results from the exploratory studies. Results will be presented from the IIP Survey and the contextual browsing data collected during the PG2 field study. In this chapter, we limit our presentation of results to the overall privacy concerns regardless of the setting.

5.3.1 Overall Impact of Factors on Privacy Comfort Levels

The survey presented participants with scenarios of varying sensitivity and asked them to give a rating of their privacy comfort level (PCL) on a scale from 1 (extremely uncomfortable) to 7 (extremely comfortable) for each of 5 potential viewer types and three levels of control. The potential viewer types included spouse, close friend, parent, colleague, and supervisor. Three levels of control were examined: the participant in control of their browser, the other person in control of their browser with the participant right there, and the other person in control of the browser with the participant leaving the room. Four scenarios were examined with varying levels of content sensitivity: one meant to be universally embarrassing, one meant to be neutral, one meant to be positive, and one where participants were asked to reflect on their usual browsing behaviour. As responses for the positive and neutral scenarios were virtually identical, only results from the neutral scenario are given.

Analyses of the IIP survey results revealed that privacy comfort levels were highly contextual overall. Privacy comfort levels were related to the potential viewers, the level of control, and the sensitivity of the content affecting the level of comfort (as shown in Figure 18). On average, participants reported that they were most comfortable when considering the neutral scenario, with their spouse/significant other as the viewer, and with themselves in control of the keyboard and mouse. On the other extreme, participants reported they were least comfortable when considering the embarrassing scenario, when leaving the room with their supervisor in control of the keyboard and mouse.

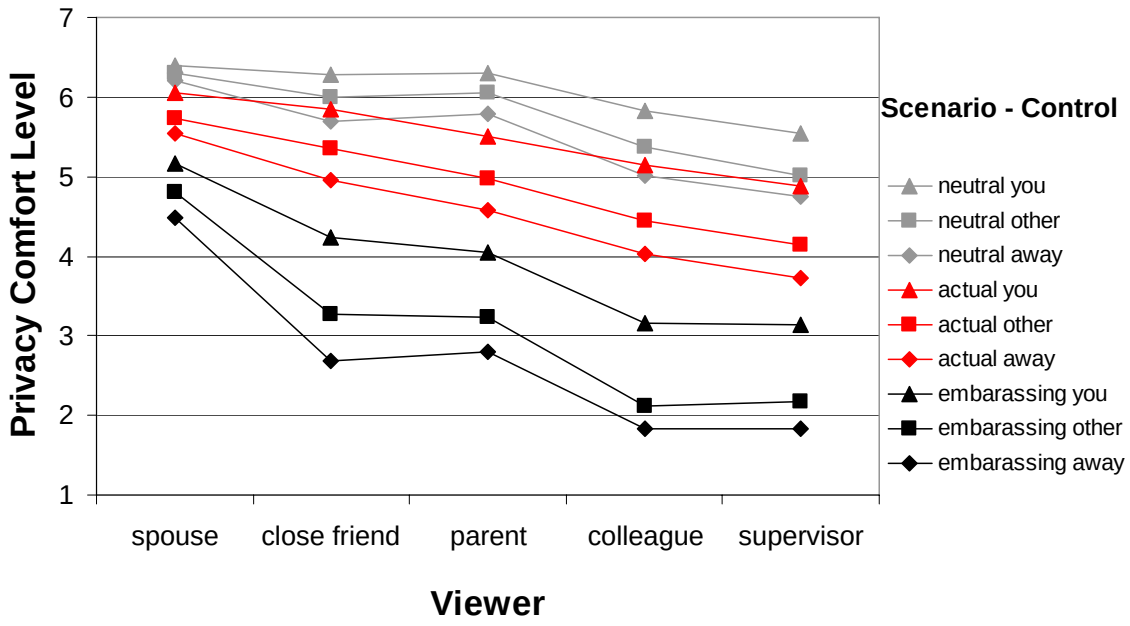


Figure 18. Comparison of privacy comfort levels (y-axis) according to the context of potential viewer (x-axis), scenario (colour of series; neutral-grey, usual browsing-red, embarrassing-black), and level of control (marker shape; triangle-you in control (you), square-other person in control with you there (other), diamond-other person in control and you leave the room (away).

We next examine the impact on the overall privacy comfort level for each these factors as well as the individual variability within the factors.

5.4 Sensitivity of Potentially Visible Content

The sensitivity of the potentially *visible content* should have an effect on privacy comfort levels. Traces of activity that are in character with the persona a user is trying to maintain [49] and are appropriate for the setting where the traces are viewed should cause little concern (e.g., non-confidential, work-related, browsing activity in the workplace). However, activities that reveal information that is not part of the persona presented (e.g., political affiliation) or that are perceived as transgressions (e.g., personal browsing if company policy does not allow it) may cause great discomfort [115]. Techniques to increase the recognition of information stored in convenience features (e.g., thumbnails of web pages in history files) may help users more easily find a desired page [82], but are also a privacy concern as they increase the visibility of incidental information.

While information sensitivity is known to be a contributing factor to privacy concerns, we needed to determine the role that content sensitivity played for our specific privacy domain: visual privacy of incidental information within web browsing. The perceived sensitivity of participants' general web browsing practices was investigated in the IIP survey through the *usual browsing scenario*. Furthermore, both field studies give us perspective about the overall privacy concerns participants had for pages they visited over the course of the studies. We next present results related to the sensitivity of potentially visible content.

5.4.1 Survey Results

Privacy comfort levels when participants reflected on their usual web browsing were lower than for the neutral scenario, but far higher than for the embarrassing scenario (as seen in Figure 18). This gives us some indication of how sensitive participants feel their typical browsing habits are. On average, 66.2% of participants rated their level of comfort higher when reflecting on their usual web browsing than when reflecting on the embarrassing scenario, 27.6% rated it the same, and 6.2% rated it lower. The embarrassing scenario was designed to give us an indication of the upper bound of participants' discomfort for traces of their web browsing activity. However, participants' actual discomfort in a given situation may depend on other factors such as their relationship to the viewer of the information or the setting in which the information is viewed. For the 33.8% of participants who indicated they would have the same comfort or less if traces of their usual web browsing were viewed, the scenario was not the most discomforting scenario imaginable or was similar to other sites they regularly visit. The medical nature of the sites given in the embarrassing scenario (e.g., www.yoursexualhealth.com/stoptheburning.html) might have mitigated some morality concerns that may have been associated with activities participants considered when reflecting on their usual browsing. Further investigation showed more participants indicated a higher level of discomfort for family viewers than for co-workers for the usual browsing scenario as compared to the embarrassing scenario. The personal nature of the embarrassing scenario may have violated the persona kept for co-workers, thereby provoking a stronger response; however, participants may have envisioned sharing medical concerns with family, but not other private activities such as erotica.

5.4.2 Results from Field Studies

During analysis of the PG1 and PG2 field studies, patterns emerged related to participants' classification of their actual web browsing activity using the 4-level privacy gradient scheme (public, semi-public, private, don't save). However, it is important to recognize that as these were field studies capturing participants actual browsing, different participants visited and classified different sets of web pages (all pages they happened to visit during that week). As such, if two people exhibited similar behaviours, it does not necessarily mean that they have similar privacy perspectives. For example, if two participants classified a large number of pages as being private, there is no way of knowing whether they both consider many types of sites to be private or whether one considers fewer types of sites to be private, but visited more of those private sites. These patterns do, however, reflect the perceived need for privacy based on the sites that an individual visited.

All participants utilized all privacy categories when classifying their visited web pages (with the exception of one user in PG1 and two users in PG2 who did not use the *don't save* category). This use of all four privacy levels validates the need for a more nuanced approach than the strict Public/Private or Save/Don't Save approach that is currently used in web browser convenience features and privacy management tools.

Of all the browsing captured in PG1 (36,170 page visits), 42% was classified as *public*, 25% as *semi-public*, 15% as *private*, and 18% as *don't save*. Results were similar in PG2 with 40% of 31,160 total page visits classified as *public*, 20% as *semi-public*, 25% as *private*, and 15% as *don't save*. A comparison of participants' classifications (normalized on a per-participant basis) with t-tests revealed no significant differences in the mean percentage of visited pages classified at each privacy level between participants in PG1 and PG2. Figure 19 shows a comparison of the mean percentages of visited pages classified with each privacy level (95% confidence interval shown) between participants in the two studies.

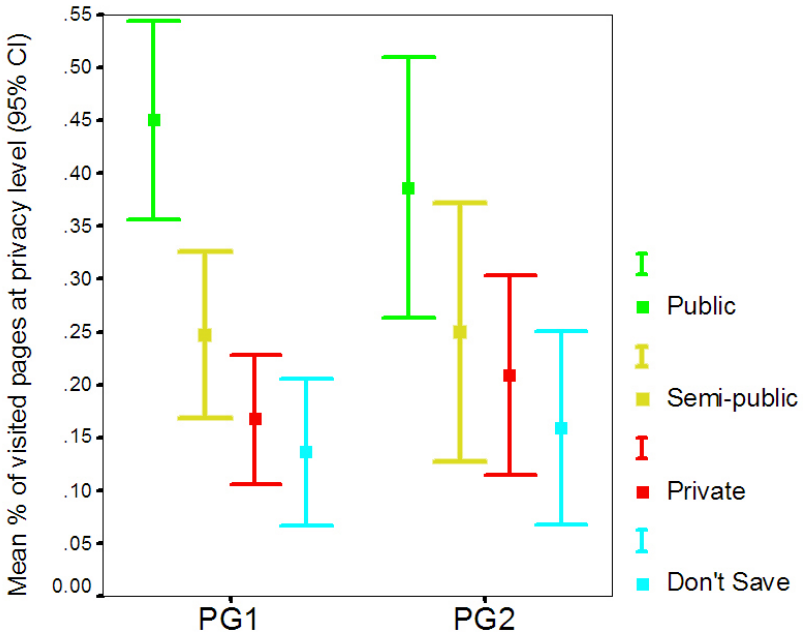


Figure 19. Comparison of the mean percentage for each privacy level between participants in PG1 and PG2 (95% confidence interval shown).

There was a great deal of variability between participants within each study as is evidenced by the large confidence intervals shown in Figure 19. In order to investigate whether common patterns in privacy application existed, we conducted k-means cluster analyses for the participants in each field study to determine whether they could be grouped based on the relative proportions of sites they classified with each privacy level. Using a data vector consisting of participant ID, % of pages classified as public, % of pages classified as semi-public, % of pages classified as private, and % of pages classified as don't save, we performed an iterative k-means cluster analyses of the participants in each field study. We assessed values for k ranging from three to six; the best fit to the data in terms of cohesion and comprehension was found when k was equal to four (see Table 14 for results from PG1). Examination of the cluster means revealed that each of the four clusters represents a group of individuals with a relatively high proportion of web browsing in one of the privacy gradients (C1-*semi-public*; C2-*private*; C3-*public*; C4-*don't save*).

Table 14. Results of cluster analysis of Privacy Gradient use in PG1.

	Clusters	C1	C2	C3	C4
Privacy Gradient	**Overall**	**Final Cluster Centers**			
Public	42%	22%	36%	62%	18%
Semi-Public	25%	58%	21%	16%	28%
Private	15%	9%	36%	11%	9%
Don't Save	18%	11%	7%	11%	46%
Number of Participants		3	5	10	2

Participants in cluster C1 had a large proportion of web sites that they considered to be semi-public. On average, participants in cluster C2 were evenly split between public and private classifications, and had a smaller number of sites that they considered to be semi-public. Although participants in this cluster are distinguished by their relatively high proportion of private sites, they still only considered 36% of the sites to be private. Participants in cluster C3 are distinguished by a higher than average amount of sites classified as public. Finally, the two participants in cluster C4 are distinguished by the number of sites they classified as don't save. It is unclear if these participants considered those sites to be extremely private or irrelevant. Analysis of PG2 data showed similar results in terms of cluster means; however, participants were more evenly divided between clusters.

Most of the participants across the two field studies (32/35) reported that the four privacy categories fit well at least most of the time; however many (17/35) reported difficulty classifying some of the visited sites (~15% of visited sites). Reasons given for the difficulty included that it depended on the person they envisioned viewing a record of the page visit (10/17), that it depended on the viewing location (7/17), that the site had multiple purposes (5/17), or that there were other reasons (5/17) (e.g., the time of day, variations in content).

5.5 Relationship to the Viewer

Previous work in other privacy domains has found that the type of viewer or receiver of information impacts privacy comfort level in a given situation. Similarly, we believed that the user's *relationship to the viewer*, or more accurately the persona that a user maintains with a viewer [49], also impacts privacy comfort levels for the viewing of incidental information. We investigated the extent to which the relationship to the viewer impacted privacy comfort

levels by examining IIP survey results and questionnaire results from the PG1 and PG2 field studies.

5.5.1 Survey Results

Analysis of the IIP survey gives us insight into privacy comfort according to participants' relationship to potential viewers. We analyzed the mean values for the 141 participants reporting privacy comfort levels (for the *usual web browsing scenario*) for each of the five viewer types. The category spouse/significant other had the highest mean privacy comfort level (5.75), followed by close friend (5.40), parent (5.03), colleague (4.53), and supervisor (4.26). Results from a Friedman two-way ANOVA showed that differences among the mean comfort levels are statistically significant (χ^2=206.30, p<.001).

Figure 20 shows the variability of participants' privacy comfort levels according to viewer. The type of viewer may also impact the degree that privacy comfort levels change according to the amount of control retained and the content sensitivity of the scenario. As was earlier seen in Figure 18, the impact of control and sensitivity of scenario did not change comfort levels for spouse/significant other to the same extent as for other potential viewers. These results are consistent with previous information privacy research such as [115] with

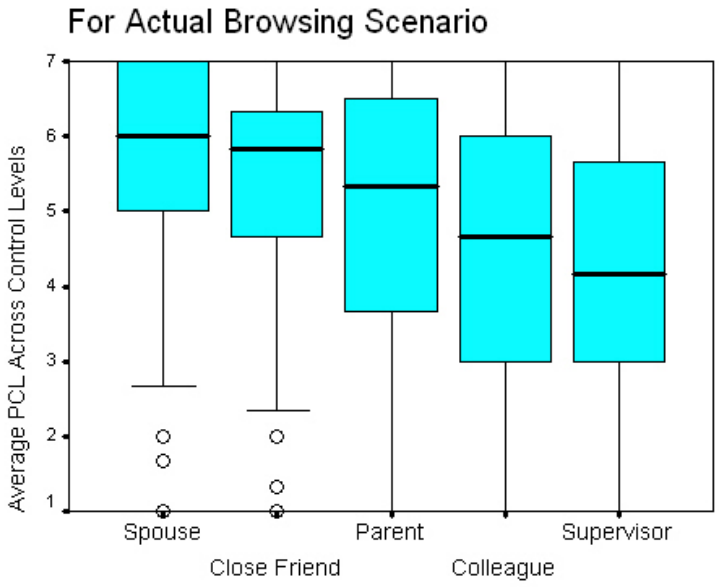


Figure 20. Box plots showing the variability of average privacy comfort levels for the five types of viewers (for usual web browsing scenario).

respect to the relative comfort levels between categories of information receivers. However, the categories used in our survey were relatively broad. Even within a viewer category, levels of trust and sharing may fluctuate according to the nature of the individual relationships. Furthermore, trust and sharing may fluctuate over time depending on the history of interpersonal interactions.

As previously discussed, all participants had people view their display at least occasionally. Trusted viewers such as spouses and close friends were regular viewers; however, some of the most frequent viewers were colleagues and supervisors, both of whom have lower overall comfort levels. It is important to note that there was variability between participants in the amount of change in privacy comfort level depending on the viewer; this factor is highly individual.

5.5.2 Results from Field Studies

We also gained some perspective about how comfortable participants in both the PG1 and PG2 field studies felt they would be for ten types of viewers seeing traces of their web activity. During the uninstall session, we asked participants to classify ten types of viewers as to what privacy level of pages they would be comfortable with them seeing (see Table 15 for question wording). It is important to note that each participant may have considered different categories of pages to fall under the classifications of public, semi-public, or private. Regardless of which types of sites participants would classify at each level, their responses give an indication of their relative comfort level for the different viewers.

Table 15. Viewer classification question.

Question: Give a classification for each of these types of viewers based on how you would feel if these viewers saw that you'd visited sites (either accidentally or on purpose) of the various types. Classify the person as "public" if you would only like them to be able to view sites you have classified as public, "semi-public" if you wouldn't mind them viewing sites you have classified as semi-public or those sites you have classified as public, "private" if you don't mind them seeing any site that you have bothered to save. ***Viewer types:*** *Parent, Spouse/Significant Other, Close Friend, Acquaintance, Colleague, Client, Supervisor, Employee/Student (Underling), Audience at a presentation, Technical Support Staff*

As can be seen in Figure 21 showing results from PG1 and Figure 22 showing results from PG2, more participants reported they would allow their spouse to see traces classified as private than any other type of viewer. Some participants would also allow a close friend or

parent to view private sites. On the other end of the spectrum, there were several viewer types that most participants would limit to viewing only those sites they'd classified as public including audience, client, underling (i.e. employee or student), supervisor and technical support staff. More participants reported that they would allow acquaintances and colleagues to view semi-public sites than would allow supervisors or underlings.

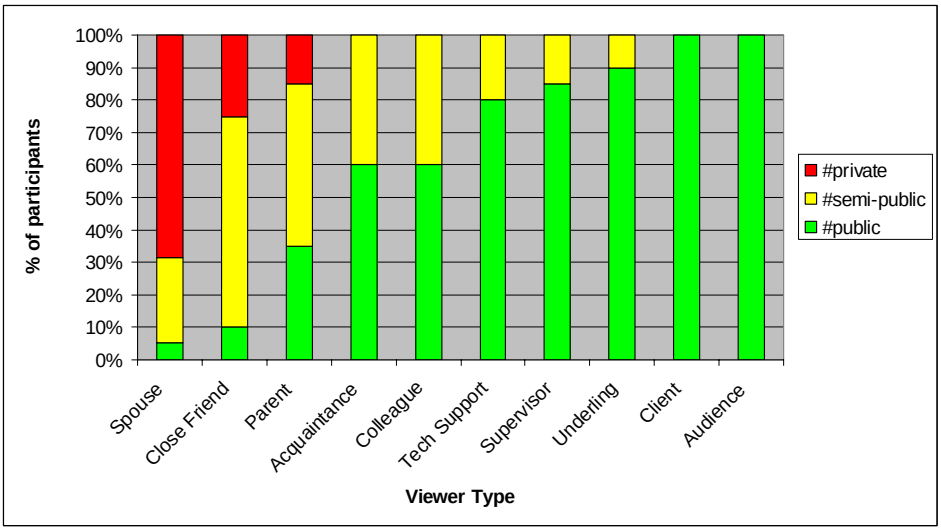


Figure 21. Viewer classification task results from PG1.

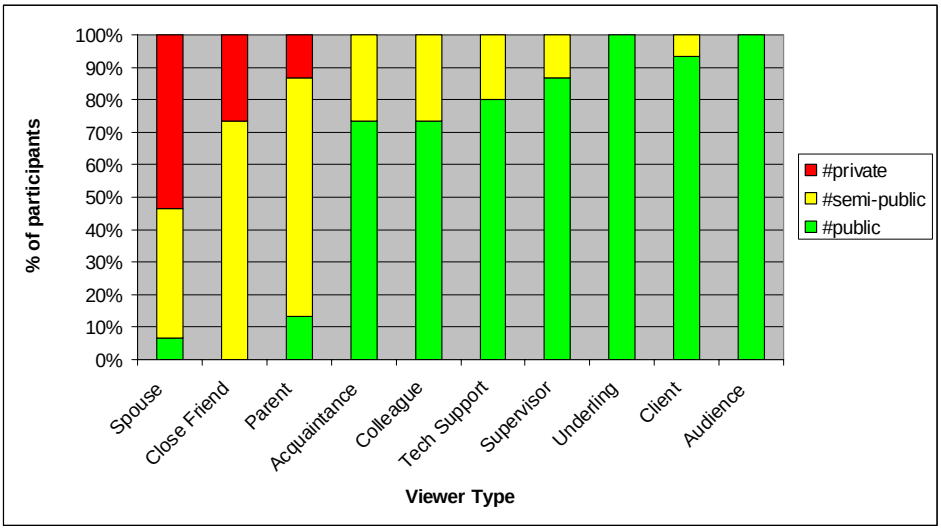


Figure 22. Viewer classification task results from PG2.

Increased privacy concerns for those in a hierarchical relationship have also been reported for location data in awareness systems [93] and content within Instant Messenger [127]. Results in the viewer classification task were highly individual with some participants

reporting they would be more restrictive than others. For example, on average those in PG1 would restrict 6/10 of viewers to sites classified as public, but there was individual variability (range 4 to 10). PG2 participants were similar (avg. 6/10 of categories public, range 4 to 8).

5.6 Level of Control

The *level of control* a person retains over what information is viewed is another important factor of privacy comfort. When considering the concept of control as it relates to privacy, most prior research discusses control in terms of which information is stored and how the information is subsequently used (e.g., what a web site may do with the personal information received). For visual privacy of incidental information within web browsers, there is no electronic transfer of the data. However, the concept of control does exist over which input is given to the applications (i.e., which input may result in incidental information being viewed). For our purposes, we refer to the level of control retained over input devices. A high amount of control (e.g., full control over input devices) should lessen privacy concerns, while lower levels of control should increase concerns. Incidental information can be hard to control due to its dynamic and temporal nature. Furthermore, users are often uncertain about what information has been saved and what may be subsequently revealed.

The IIP survey examined the impact of level of control on participant's privacy comfort levels. We analyzed the mean values for the 154 participants reporting privacy comfort levels for the *usual web browsing scenario*. When participants envisioned themselves in control of the keyboard and mouse, they had the highest privacy comfort level across the viewing audience (mean 5.50). As control was lost, the privacy comfort level decreased. Participants reported a mean privacy comfort level of 4.94 if the other person was in control of the input devices, and a mean privacy comfort level of 4.58 if the other person was left alone at the computer. Results from a Friedman two-way ANOVA showed that differences among the mean comfort levels are statistically significant (χ^2=134.74, p<.001). Figure 23 shows the variability of participants' privacy comfort levels according to level of control. As could be seen previously in Figure 17, both the viewer and the scenario impact the magnitude of the change in PCL according to level of control retained. It is also important to note that not all participants were concerned along this factor.

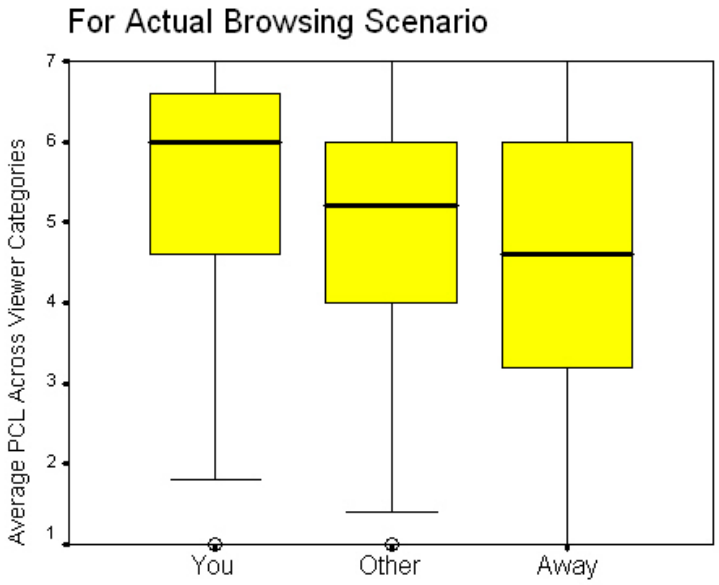


Figure 23. Box plots showing the variability of average privacy comfort levels for the three levels of control (for usual web browsing scenario).

5.7 Inherent Privacy Concerns

As previous research has described, the *inherent privacy concerns* of an individual will have a large effect on privacy comfort level in a given situation. By partitioning participants into a privacy classification scheme such as the Westin-Harris segmentation model [116], an indication of participants' inherent privacy concerns may be found. The Westin-Harris segmentation model explores consumers' confidence in how personal information is collected and used by companies. It partitions consumers into three privacy categories: privacy fundamentalist, privacy pragmatist, and privacy unconcerned. Such classifications could also be used as an initial predictor of privacy preferences. Participants that are privacy unconcerned should have relatively high comfort levels regardless of the context; similarly, fundamentalists will have relatively low comfort levels. Privacy pragmatists, however, will likely have varying comfort levels depending on visible content, level of control, and viewers.

The IIP survey did not contain any questions that attempted to determine participants' inherent privacy concerns. In order to estimate their inherent privacy concerns, the privacy comfort levels that the IIP survey participants gave for the embarrassing scenario were examined. This scenario was most likely to provoke discomfort in participants and exhibited large comfort differences by context (level of control, type of viewer). Privacy fundamentalists would be expected to have a low privacy comfort level regardless of context

and the privacy unconcerned to have a relatively high privacy comfort level. However, pragmatists might have differences in their privacy comfort level depending on the context.

In an initial attempt at discerning inherent privacy concerns through privacy segmentation of participants, an iterative k-means cluster analysis was performed on participants' median comfort levels and the magnitude of their contextual differences for the embarrassing scenario. The data vector consisted of the participant ID, their median privacy comfort level for the embarrassing scenario (across all 15 viewer/control combinations for the embarrassing scenario), and magnitude of the difference between their minimum and maximum privacy comfort level (across all 15 viewer/control combinations for the embarrassing scenario). The initial cluster centers for the iterative k-means cluster analyses (k=3) were selected to correspond with low privacy comfort level/low contextual differences (fundamentalists), high privacy comfort level/low contextual differences (unconcerned), and moderate privacy comfort level/moderate contextual differences (pragmatists). Participants clustered into the three groups as follows: 28% as fundamentalists, 64% as pragmatists and 8% as unconcerned. Figure 24 shows a conceptual diagram of the privacy segmentation.

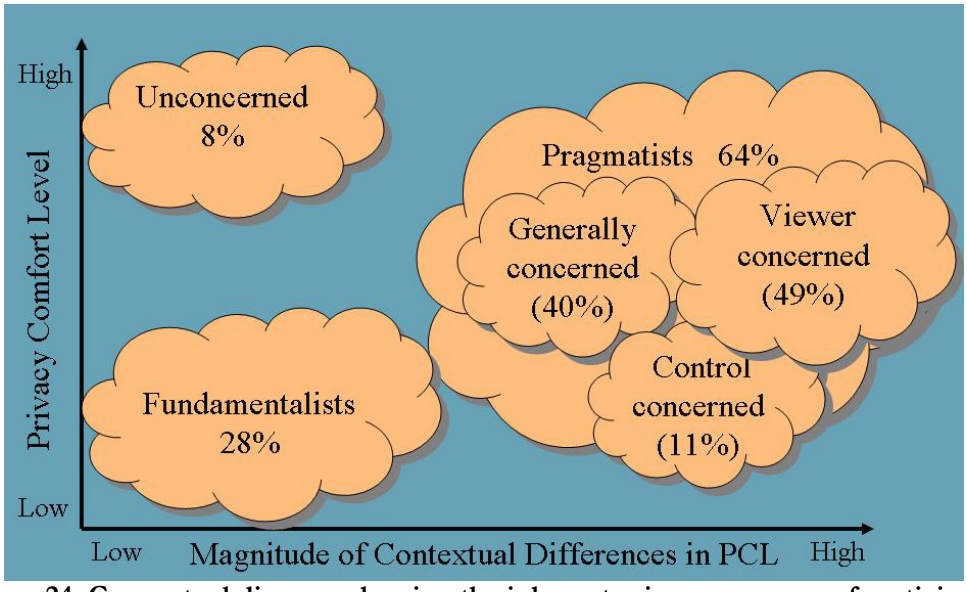


Figure 24. Conceptual diagram showing the inherent privacy concerns of participants according to their overall level of concern and the magnitude of difference in that comfort depending on the viewing context. Note: within the pragmatist cloud, the various subdivisions occur at similar levels of concerns and magnitudes of difference.

We performed a further iterative k-means cluster analyses (k=3) to determine whether the participants classified as privacy pragmatists were concerned along only one or both of the factors (level of control, relationship to viewer). The data vector consisted of the participant ID, the magnitude of difference between their minimum and maximum privacy comfort level by level of control retained over input devices (averaged across all viewer types), and the magnitude of difference between their minimum and maximum privacy comfort level by relationship to viewer (averaged across all levels of control). The initial cluster centers were selected to correspond with low level of control/high relationship to viewer differences, high level of control, low relationship to viewer differences, and moderate level of control/moderate relationship to viewer differences. When clustered solely by the magnitude of their differences in privacy comfort level, 40% of the 100 participants in the pragmatists cluster were concerned along both factors, while 60% were concerned along only one factor (control or viewer). For those 60 pragmatists concerned with only one factor, most (49) had high differences for viewers and low differences for level of control and the converse was noted for the remaining 11 pragmatists. Note that within the pragmatist cloud in the diagram, the various subdivisions occur at similar levels of concerns and magnitudes of difference.

In a similar fashion to the subdivision of pragmatists in the consumer privacy domain into identity concerned and profiling averse based on their areas of concern [139], it may be useful when modeling incidental information privacy to consider pragmatists in the categories *control concerned*, *viewer concerned*, and *generally concerned*. If we can determine an individual's inherent privacy concerns we may be able to simplify configuration of a privacy management scheme. If a user is classified as a privacy fundamentalist, then the system should provide maximum privacy protection without requiring ongoing interaction. If a user is classified as a pragmatist, then knowing along which factors a user is concerned may allow the interface to be tailored to those concerns. Those that are privacy unconcerned would have little use for such a system.

Our initial examination of inherent privacy concerns looked only at the embarrassing scenario. However, many participants could be considered *content concerned* as they did not exhibit the same high concerns with other content scenarios that they did with the embarrassing scenario. Therefore, the fundamentalist group may have been inflated in our

initial privacy segmentation. Similarly, solely looking at the neutral or positive scenarios may over-classify participants as privacy unconcerned. Using the *usual browsing scenario* as the basis for classification is problematic as the sensitivity of the content will vary according to browsing practices (e.g. a participant may be concerned because they have recently conducted some particularly sensitive browsing or because they have inherently high privacy concerns).

In order to account for content, control, and viewer sensitivities we examined the magnitude of differences according to context for the averaged neutral and embarrassing scenarios as these two scenarios give the range of browsing sensitivity for most users. Figure 25 shows the distribution of the average privacy comfort level across contexts for the averaged neutral and embarrassing scenarios. The mean privacy comfort level is 4.52 (between neutral and slightly comfortable on the privacy comfort scale) with a standard deviation of 0.99. We used the mean privacy comfort level as the cutoff between uncomfortable and comfortable for the purpose of segmenting participants Those participants with a privacy value below the mean were classified as being uncomfortable and those above the mean were classified as being comfortable.

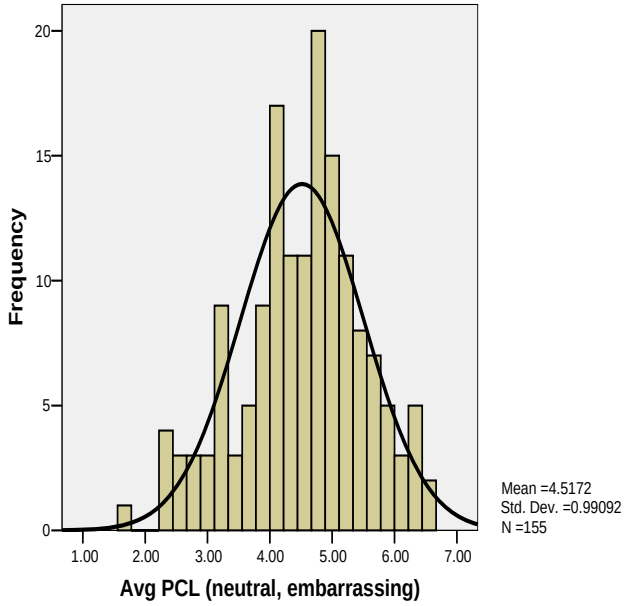


Figure 25. The average privacy comfort level across the neutral and embarrassing scenarios.

Figure 26, shows the distribution of the magnitude of differences according to context. The mean of these differences was 1.90 with a standard deviation of 0.60. As we wanted to identify those participants with very low levels of contextual differences (i.e., fundamentalists, unconcerned), we used 1 standard deviation below the mean (less than 1.30) as the cut-off point.

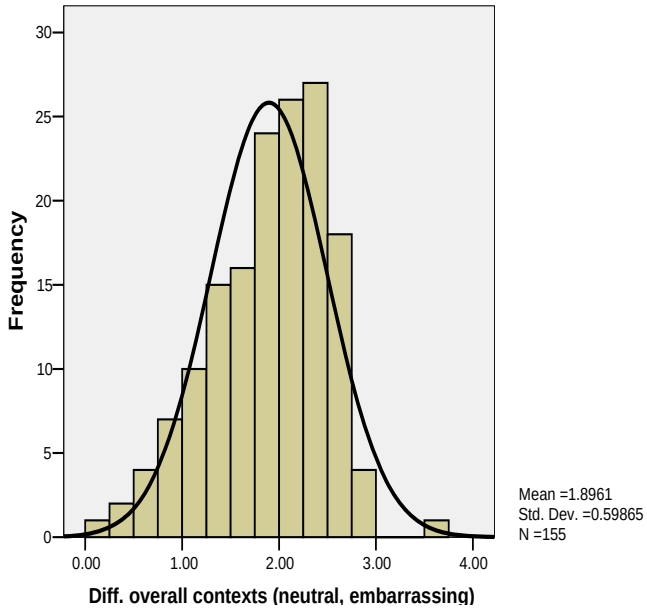


Figure 26. Magnitude of the differences in privacy comfort level for the averaged neutral and embarrassing scenarios.

We segmented participants according to their comfort and magnitude of contextual differences. Table 16 shows the results of the participant segmentation. The majority of participants (83.9%) using this segmentation scheme are considered to be pragmatists, having a privacy comfort level that changes depending on the viewing context. Those participants with a low magnitude of contextual differences were divided into unconcerned (10.3%) and fundamentalists (5.8%) based on their overall level of comfort.

Table 16. Results of participant segmentation using the average of the neutral and embarrassing scenarios.

Level of Comfort **Contextual Differences**	**Comfortable**	**Uncomfortable**
Very low differences	Unconcerned (16 - 10.3%)	Fundamentalist (9 – 5.8%)
Other	Pragmatist (130 – 83.9%)	

In order to subdivide pragmatists as being overall concerned (equally concerned across contexts) or having some subset of concerns, we examined the effect of scenario, level of control, and viewer on their privacy comfort levels. While we used normalized data to find appropriate thresholds for the overall segmentation, we must consider that control, viewer, and scenario do not impact privacy comfort level equally as can be seen in the histograms shown in Figure **27**.

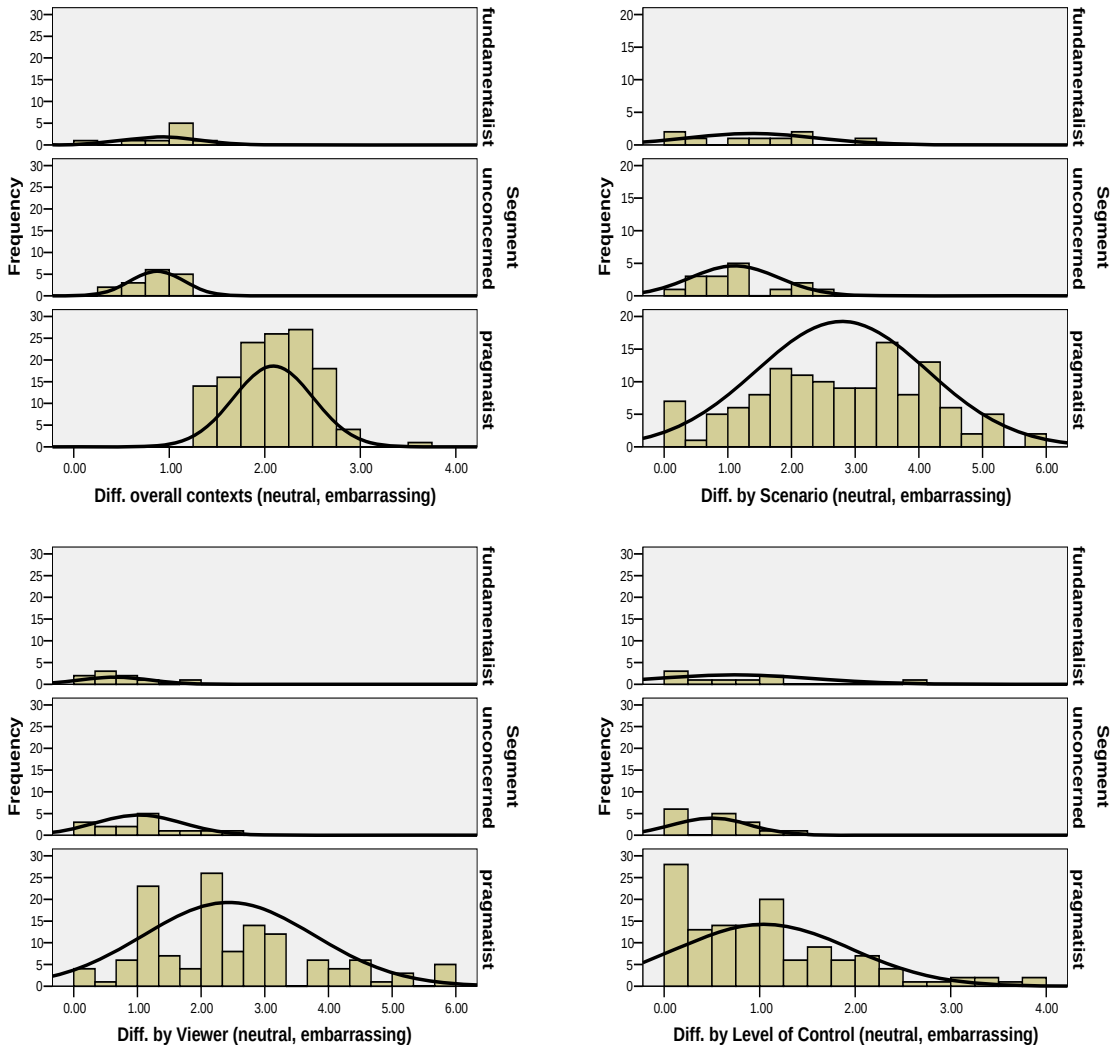


Figure 27. Magnitude of contextual differences by privacy segmentation for overall contexts (top left), scenario (top right), viewer (bottom left), and level of control (bottom right).

As level of control has a lesser impact on privacy comfort level than viewer or scenario, it would be inappropriate to use thresholds based on normalized values for each factor. Instead, we calculate a normalized value for each participant for the relative impact

each factor had on their PCL. We do this by summing the differences across all three factors and then calculating for each factor the percentage of the difference. For example, a participant with a difference of 1 for control, 1 for viewer, and 4 for scenario has 67% of their difference due to scenario, 16% for control, and 16% for viewer.

We ran a non-iterative k-means classification (k=7) of these percentages to determine for our pragmatists the nature of their concerns. The data vector consisted of participant ID, their percentage of differences in PCL for level of control, their percentage of differences in PCL for viewer, and their percentage of differences in PCL for scenario (computed as described above). The initial cluster centers given were based on 100% concerned along a factor (control concerned, viewer concerned, scenario concerned), 50% concerned along 2 factors (control/viewer concerned, viewer/scenario concerned, control/scenario concerned) or evenly split (33% each) along all 3 factors (overall concerned). As can be seen in Table 17, the majority of pragmatists (43.8%) can be considered viewer/scenario concerned, while 26.9% are concerned across all contexts.

Table 17. Final cluster centers for pragmatists.

	Cluster						
	1 Control concerned	*2 Viewer concerned*	*3 Scenario concerned*	*4 Control/ Viewer concerned*	*5 Viewer/ Scenario concerned*	*6 Control/ Scenario concerned*	*7 Overall concerned*
Control (% of total differences)	1.00	.07	.07	.38	.07	.33	.29
Viewer (% of total differences)	.00	.83	.16	.55	.42	.13	.36
Scenario (% of total differences)	.00	.10	.77	.08	.51	.54	.34
Total participants (130)	1	8	13	5	57	11	35
% of pragmatists	0.8%	6.2%	10.0%	3.8%	43.8%	8.5%	26.9%

The final breakdown of privacy pragmatists by the nature of their concerns can be characterized in the Venn diagram shown in Figure 28. Segmenting participants this way illustrates how the situational factors of level of control, relationship to the viewer and sensitivity of the content impact privacy comfort level, but gives no perspective on the

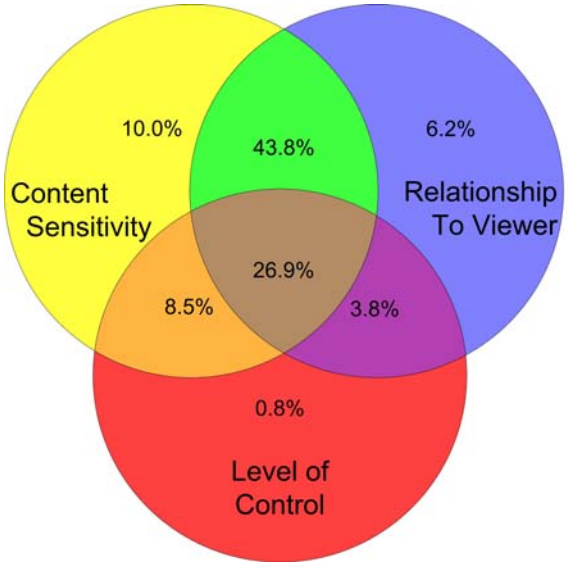


Figure 28. Venn diagram showing privacy pragmatists subdivided for their privacy concerns by the relative impact of level of control, relationship to viewer and content sensitvity.

overall level of comfort (with the exception of those considered to be privacy fundamentalists or unconcerned).

Alternatively, we could segment our IIP survey participants in a similar fashion to Sheehan [137], taking into consideration the overall level of privacy concern when subdividing the pragmatists. We initially divide our participants as before based on their overall level of privacy concern (the average of their privacy comfort levels for the neutral and embarrassing scenarios) and the magnitude of their difference to partition the pragmatists from those with little contextual differences (i.e. the privacy unconcerned and fundamentalists). We then subdivide the pragmatists according to their overall level of privacy concern, using Sheehan's terminology of circumspect for those exhibiting a higher overall level of comfort and wary for those exhibiting a lower overall level of comfort. This segmentation results in a fairly even division of the pragmatists as shown in Table 18.

Table 18. Inherent privacy concerns, pragmatists subdivided according to level of concern.

	Level of Comfort	
Contextual Differences	**Comfortable**	**Uncomfortable**
Very low differences	Unconcerned (16 - 10.3%)	Fundamentalist (9 – 5.8%)
Moderate to high differences	Circumspect (66 – 42.6%)	Wary (64 – 41.3%)

The first segmentation (focusing on the impact of contextual factors) could be used by an intelligent system to simplify which contextual factors to take into account when determining a privacy level. The second type of segmentation (focusing on the level of privacy concern) may be useful to help determine which content should be shown.

5.8 Re-examining the Factors of Incidental Information Privacy

Our privacy segmentation analysis shows how individuals are concerned along differing privacy factors, with varying amounts of their privacy comfort level in a given situation dependent on the situational factors of content sensitivity, level of control, and relationship to the viewer. While our initial conceptual model of the factors of incidental information privacy showed inherent privacy concerns impacting privacy comfort level along-side of the other three factors, it may be better to consider inherent privacy concerns as encompassing the other factors (as shown in the conceptual diagram in Figure 29). The inherent privacy concerns appear to affect not just the level of comfort, but also appear to moderate the effect of the other three factors (content sensitivity, level of control, relationship to the viewer).

If we can classify users' inherent privacy concerns with respect to the visual privacy of incidental information, we could potentially increase the effectiveness of privacy management systems that automate the filtering of appropriate content. It may be possible to establish an appropriate weighting mechanism for each user for the contextual factors of viewer, content sensitivity, and level of control. Then in a given situation, an appropriate level of concern could be calculated which could then be used to filter the traces of prior activity appropriately.

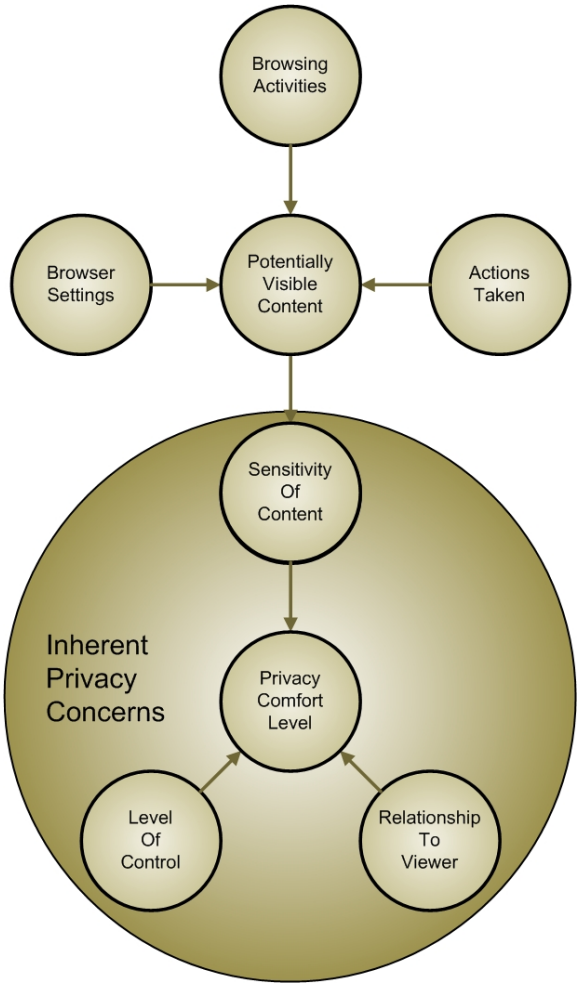


Figure 29. Revised conceptual model of the incidental information privacy factors.

5.9 Summary

In this chapter we have presented general results concerning incidental information privacy in web browsing. Our examination of the scope of the problem in section 5.1 revealed that most of our participants had occasions when others could view their displays as they worked closely together and that many of them would take steps to limit what traces of incidental information may be visible in their web browsers if given advanced notice of collaboration. While the visual privacy of incidental information in web browsers is not a concern of every person that uses web browsers, we have found it to be a concern for many.

In section 5.2, we presented results showing participants' privacy concerns about their usual web browsing during the field studies. Our field study results showed patterns in our participants' privacy perceptions, both by content category and temporally, at the

browser window level. It may be feasible to leverage these patterns to assist with classification of generated traces of browsing activity. An examination of the feasibility of using these patterns to reduce the burden of content classification will be presented in Chapter 7.

Results from the three studies provided insight into the factors of incidental privacy, including the impact of level of control, relationship to viewer, sensitivity of potentially visible content, and inherent privacy concerns on an individual's privacy comfort level in a given situation. Results are summarized in Table 19, including the design implications for a visual privacy enhancing web browser. We have shown that each of these factors is highly variable, emphasizing the need for personalized or flexible solutions to privacy management in this domain. Our analyses of inherent privacy concerns led to a revised conceptual model of incidental information privacy which may be applicable to other privacy domains. Not only do the inherent privacy concerns of users impact their privacy comfort level, they also impact the extent to which the other factors (relationship to the viewer, sensitivity of potentially visible content, and level of control) are considered. Our results highlight the importance of considering inherent privacy concerns within the context of the privacy domain.

Next, in Chapter 6, we will present results concerning the impact of other dispositional factors (e.g., gender, age, technical level) and situational factors (e.g., location, device) on the factors of incidental information privacy including. This will allow us to further develop our conceptual understanding of incidental information privacy.

Table 19. Summary of results investigating the impact of primary factors of IIP on participants' privacy comfort levels, including implications for design of visual privacy management systems.

Concept	Section	Study	Findings	Design Implications
Privacy comfort level is contextual	5.3.1	IIP survey	PCLs varied depending on the viewer, the level of control retained and the sensitivity of the scenario.	A privacy management system should be able to respond to changing viewing situations (viewer, level of control over input device, sensitivity of traces in convenience features) so that content can be filtered appropriately.
Sensitivity of potentially visible traces.	5.4.1	IIP survey	PCLs for usual privacy scenario indicate that typical web browsing is considered to be as or more sensitive than the embarrassing scenario for 1/3 of participants.	A privacy management system must be able to protect that browsing which users consider to be very private (or private for some viewing contexts) while allowing them to use their convenience features for the purpose of revisitation.
			Impact of this factor on PCL is highly individual.	
	5.4.2	PG1, PG2	For the two studies (PG1/PG2) participants classified 42%/40% of pages public, 25%/20% as semi-public, 15%/25% as private, 18%/15% as don't save.	
	5.4.2	PG1, PG2	All 4 privacy levels used by almost all participants	Validates the needs for a more nuanced approach than standard public/private or save/don't save
Relationship to the viewer	5.5.1	IIP Survey	Sig. differences in PCL for the 5 viewer types.	A visual privacy management system should allow users to filter content appropriately for different types of viewers according to their individual concern for that viewer.
			Impact of this factor on PCL is highly individual.	
	5.5.2	PG1, PG2	Spouse reported most likely to be allowed to view private browsing, followed by close friend and parents (mostly semi-public), other categories most would only allow to view public sites.	
Level of control retained	5.6	IIP Survey	Sig. differences in PCL for the three levels of control.	A visual privacy management system should guard privacy when the user is in control of the input devices and also when they are away from the system.
			Impact of this factor on PCL is highly individual.	
Inherent privacy concerns	5.7	IIP Survey	Segmented participants by PCL and magnitude of contextual differences.	If users can be grouped according to inherent privacy concerns, may be able to set intelligent defaults per group or use group membership as the basis for subsequent personalization.
			Segmented participants by PCL and applicable factors of contextual differences.	

Chapter 6
Results: Examining the Impact of Browsing Context

In Chapter 5 we presented results showing how participants' privacy comfort level in a viewing situation depended on their relationship to the viewer, level of control over input, and sensitivity of the content generated and was furthermore influenced by their inherent privacy concerns. This chapter presents further analysis of the data collected during our exploratory studies (Figure 30 shows the scope of this analysis). We investigated how browsing activities, web browser settings, and actions taken to preserve privacy combine to produce the potentially visible content in web browsers. We also explored how these browsing behaviours and privacy concerns in a given situation are impacted by the

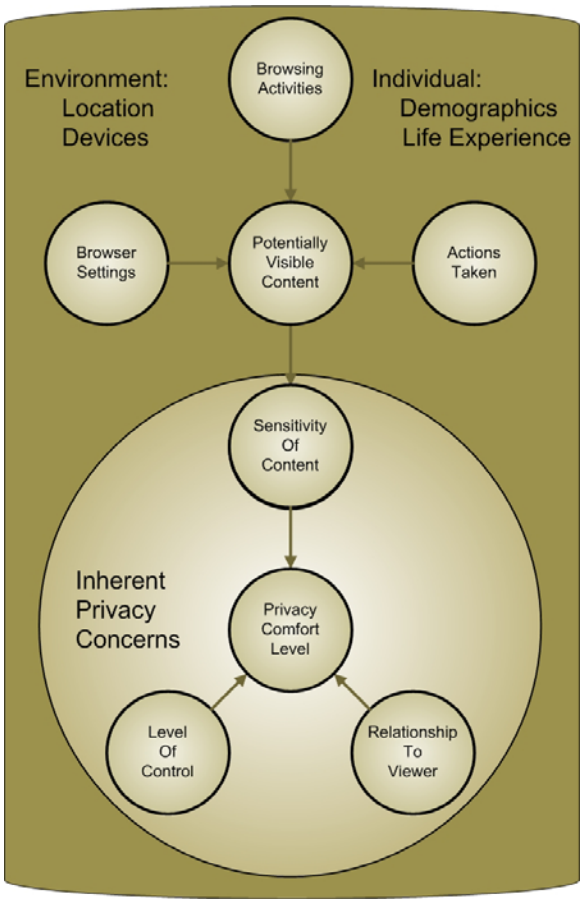


Figure 30. Conceptual model of the environmental context and an individual's attributes shaping web browser activities and privacy concerns.

environmental context (location, device) and an individual's attributes including demographics and life experience.

Privacy comfort in a given situation depends not only on a person's disposition to privacy, but also on the context of the situation. While inherent privacy concern indicates someone's overall privacy preferences, the situational context will determine what decision is made as to which information is appropriate to reveal [78, 150]. For example, in a study examining online disclosure of information, independent pathways were found for the dispositional variable of participant's general privacy concerns as well as the situational variables of perceived privacy (in terms of anonymity and confidentiality) and participants' trust in the receiver of the information [78]. To develop effective approaches to incidental information privacy management, we must understand as much as possible how the dispositional and situational variables affect privacy comfort level in a given situation.

Situational variables are those which may vary according to the usage environment. Situational variables for incidental information privacy in web browsers include the computing device used and the location of use. Furthermore, within each location there may be other variables that change. These include the current role of the user, social norms for the location, rules for personal web browsing activities, and different types of viewers of the display and users of the device. These variables may constrain or shape browsing activities and privacy concerns. For example, someone with both a home and a work computer may refrain from conducting many personal activities while at work, while someone with only access at work may conduct a broader range of activities. A laptop user may perform the majority of their browsing activities on their laptop, but their viewing concerns may change as they move between different locations with different social norms. One's browser settings and preventative actions taken may also change depending on the usage environment. Beyond which traces are potentially visible as a result of these changes, the perceived sensitivity of the traces may also change as a result of the viewing situation. The cost and benefit of disclosure depends on the specifics of each situation [78].

Dispositional variables are those that affect an individual's disposition to privacy. A person's demographics such as age and gender may affect their privacy disposition. However, disposition to privacy, what we have been referring to as their inherent privacy concerns, is also grounded in an individual's life experience. For example, their technical

level or computer experience may impact their inherent privacy concerns. Additionally, dispositional variables may moderate the effect of the situational variables. Someone with strong inherent privacy concerns may always be very private, someone with weak concerns may be less private, others may be more pragmatic and may more often modify their privacy comfort and browsing activities in response to the state of the environment.

In this chapter we explore the inter-relationship of dispositional and situational variables and their impact on participants' activities and privacy concerns. We present results from the IIP survey and the contextual data captured during the PG2 field study. In section 6.1, we examine the impact of dispositional variables such as our participants' demographics and life experiences on their inherent privacy concerns. In section 6.2, we examine the impact of situational variables such as location and device on inherent privacy concerns. In section 6.3, we examine the impact of the environmental context on the overall application of privacy levels by participants in the PG2 field study. We then breakdown the possible causes for the differences found in the subsequent sections. We examine how the environmental context affects browsing activities (section 6.4), browser convenience features settings (section 6.5) and the post-browsing privacy preserving actions taken if given advance notice of collaboration (section 6.6), all of which contribute to what content is potentially visible within the traces. Finally, in section 6.7, we examine whether the same types of content are perceived as having differing privacy concerns across usage contexts.

6.1 Dispositional Variables and Inherent Privacy Concerns

We examined the IIP survey participants' inherent privacy concerns with respect to the dispositional variables of age, gender, education level, technical level, and computer experience. For each variable we compare our results with previous research that has investigated differences among privacy concerns according to various demographic or dispositional variables. For example, O'Neil [113] used data from the 1998 GVU survey and examined differences among on-line privacy concern by sex, education level, income, and race. Sheehan [137] examined the demographic profiles of four groups of participants (unconcerned, circumspect, wary, alarmed) segmented by their online privacy concerns regarding the collection and use of personally identifiable data.

In section 5.7 we subdivided pragmatists according to their level of contextual differences (i.e. the magnitude of their differences in privacy comfort level) according to the impact of viewer, level of control, and content sensitivity. This was a useful division for the purpose of understanding how the situational factors of viewer, level of control, and content sensitivity impacted participants. However, the resulting small group sizes made analysis of the effect of dispositional variables through CHI-Square tests inappropriate as there were many small or empty cells. For the purpose of our current analysis, we separately investigated the two components of inherent privacy concern: the *level of the contextual differences* (which ranged from 0 to 3.53, higher values indicate larger contextual differences) and the *overall privacy comfort level* for the average of the neutral and embarrassing scenarios (which ranged from 1.73 to 6.50, higher values indicate higher comfort levels).

For those variables where we have continuous data (e.g. age), we performed Pearson correlation analyses using the raw scores for the level of contextual differences and overall privacy comfort level. For the categorical variables (e.g. technical level), we performed Chi-Square analyses to determine if membership differed between low, medium, and high levels of the contextual differences and overall privacy comfort level. In order to avoid issues with small cell size, the divisions for low, medium, and high were selected to provide three groups of equivalent size (shown in Table 20). We also performed t-tests and analysis of variance where appropriate to further examine the differences between categorical groups.

Table 20. Cut-off points for low, medium, and high levels of contextual differences and overall privacy comfort level.

Category	Level of Contextual Differences	Overall Privacy Comfort Level
Low	0.00-1.70 (n=51)	1.73-4.13 (n=52)
Medium	1.71-2.23 (n=52)	4.14-4.95 (n=51)
High	2.24-3.53 (n=52)	4.96-6.50 (n=52)

6.1.1 Age

We found no correlation between age and the overall privacy comfort level or the level of contextual differences. Age was not normally distributed and we observed differences in the distribution of participants' age by gender (Figure 31). The 89 male participants in the IIP survey had a mean age of 29.4 and many were in their early twenties, while the 66 female participants had a mean age of 34.2 and were more evenly distributed.

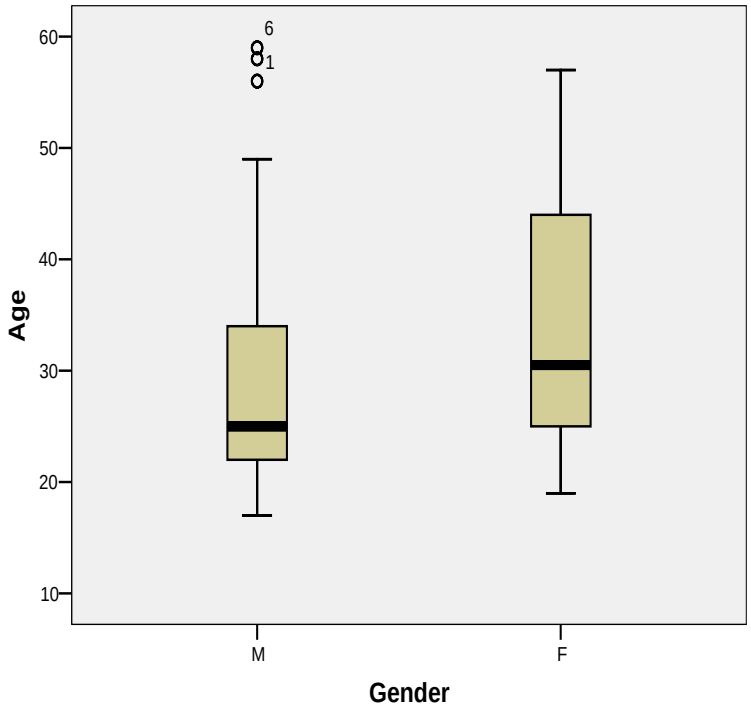


Figure 31. Box plot showing distribution of age by gender.

This may have impacted our ability to find correlations between age and inherent privacy concerns.

Some connection between age and privacy or security concerns has been reported in the literature. In a 1998 survey, Westin and Maurici [151] found that the youngest and oldest segments of the population expressed fewer offline and online consumer-business privacy concerns than the general population. Age was a factor examined by Dourish et al. [42] as they investigated the security practices of 20 participants, recruited from an academic institution and an industrial research lab, whose jobs required a measure of confidentiality. Their results indicate that age and experience were correlated with attitudes towards security. In particular, younger subjects with a longer exposure to computers (i.e. childhood exposure) were observed to have greater confidence in their abilities and were more pragmatic about their security needs and were more likely to examine the costs as well of the benefits of security. These younger subjects were also more nuanced in their discussions of security scenarios. More recently, Hutchings et al. [73] had 28 participants in an online survey rate the perceived sensitivity of risk when using public devices using a scale from 0 to 5. Overall, participants indicated they were mildly to somewhat concerned (avg. rating 2.6); however those participants over the age of 50 had increased concerns (avg. rating 3.8) while those in

their 20's has an average level of concern of only 1.8. The authors speculate that the older participants may have had a lower comfort with technology, an increased awareness of risk, or had more to lose as a result of a privacy violation.

6.1.2 Gender

A Chi-Square analysis revealed a marginal difference in the distribution of overall privacy comfort level by gender ($\chi^2(2, N=155) = 5.349$, $p=0.069$). Males were most likely to be in the top third of participants for comfort, followed by the bottom third and least likely to be in the middle third; while females were most likely to be in the middle third, followed by the bottom third, and least likely to be in the top third. Despite these differing distributions, a t-test revealed no significant differences in the mean overall privacy comfort level by gender (male = 4.54, female = 4.48).

A Chi-Square analysis revealed a significant difference in the distribution of the level of contextual differences by gender ($\chi^2(2, N=155) = 6.422$, $p=0.040$). Males were most likely to be in the bottom third and least likely to be in the top third, while the reverse was true for females. A t-test revealed significant differences in the mean level of contextual differences (male = 1.8, female = 2.02; $t(153) = -2.317$, $p=.022$).

Prior research in related privacy domains has found some differences in privacy concerns attributed to gender. O'Neil [113] found small but significant differences in concern depending on gender: 83.9% of women and 79.2% of men reported being very or somewhat concerned about privacy, and 55.9% of women and 52.6% of men reported being very concerned. Also, while both women and men were found to value privacy over convenience, gender again played a small role: more women valued privacy (82.9%) than men (76.6%). Cvrcek et al. [38] examined the value of location privacy for mobile devices according to gender. They found gender did not play a role when location information was given for academic purposes (ratio 1:1), but females placed a higher value on their privacy than males when the information was given for commercial purposes (ratio 1.4:1), and an even higher value when the information was given for commercial purposes over the long-term (ratio 1.8:1). Sheehan [137] found that in 5/15 online marketing situations, female participants rated their level of concern at a significantly higher level than the male

participants did. In particular, women appeared to be more concerned about unsolicited email and secondary usage of information than men.

6.1.3 Education Level

Participants in the IIP survey reported their highest educational level achieved using one of seven levels: less than high school, high school, technical school, some university, Bachelor's degree, some graduate school, and graduate degree. For this analysis we reclassified participants as having less than a university education (n=26), an undergraduate education (n=77), and a graduate level of education (n=52).

A Chi-Square analysis revealed a marginal difference in the distribution of overall privacy comfort level by educational level ($\chi^2(4, N=155) = 7.988$, $p=0.092$). There was no clear pattern for those with less than a university education; however those at the undergraduate level were most likely to be in the bottom third for privacy comfort, while those at the graduate level were most likely to be in the top third.

A Chi-Square analysis revealed no significant differences in the distribution of the level of contextual differences by education level.

O'Neil [113] also found differences according to educational level. Results were similar to ours as the levels of concern were not found to systematically increase or decrease according to the level of education achieved.

6.1.4 Technical Level

The technical level of participants was assigned based on their declared field of study or their reported position of employment. We assigned three levels: those in a computer science related field were classified as technical (n=50), those in a scientific field were classified as semi-technical (n=25), and those in a non-science or computer science field were classified as being non-technical (n=61). We were unable to classify 19 participants.

Chi-Square analyses revealed no significant differences in the distribution of either overall privacy comfort level or the level of contextual differences by technical level using the three levels (technical, semi-technical, non-technical). An examination of just those participants classified as either technical or non-technical, also showed no significant difference in the distribution of participants for the overall privacy comfort level of the

magnitude of the contextual differences. However, there was a small but significant difference in the mean level of the contextual differences (technical = 1.78, non-technical = 2.04; t(109) = -2.303, p=.023).

Technical level had an inter-relationship with age. A one-way analysis of variance found that the mean age of participants varied by technical level ($F_{2, 133}$ = 10.375, p=.000). Post hoc analysis, using a Bonferroni-corrected alpha level of p < .013, showed that the technical participants (mean age 27.67) were significantly younger than both semi-technical participants (mean age 32.92) and non-technical participants (mean age 36.18); however, no significant difference was found between the semi-technical and the non-technical participants. Similarly, technical level was found to be inter-related with gender. Chi-Square analysis revealed a significant difference in the distribution of participants between technical levels by gender (χ^2(2, N=36) = 35.579, p=0.000). Males were more likely to be classified as technical (48 males, 13 females), while females were more likely to be classified as non-technical (11 males, 39 females); there was no difference in the distribution of males and females for those classified as semi-technical (14 males, 11 females). Figure 32 shows the distribution in age by technical level, split by gender. Technical participants tended to be younger males, while non-technical participants tended to be females with a broad age range.

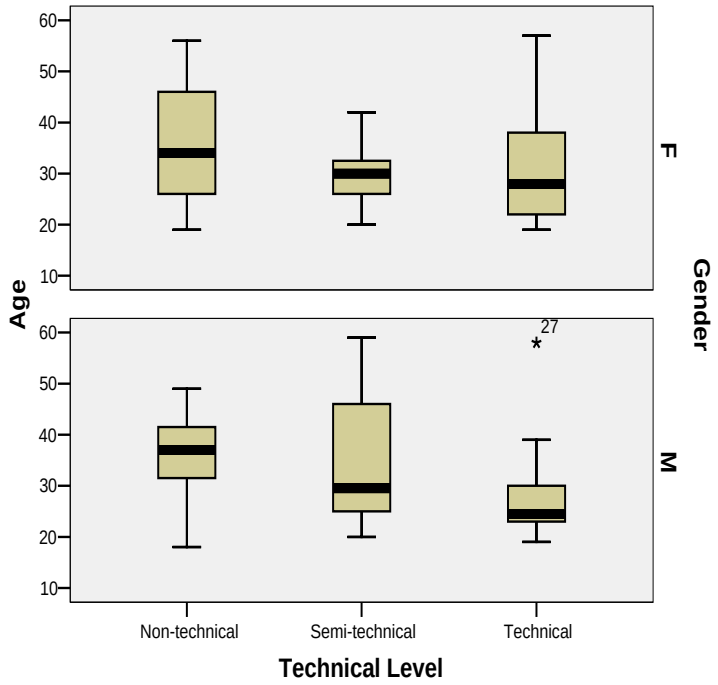


Figure 32. Box plot showing distribution of age by technical level, split by gender.

In prior research, Patil and Kobsa [127] found a significant positive effect of technical savviness on privacy concerns. They measured technical competency based on answers that participants gave to technical questions.

6.1.5 Computer Experience

There was a modest positive correlation (r=.240, p=.015, n=154) between years of computer experience and the overall privacy comfort level. No correlation was found between computer experience and the level of contextual differences. As computer experience was also found to have a moderate correlation with age (r=.523, p=.000, n=154), we performed a partial correlation controlling for the age of the participant. We found that the strength of the correlation between computer experience and overall privacy comfort level increased slightly (r=.255, p=.001) when age was controlled. No correlation was found between computer experience and the level of contextual differences.

6.1.6 Summary

Table 21 shows a summary of the impact of dispositional variables on inherent privacy concerns, as well as their interrelationships. Participants who were female had greater contextual differences than those who were male as did those who were non-technical rather than technical. Participants with a greater amount of computer experience had a higher overall privacy comfort level. However, it is clear that the dispositional variables we examined account for only a small portion of the variability of inherent privacy concerns.

Table 21. Impact of dispositional variables on inherent privacy concerns.

Dispositional Variable	***Impact on Level of Contextual Differences***	***Impact on Overall Privacy Comfort Level***
Age		
Gender	Women > men	Marginal, no pattern
Education Level		Marginal, no pattern
Technical Level	Non-technical > technical	
Computer Experience		Positive relationship

For our survey population, age, gender, and technical level were found to inter-relate. In order to more fully investigate the relationship of these dispositional variables on privacy concerns, participants should be recruited so that they are balanced in terms of age, gender and technical level. Given that our non-technical participants were primarily women, it is difficult to determine whether both factors impact the amount of contextual differences.

6.2 Situational Variables and Inherent Privacy Concerns

There are several situational variables which may constrain or shape participants' browsing activities and inherent privacy concerns. People use web browsers at work, at school, at home and in other locations such as coffee shops and libraries. Within these locations there are different devices available for use. At some locations, people have access to a dedicated computer; at other locations they may have to share a computer. Some people move between locations with their laptops or other web-enabled mobile devices. There are also different rules and social norms associated with each location. Workplaces often restrict personal browsing or limit it to break times. Some workplaces physically block access to sites with potentially objectionable content or have formal policies against such browsing. Other environments are more amenable to personal browsing, although social norms may limit what activities people will engage in if their screens can be viewed. Furthermore, some people choose to erect more well-defined boundaries between their work and non-work roles [114]. In each location there are also different types of people that may be able to view one's computer display or use the device itself. There is also the role of the person conducting the browsing to consider. Potential consequences of privacy violations may be dependent on the persona one is trying to maintain.

We next present results from the IIP survey pertaining to the impact of situational variables on inherent privacy concerns. Where relevant, we include the inter-relationship between the dispositional variables and these situational variables. We use the same breakdown as in Table 20 for low, medium, and, high levels of contextual differences and overall privacy comfort when examining the distribution of participants by categorical variables.

6.2.1 Location

Our IIP survey asked participants to indicate their primary location of web browser use from a choice of work, home, and school. When recruiting participants, we were targeting people from each location, although often a person targeted in one location may have had another primary location of use. For instance, we contacted various mailing lists at Dalhousie University in an effort to recruit participants conducting the majority of their browsing at school; however this often resulted in participants who conducted the majority

of their browsing at home. We had hoped to recruit equal numbers at each location. In the end, we stopped recruiting with 88 participants reporting that they conducted the majority of their browsing at home, 44 at work, and 23 at school. For some of the analyses in this chapter, we collapse participants to the categories of home (n=88) and away (n=67).

Chi-Square analyses revealed no significant differences in the distribution of either overall privacy comfort level or the level of contextual differences by the majority location of use, regardless of whether participants were divided by home/work/school or home/away. Few of the IIP survey participants used web browsers in a single location: only 2.9% of participants reported never using a web browser at home and only 6.5% reported never using one when away from home. This use of web browsers in multiple locations regardless of where the majority of web browsing occurred may have impacted our ability to detect any differences in inherent privacy concern as a result of conditioned privacy experiences in one location or the other.

6.2.2 Devices in Use

Our IIP survey asked participants to indicate the type of device on which they performed the majority of their browsing activity (i.e., their *majority computer of use*) from a choice of laptop computer, single user desktop computer, and shared desktop computer. As with the primary location of use, we were attempting to balance participants by device. However, we had a difficult time recruiting participants who performed the majority of their browsing on shared desktop computers. In the end, we stopped recruiting with 78 participants who performed the majority of their browsing on a single user desktop PC, 60 participants who performed the majority of their browsing on a laptop computer, and 17 participants who performed the majority of their browsing on a shared desktop PC. We also asked participants whether or not they used a laptop computer, 98 (65.9%) of our participants did some of their browsing on a laptop computer. Furthermore, 81 (87.9%) of these laptop users reported using their laptop computers in multiple locations.

Chi-Square analyses revealed no significant differences in the distribution of overall privacy comfort level by majority computer of use or by whether or not a participant reported using a laptop computer. However, significant differences in the distribution of overall privacy comfort level were found for laptop users depending on whether or not they

used their laptop computers in multiple locations ($\chi^2(2, N=98) = 10.224$, $p=0.006$). Participants who did not use their laptop computers in multiple locations were more likely to have a high overall privacy comfort level, followed by a medium privacy comfort level, and then low. The reverse was true for participants who reported using their laptops in multiple locations.

Chi-Square analyses revealed no significant differences in the distribution of the level of contextual differences by the majority computer of use, by laptop use, or by laptop use in multiple locations.

Most participants reported using multiple computers for web browsing: 92.3% reported using more than one computing device, 38.7% more than two, 16.1% more than three, and 6.5% more than four. A variety of computer types were regularly used: laptops, single user PCs, and shared PCs, both at home and away from home (see Table 22 for a breakdown by location and device type). This diversity of computers in regular use may have impacted our ability to detect significant differences by the majority computer in use.

Table 22. Percentage of participants that use each device type in each location.

	Single User PC	Shared PC	Laptop	Other
Home	33.6%	41.8%	50.0%	7.5%
Away	51.2%	38.8%	38.0%	3.1%

We examined participants' inherent privacy concerns by the total number of devices they used. We coded the total number of devices used for web browsing across locations as low (1-2 devices), medium (3-4 devices) and high (4 or more devices). Chi-Square analysis revealed significant differences in the distribution of overall privacy comfort level by the total number of devices ($\chi^2(4, N=155) = 13.236$, $p=0.010$). Participants with a higher number of total devices were more frequently found in the cells with low overall privacy comfort levels, while those with a low number of total devices were more frequently found in the cells with high overall privacy comfort levels. We should note that the total number of devices also correlates with both gender and technical level. Therefore, differences in inherent privacy concerns may be less as a result of situational usage patterns and more as a result of the dispositional factors that lead to those situations of use. No differences in distribution were found when examining the amount of contextual differences by the number of devices.

6.2.3 Potential Viewers/Users of Display

Different categories of viewers may be able to view the computer display in different usage environments. For example, a spouse or parent may be less likely to view a display at one's workplace than at home. An employer or colleague may be more likely to view a display at work. Close friends may be found in either location. Similarly, different categories of people may be able to sequentially use one's computer in the different locations.

As presented in section 5.1.1, participants indicated the frequency with which ten categories of people could view their display or use their computer. We examined whether viewer and user frequency had an impact on participants' inherent privacy concerns. We anticipated that participants with frequent viewers/users, particularly those viewers/users in a hierarchical relationship (i.e., supervisor, employee) may have heightened privacy concerns. However, Chi-Square testing of overall privacy comfort level and the level of contextual differences by each viewer/user type revealed no systematic patterns. Only two tests were significant (overall privacy comfort level by technical support viewer, amount of contextual differences by close friend), but neither showed monotonic patterns upon inspection of the cell distributions.

6.2.4 Role of Person

One aspect of location that our exploratory studies did not investigate directly is the role of the person in the environment. Since Goffman's early work ([49]), there has been a great deal of social research about how individuals maintain different personas for different situations. Indeed, many approaches to privacy management are role based, providing users with an opportunity to create different persona's for different usage contexts (e.g. Lederer et al.'s metaphorical faces [93]).

There may be different consequences of privacy violations depending on the individual's role. For example, a job search viewed in the AutoCompletes may be expected for a co-op student, while the same search may be considered to be very sensitive for an employee or a supervisor. One of our participants in the PrivateBits study (presented in Chapter 8) discussed how different content would be appropriate in different work situations depending on her role, even if the viewer was the same person. She felt that it would be highly inappropriate if traces related to personal browsing were displayed during a

work-related meeting; however, if traces of the same browsing were displayed in a social context (i.e. when on lunch break), she would feel much more comfortable.

6.2.5 Summary

Table 23 summarizes the results presented in this section as we investigated the impact on situational variables on two factors of participants' inherent privacy concerns: their overall privacy comfort level and the magnitude of their contextual differences for their privacy comfort level. We did not find any differences in inherent privacy concerns based on which device was used for the majority of web browsing. Similarly, analysis of the relationship between majority location of use and inherent privacy concerns revealed no significant differences. It may be the case that measures of a different granularity would have yielded different results. For example, we did not enquire about the rules and social norms associated with locations of use. It could be that participants working in very formal environments with strict policies about internet usage have heightened privacy concerns for any browsing that they do conduct, or that they have very low concerns as they self-regulate their browsing accordingly. It would be interesting in future work to investigate how the social norms of a location and the desired personas in that location impact incidental information privacy.

Table 23. Impact of situational variables on inherent privacy concerns.

Situational Variable	Measure	Impact on Level of Contextual Differences	Impact on Overall Privacy Comfort Level
Location (6.2.1)	Majority location of use (home/away)	None found	None found
	Majority location of use (home/work/school)	None found	None found
Devices in Use (6.2.2)	Majority computer in use (single user desktop/shared desktop/laptop)	None found	None found
	Laptop user	None found	None found
	Laptop user in multiple locations	None found	Lower PCL if laptop used in multiple locations
	Total # devices used (correlated with gender and technical level)	None found	Lower PCL for users with higher # of devices in use
Potential Viewers/Users of Display (6.2.3)	Frequency with which 10 categories of people could view display/use computer	Only for close friend (no pattern)	Only for technical support (no pattern)

We found correlations between regular usage situations and inherent privacy concerns for only a subset of the situational factors investigated. We did find that laptop users that conducted browsing in multiple locations had heightened privacy concerns with respect to overall privacy comfort level as did participants who used a large number of devices. However, an examination of frequency of viewers and users of displays did not reveal significant differences in inherent privacy concerns.

6.3 Impact of Context on Privacy Level Application

For the next four sections, in addition to IIP survey data, we examine the data captured during the PG2 field study. As described in Chapter 3, we recruited five non-technical laptop users, five non-technical desktop users, and five technical desktop users and installed logging software on their regular computer(s). Participant IDs were assigned based on their technical ability and device use (TD = technical desktop user, NTD = non-technical desktop user, NTL = non-technical laptop user). One goal during this field study was to examine the impact of context (page content, device, location) on participants' application of privacy levels. Results concerning page content were described in section 5.2.1. In this section, we examine the impact of location on participants' overall application of the four privacy levels (public, semi-public, private, don't save).

We found in the PG2 field study that the overall application of privacy levels did vary depending on the location, as seen in Table 24. Browsing conducted at home tended to be less often classified as private and more often classified as don't save than browsing conducted away from home. The application of privacy levels also varied between work and school as away locations. Browsing conducted at work tended to be less often classified as public or private and more often classified as semi-public than browsing conducted at school. However, it would be unwise to draw conclusions given the small number of participants and the potential for individual differences with inherent privacy concerns, potential users and viewers in the location, and the social norms of the location. Furthermore, as not all participants performed browsing in all locations, we cannot perform statistical comparisons.

Table 24. Overall application of privacy levels in PG2 study by location of browsing.

	public	semi-public	private	don't save
overall	40.0%	19.6%	25.3%	15.1%
away	36.3%	19.0%	37.1%	7.5%
work	27.3%	31.7%	30.6%	10.5%
school	42.6%	10.2%	41.7%	5.5%
home	42.8%	20.0%	16.2%	20.9%

During the PG2 field study (n=15), five participants did some browsing at work, accounting for 17.9% of all visited pages; four participants did some browsing at school, accounting for 25.8% of all visited pages; and twelve participants did some browsing at home, accounting for 56.3% of all visited pages. There were six participants who had browsing captured in more than one location: three participants were split between home and work and three participants were split between home and school.

If we examine participants that browsed in multiple locations (Table 25), we can see that their individual privacy patterns did change according to location of use. The light gold shading highlights those cells showing browsing away from home that are above 5 percentage points from the home levels; the darker blue shading highlights those below 5 percentage points from home levels. As can be seen, although the privacy applications differ between locations for those users browsing in multiple locations, there is no pattern to how they differ between users. The difference in perception of sensitivity of browsing may be due to varying inherent concerns, differences in browsing activities, or differences in the types of potential users or viewers and the implications of privacy violations in that location.

Table 25. Comparison of home and away browsing for participants with activities in both locations. Light gold shading indicates an increase from home, darker blue shading indicates a decrease.

	Home				Away			
	public	semi-public	private	don't save	public	semi-public	private	don't save
ID					**Away = work**			
TD5	83.4%	9.5%	7.1%	0.0%	100.0%	0.0%	0.0%	0.0%
NTD2	14.6%	10.3%	27.4%	47.8%	38.9%	21.8%	16.8%	22.5%
NTD5	32.0%	6.7%	31.0%	30.3%	25.2%	21.0%	31.2%	22.7%
					Away = school			
NTL3	51.1%	30.1%	8.5%	10.3%	88.5%	6.6%	0.0%	4.9%
TD3	91.9%	0.0%	8.1%	0.0%	70.5%	1.1%	28.3%	0.1%
TD1	49.1%	6.0%	28.6%	16.4%	36.3%	13.1%	26.7%	23.9%

6.4 Impact of Context on Browsing Activity

Location can impact what kind of browsing activities are conducted. One of the IIP survey questions asked participants to approximate how much of their web browsing activity was conducted for personal reasons, educational reasons, or work reasons. While some participants only reported a subset of reasons, others reported all three. We therefore cannot perform statistical analysis on this data.

As can be seen from the box plots in Figure 33 which show the mean values as well as the distribution, the primary location of use appears to be related to the purpose of browsing. While the amount of educational browsing was consistent between home and away (mean 27.0% home, 26.9% away), the amount of personal browsing increased when the majority of browsing was conducted at home (51.1%) rather than away from home (34.6%). Correspondingly, the amount of work browsing decreased when the majority of browsing was conducted at home (21.5% home, 38.5% away). This difference in the amount of personal browsing depending on the primary location of use makes sense given that people may self-regulate their browsing activities in the workplace as a mechanism to neutralize surveillance by others [101].

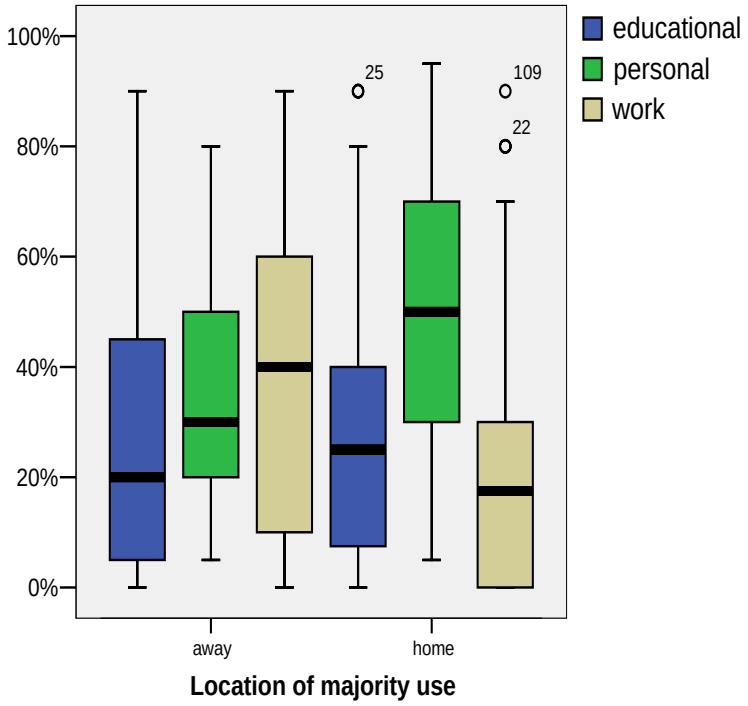


Figure 33. Purpose of browsing, by location of majority use.

It is interesting to note that there is a great deal of variability in the levels of browsing. For example, some participants conducting the majority of their browsing from home have a primarily work-related reason so are probably people that work from home. Some participants that conduct the majority of their browsing away from home have a high percentage of personal browsing, perhaps because they have limited access at home. Clearly not all browsing conducted at work is work related, nor is all browsing conducted at home, personal.

What devices are available for use may also impact where participants performed their personal browsing. Overall, 65.9% of IIP Survey participants answered questions about laptop use. Most of these participants also reported using other computers at times (87.9%). We asked laptop users to report which devices they would use to conduct web browsing of a personal nature. Most reported they would use their laptop computer (86.3%) for browsing of a personal nature, although home desktop computers and away desktop computers were also used for personal browsing by many participants (55.0% and 33.8% respectively).

6.4.1 General Browsing Activities

We now examine how the general browsing activities IIP survey participants reported conducting (as first examined in section 4.5.1) vary according to the browsing environment. Participants did vary their reported activities depending on the location of their computer (regardless of type of computer). Most participants that use their web browsers for a given activity will do that activity while at home: 73.8% for technical support forums, 91.3-98.5% for the remainder of the activities including email, entertainment, banking, personal improvement forums, shopping, medical information, games, and erotica. However, only technical support forums and email are accessed similarly at home and away. The remaining activities, which are more personal in nature, are much less likely to occur when participants are away from home (6.2% for erotica, 40.9-55.2% for remainder).

We were interested to determine if participants partitioned their web browsing activities according to location or conducted the browsing in both locations. In order to reflect participants' choice rather than circumstance, we omitted participants that only indicated browsing activity for one of the locations. While activities such as email, technical support forums, and entertainment browsing often occurred in both locations, the more

personal the type of activity, the more likely the activity was conducted only at home (Figure 34). With the exception of technical support forums, few users only conducted browsing activities while away from home. Interestingly, all participants who indicated viewing erotica away from home were laptop users. It is apparent that traces of prior browsing activities of differing sensitivities may be generated in each location of use; this may increase uncertainty about which traces have been saved. In particular, this may impact laptop users who do the majority of their browsing on a single device that moves between settings, as well as users who consolidate their Histories and Favorites online for use in multiple locations.

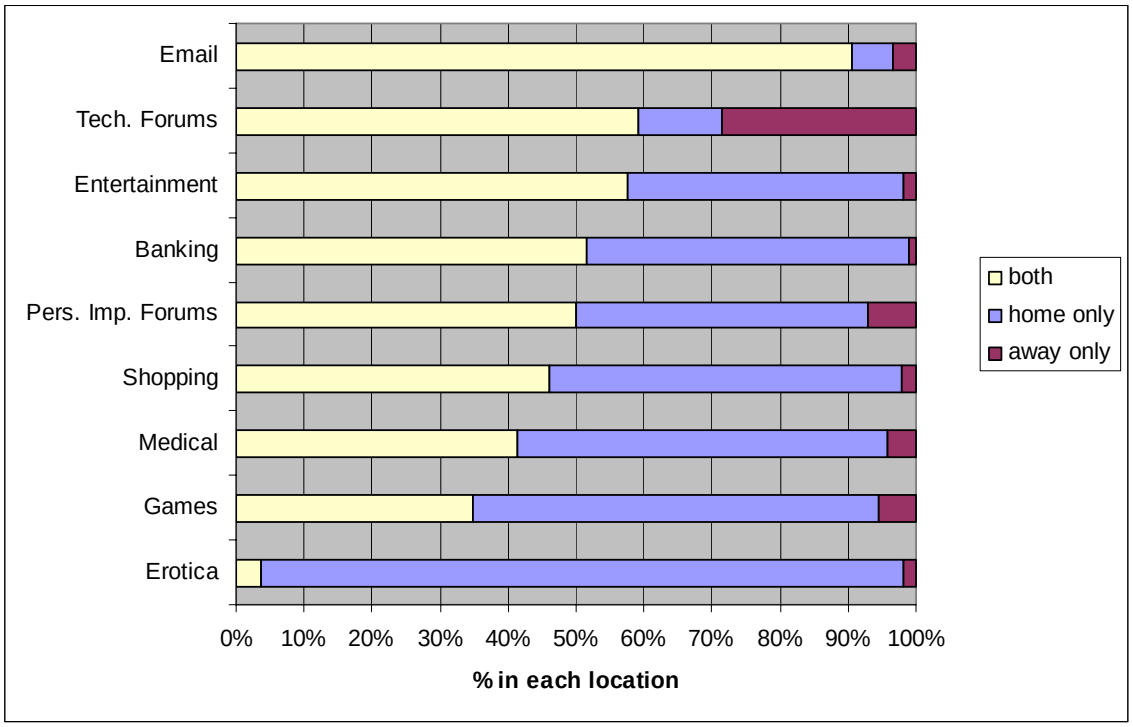


Figure 34. The proportion of participants reporting each activity, who conduct the activity both at home and away, only at home, or only away from home.

We examined the average privacy comfort levels assigned by participants for the scenario that had them reflect on their *usual browsing activities* using a two-way ANOVA with primary location of use and majority computer of use as the independent variables. A main effect was found for the primary location of use (home, away from home). Participants who performed the majority of their browsing at home gave lower average privacy comfort level ratings for the usual browsing scenario than those performing the majority of their browsing away from home ($F(1, 148) = 7.45$, $p=0.007$). The difference may be due to the wider range

of personal activities that participants stated they engage in on their home computers. The main effect for the type of computer used ($F(2, 148) = 0.85$, N.S.) and the interaction effect ($F(2, 148) = 1.76$, N.S.) did not reach statistical significance. The diversity of computers in regular use (Table 22) may have impacted our ability to detect significant differences in privacy comfort level by computer types.

6.4.2 Impact of Context on Content Categories of Visited Pages

In the last section, we reported on the general types of browsing activity that occur as reported by participants in the IIP survey. Participants' reported different types of activities between locations, with more personal or non-work related browsing often occurring only at home. We now present results from the PG2 field study examining how the logged browsing activity varied between locations and devices. Caution must be taken when generalizing the results as not all participants browsed in all locations and we have a limited sample size.

We examined 17 frequently visited categories of page visits across all participants. These categories include the 16 most frequently visited categories (i.e., more than 400 page visits total) and the pornography category (285 page visits). Figure 35 shows the percentage of visits within each of the categories at each location. We included pornography in this

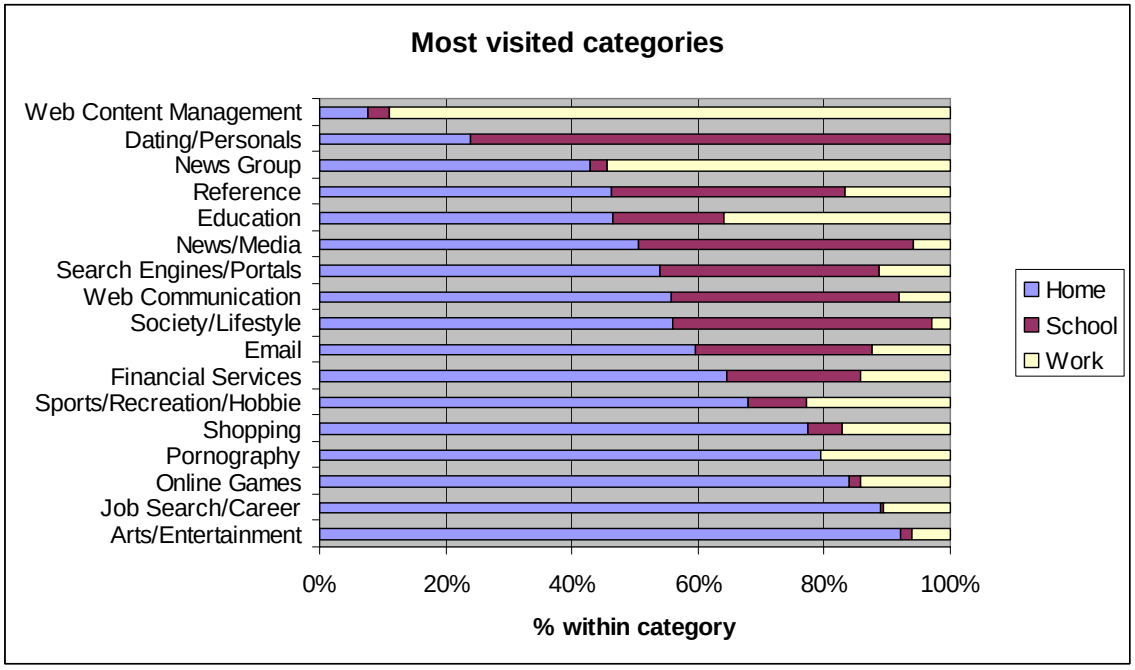


Figure 35. Most visited categories of web pages during the PG2 field study, by location of browsing.

investigation as IIP survey participants most frequently partitioned their viewing of erotica between locations and we wanted to be able to confirm whether this occurred with the PG2 field study participants as well. As can be seen, some categories occur much more frequently in one location than another. Interestingly, the relatively high percentage of dating/personals at school can be attributed to a single participant who did not have access to a home computer.

We examined the page visits of the four participants with more than 5% of their browsing in a second location to get a clearer picture of how browsing activities changed between locations for those participants. The following participants were included in this within participant analysis: NTD2 (86.5% home, 13.5% work), NTD5 (39.7% home, 60.3% work), TD3 (43.3% home, 56.7% school), and TD1 (23.8% home, 76.2% school). For each participant, we included those categories accounting for at least 1.0% of browsing within one of the locations. Figure 36 shows the division of browsing activities for participant NTD2. This participant exhibited a high amount of partitioning with 9/10 of their most frequent categories having more than 75% of page visits in a single location (home or work). Participant NTD5 had 9/15 of their most frequent categories similarly partitioned between home and work, TD3 had 9/16 categories partitioned between home and school, and TD1 had 7/12 categories partitioned between home and school.

We had intended to examine the actual browsing data from the PG2 field study to determine if the types of activities did indeed change depending on the type of computer. Although the five laptop users in the study used their laptops in differing locations, they had

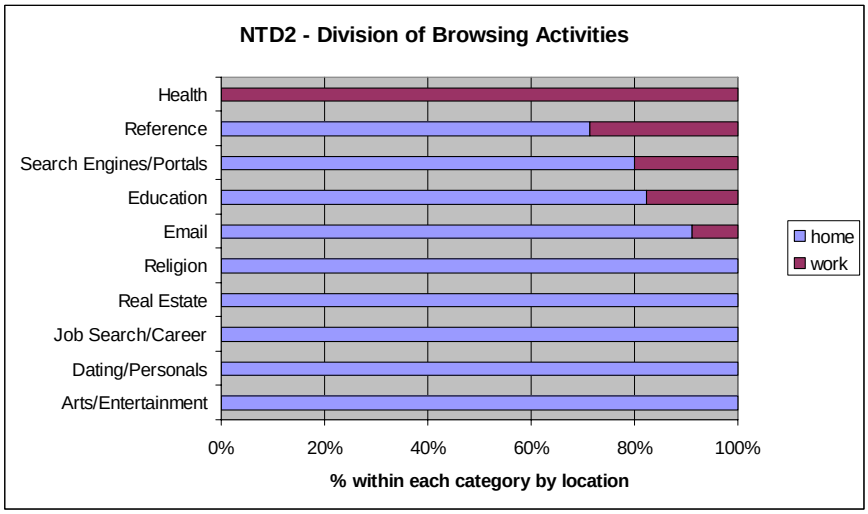


Figure 36. Division of browsing activities between home and work for participant NTD2.

limited internet access when away from home, so we are unable to contrast their usage with that of the desktop users to gain additional insight on the effect of device context on perceived content sensitivity across browsing locations.

6.4.3 Summary

Browsing activities varied according to the usage contexts. General browsing activities and purposes of browsing were found to vary according to device and location. The categories of visited web pages in the PG2 study also differed between locations. Overall, this can be seen in the different percentages of page visits within a location (across categories) and well as different percentages of page visits within a category (across locations). The four participants with more than 5% of page visits occurring in a second location also exhibited some degree of partitioning of activities between locations

6.5 Impact of Context on Browser Settings

As we have just presented in section 6.4, browsing activities appear to change depending on the location and the device in use. However, whether or not an activity results in a stored trace depends on the browser convenience feature settings. Browser settings (e.g., saving 0 days history) can reduce which browsing activities are recorded and may appear as visible content.

One problem for users who attempt to manage incidental information privacy in their web browser is that it is not always clear what traces will be revealed. With multiple devices, there may be increased uncertainty, particularly for those users that don't partition their browsing activities between locations and devices. As we next present, many participants indicated that they used their web browser convenience features differently for each computer. Participants tended to be less likely to use the convenience features on their desktop PCs away from home than on their home desktops or laptop computers. This lack of standardized settings across computers could add to the uncertainty about what will be revealed for each computer. Before tools can be developed to help users maintain incidental information privacy in their web browsers, we require an understanding of how users currently manage the tradeoff between convenience and privacy during collaboration. We next present an analysis of the impact of context on current privacy management practices using data obtained from a common set of questions administered during the IIP survey and

the PG1 and PG2 field studies. As one participant took part in both field studies, his responses from the first field study were omitted resulting in a total of 189 participants for this analysis.

This portion of the questionnaire began with the statement: "Web browsers offer various convenience features such as Favorites/Bookmarks, History, and Auto-completion to allow for easier web browsing; but these features may also display material that can be inappropriate. Please think about how you handle the tradeoff between convenience and privacy." For each of the questions (see Table 26), participants were asked to respond for each of their computers in regular use (home PCs, work/school PCs, laptops). Following each question, participants were given space to describe "How would you like to be able to manage this feature."

Table 26. Questions investigating web browser convenience features use and their possible answer choices.

Question 1:	***Question 2***:	***Question 3***:
The Favorites/Bookmarks feature allows you save the title and web address of web pages that you would like to re-visit. How do you use this feature? (check one).	The History feature allows you to keep a record of URL's visited. How is this feature set? (check one)	The Auto Complete feature stores previous entries and lists possible matches from entries you've typed before. How do you currently have this feature set? (check all that apply).
Answer choices:	***Answer choices:***	***Answer choices:***
1) Use it to save web addresses with default/ accurate names; 2) Use it to save web addresses, but rename some to conceal the identity; 3) Don't use.	*1) Unsure; 2) Default setting; 3) Set for 0 days history to be stored; 4) Set for some number of days history to be stored. (Specify number of days (if known)).*	*1) Unsure; 2) Default setting; 3) Use for web addresses; 4) Use for forms; 5) Use for user names and passwords on forms; 6) Don't use.*

For each convenience feature studied, we present how participants reported using the feature to manage their privacy and how they indicated they would like to enhance the feature. Comments about potential enhancements describe both how participants would like to manage their privacy within the feature and how they would like to change the feature's functionality as a revisitation tool. As privacy is a secondary consideration to the primary purpose of these tools for revisitation, it is important that we ground our examination of privacy within the desired usage of these convenience features. Numbers in parentheses represent the number of participants responding.

We rely on descriptive statistics to report this information. Not all participants reported browsing in all locations and on all devices, so within participant statistics would omit most participants from comparisons. Furthermore, most participants had more than

one browsing context, so between participants comparisons for each context of use are inappropriate due to the lack of independence. For each feature we give descriptive statistics for overall patterns of use between contexts and provide within participant values for those participants with browsing in more than one location.

6.5.1 Favorites

For Favorites use, there were 141 responses for home PCs, 136 for work/school PCs, and 126 for laptops. Most participants reported that they use default or accurate names when saving web pages in their Favorites (Figure 37). Responses indicate they were less likely to use Favorites when working on a PC at work/school than on either a PC at home or a laptop and were also less likely to change some names to conceal the identity of pages.

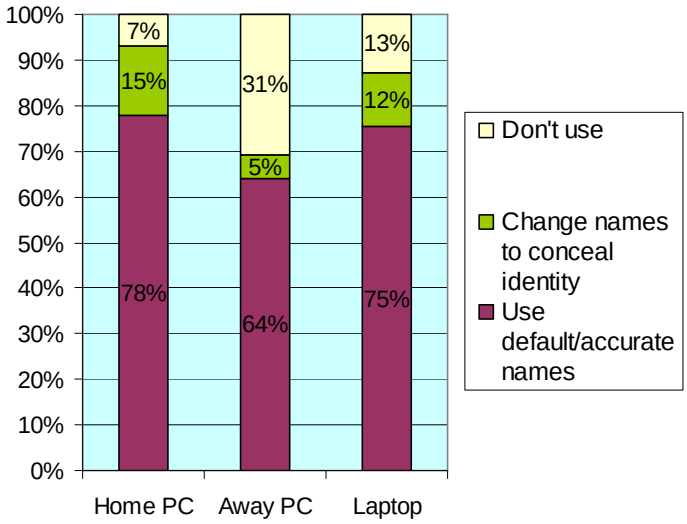


Figure 37. Participants' reported usage of Favorites.

A total of 138 participants gave responses for more than one computer, and of those 42 (30%) reported that they used Favorites differently depending on the computer. The main differences reported were between PCs used at home and work/school (n=38). Thirty participants reported not using Favorites at work/school and either using default/accurate names (23) or concealing the identity of some sites at home (7). Others (8) used only default or accurate names at work/school, but either concealed the identity of some sites (7) or didn't use Favorites (1) at home.

Many participants (38) made suggestions to enhance Favorites. Interestingly, many of these were not directly related to privacy. This may be because Favorites already give users

flexibility in what is recorded. Users can opt to rename entries and must explicitly take the step of saving a web page in Favorites so there is awareness of what may appear later. Indeed, three participants stated they only saved entries they would not mind others seeing later. Privacy related suggestions included password protection of Favorites (4), the ability to have profiles within the browser (3), and a desire for more control (2).

Many of the remaining suggestions concerned improvements for managing Favorites. Improvements at the time of recording included more meaningful default names (8), methods of classifying the saved entries (4), blocking auto-entries from Java script (1), auto-adding sites that are frequently visited (1) and allowing multiple entries under different classifications (1). Improvements related to re-finding entries included having them available on-line so that they could be used from multiple computers (2) and making Favorites searchable (1). In terms of keeping Favorites uncluttered, one participant suggested having auto-deletion of entries after a certain length of time and another suggested that the browser periodically check whether the entries were still valid. Four participants said they would not want to change the functionality, one said that they used their Favorites when monitoring frequently visited sites and one reported not using Favorites because they relied on Google to re-find visited sites.

6.5.2 History

For History use, there were 146 responses for home PCs, 142 for work/school PCs, and 128 for laptops. As seen in Figure 38, most participants reported that they use the default setting for the number of days to store their browsing History (20 days in IE). Overall, participants' responses indicate they were more likely to set their History to store 0 days when working on a PC at work/school than on either a PC at home or a laptop computer; they were also less likely to be sure of what the exact setting was or to have set their own number of days (other than 0). The median number of days set was 7 for PCs (ranging 1 to 9999 for home and 1 to 30 for away) and 9 for laptops (2 to 9999).

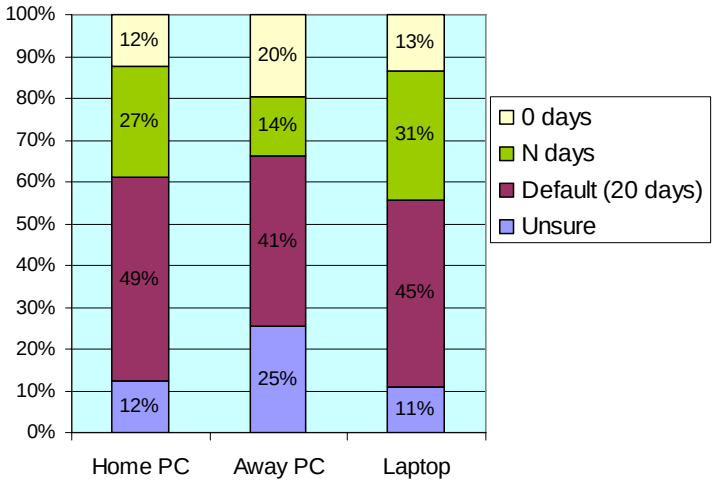


Figure 38. Participants' reported History settings.

Participants that specified their own number of days (0 or N) overwhelmingly chose to save fewer than 20 days History (as shown in Figure 39). Responses given for number of days saved in History for work/school PCs indicated a more conservative value than for home PCs or laptops. This desire for fewer days saved suggests that the default of 20 days may be less appropriate.

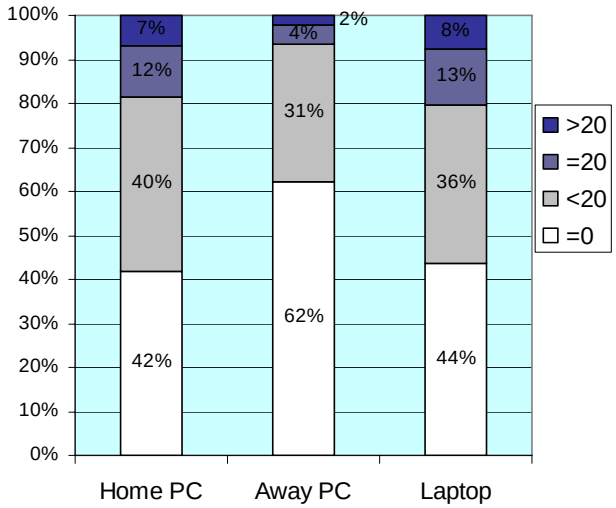


Figure 39. Number of days History saved for participants who specified a non-default value.

A total of 142 participants gave responses for more than one computer, and of those 78 (55%) reported that they had different History settings depending on which computer was in use. As in the overall findings, participants were more likely to have a lower number of days saved on a work/school PC than at home or on their laptop.

Seventy-two participants made comments or suggestions about the History feature; many were directly related to privacy. A common concern was the inability to selectively save or delete History entries. Participants stated they wanted more fine-grained control at the item level (16), to toggle the recording of pages on and off within sessions (6), to enable/disable History at the session level (8), and to set History to automatically clear when the computer shuts down (2). A further six indicated that they manually delete history at the end of a session and three others expressed a desire for an easier method of clearing History. Participants also expressed a desire for password protection (5) and profiles (2).

In more general comments, several (11) participants thought that the History worked well as is and one expressed a desire for greater awareness of what pages would be stored. There was also a lack of awareness about how this feature could be configured: seven expressed a desire to set the number of days, apparently unaware that this ability existed.

6.5.3 Auto Complete

For Auto Complete settings, there were 143 responses for home PCs, 140 for work/school PCs, and 126 for laptops. The question asked participants to report on their settings for this feature. The option "don't use" was intended to mean that the feature was disabled (i.e. don't use to store data) rather than that participants chose not to make use of the presented Auto Complete text. While we can not be sure if there is ambiguity in how participants interpreted "don't use", we did have 5 participants who mentioned in the general comments for this feature that they don't make use of Auto Completes although they had not disabled them via the settings. For these participants, traces of prior activities will still be visible as text in their Auto Completes; they just choose not to select an entry from those offered in the selection box. For the sake of this analysis, we will consider those participants who selected "don't use" to have indicated that they have disabled the feature.

As seen in Figure 40 most participants reported using this feature either with the default settings (storing and displaying text for web addresses, user names and passwords) or with participant-specified settings (some combination of web addresses, forms, and user names and passwords being saved). Laptop users were more likely to specify their own Auto Complete settings and were most likely to use this feature; those that use PCs at work/school were least likely to specify their own settings or use this feature.

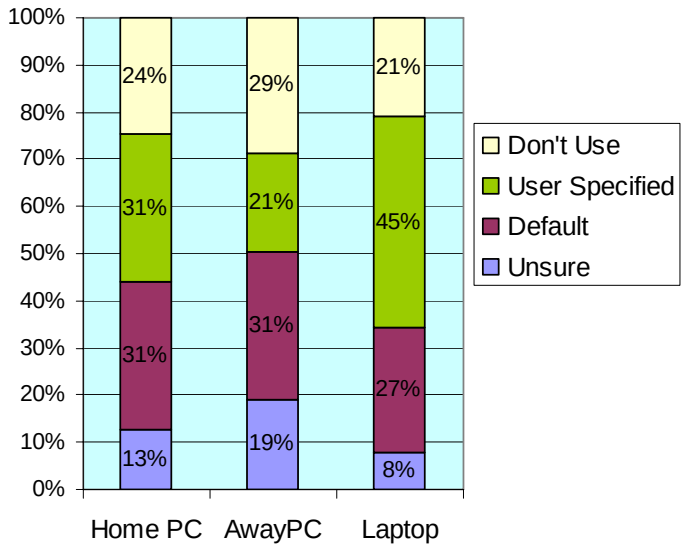


Figure 40. Participants' reported Auto Complete settings.

In over half the instances, participants reported that they opted to change the default settings (i.e. not default or unsure). The most common setting reported (see Figure 41) was to opt to not save any of the text, thus disabling the feature. Interestingly, the default setting (storing URLS and passwords) was rarely given by those participants indicating precise settings. It is unclear whether this is because they were aware of the default settings and approved of them (i.e. no need to change) or because the default setting is not optimal. One participant did suggest that the default setting should be to have the feature disabled.

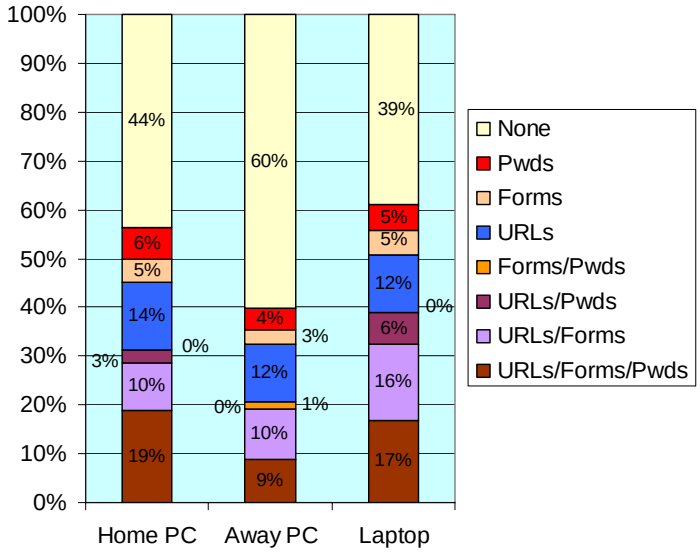


Figure 41. Types of data saved in Auto Complete for participants reporting a setting other than unknown or default.

Forty-seven participants made comments or suggestions about the Auto Complete feature. Many of these were directly related to privacy. As with the History feature, one of the key privacy problems identified was the inability to selectively save or delete entries. Participants stated they wanted more fine-grained control at the item level (4), the ability to toggle the recording of pages on and off within sessions (6), and the ability to enable/disable additions to the Auto Complete entries at the session level (2). This similarity to the desired functionality for the History feature is expected due to the dual purpose of the entries in the History files (both for History functionality and Auto Complete). Participants also suggested password protection (4) and the use of profiles (3). Again, participants (9) showed a lack of awareness of functionality, expressing a desire for an existing ability. A few participants suggested some automated functionality should be provided. Examples of such functionality included not recording credit card information or form data from secure pages (2) and specifying keywords that would be used to filter what information gets stored (1).

In more general comments, some participants thought that the Auto Complete function worked well as is (7). Again, there was a lack of awareness about how this feature could be configured with nine participants expressing a desire to disable some of the functionality (e.g., "to be able to disable that for my user names and passwords"), apparently unaware that this ability already existed.

6.5.4 Limitations

One of the limitations of questionnaire data is that it relies on participants' ability to accurately report their data. A benefit to the mixed methodology approach we took in this research is that we were able to triangulate the data and compare results between methods. In the second field study, participants completed the questionnaires on paper while the researcher installed the logging software on their computer. We installed the software on 20 computers (7 home PCs, 8 school/work PCs, 5 laptops).We took this opportunity to check participants' actual settings for History and Auto Complete. When we compared the settings we recorded with participants' self-reported data, we found inaccuracies between what the participants' believed the settings were and what they actually were in 22.5% (9/40) of the recorded settings.

Only 5% (1/20) of the instances of History settings were reported incorrectly. The sole participant was quite confused about their History settings: they thought they were using the default setting, actually had their computer set to record 0 days, and had commented that they wanted the maximum setting. Participants were much less accurate when reporting their Auto Complete settings; 40% (8/20) instances were reported incorrectly. Errors included two participants thinking that they were storing nothing, but using the default settings (1) or saving all three types of data (1); three thinking they were using the default settings, but also having forms set (3); and three thinking they were storing just URLs, but also saving passwords (2) or saving all three types (1). There were also 2 instances of "unsure" for History and 3 instances of "unsure" for Auto Complete; in all cases, these participants were using default settings. The Auto Complete error rates were consistent across computer types/locations. The high error rates for Auto Complete are an indication of the complexity of the feature's settings.

It should be noted that as the field study participants did not have access to their computer while completing the questionnaires, this error rate is likely higher than for the participants who completed the on-line survey who may have checked their actual settings while responding, at least for the device in use while they completed the survey. Due to the potential for high error rates, particularly for Auto Complete, it may be best to view those usage results as an indication of participants' preferences for settings rather than their actual settings.

6.5.5 Design Implications for Enhanced Browser Convenience Features

There are several design implications that arise from our analysis of participants' convenience feature settings that should be considered by those developing enhanced browser convenience features. These are summarized in Table 27.

Table 27. Summary of convenience feature settings results and their implications for general design of enhanced browser convenience features.

Concept	Section	Study	Findings	Design Implications
Awareness of settings	6.5.1, 6.5.2, 6.5.3, 6.5.4	IIP Survey, PG1, PG2	Many participants were not aware of their current browser settings.	Increased visualization of settings needed.
Understanding of settings	6.5.1, 6.5.2, 6.5.3, 6.5.4		Many participants were incorrect in their understanding of settings, particularly for Auto Complete.	Increased visualization of the impact of the setting on browser actions.
				Clearer explanations of feature functionality that are readily accessible to the user.
Default values for settings	6.5.1, 6.5.2, 6.5.3		Analysis revealed participants' reported changed in their settings from the default provided.	There is an opportunity to provide more intelligent defaults based on user preference.
			We found differences in the settings between locations of use.	Different default profiles could be developed for typical situations of use (e.g., web browser at work).

6.6 Impact of Context on Post Browsing Actions

Preventative actions such as erasing all traces in the History feature can also limit what traces of browsing activity are potentially visible. As initially discussed in section 5.1.2, during both of the field studies as well as the survey, participants were asked to reflect on what actions they might take to conceal potentially sensitive information if given advanced warning that somebody else would be working closely with them. Further examination of the results revealed differences in how some of the features tend to be used depending on the location/device setting (home PC, away PC, laptop).

Laptop users in the IIP survey were most likely to report they would take some action if given advanced warning that someone could view their display (66.4% of home PC users, 60.2% of away PC users, and 76.6% of laptop users). Furthermore, in the PG1 field study (consisting of 20 laptop users), 19/20 participants indicated that they would take some actions. In the PG2 field study, 5/5 non-technical laptop users, 7/8 desktop PC users at work, and 7/8 desktop PC users at home indicated they would take some actions. As stated in section 5.1.2, the higher rate for actions found in the two field studies were likely due

heightened privacy awareness. The field study participants were required to have occasions where their browser window could be viewed by others so that they would be able to relate to this type of privacy concern. Across all three studies, 67.6% of home PC users, 60.6% of away PC users, and 72.6% of laptop users reported taking some actions.

We next discuss the differences in the actions that participants reported that they take to preserve their privacy when given advance notification that somebody will working closely with them and will be able to look closely at their display.

Figure 42 shows the actions that participants indicated they would take across the three studies.

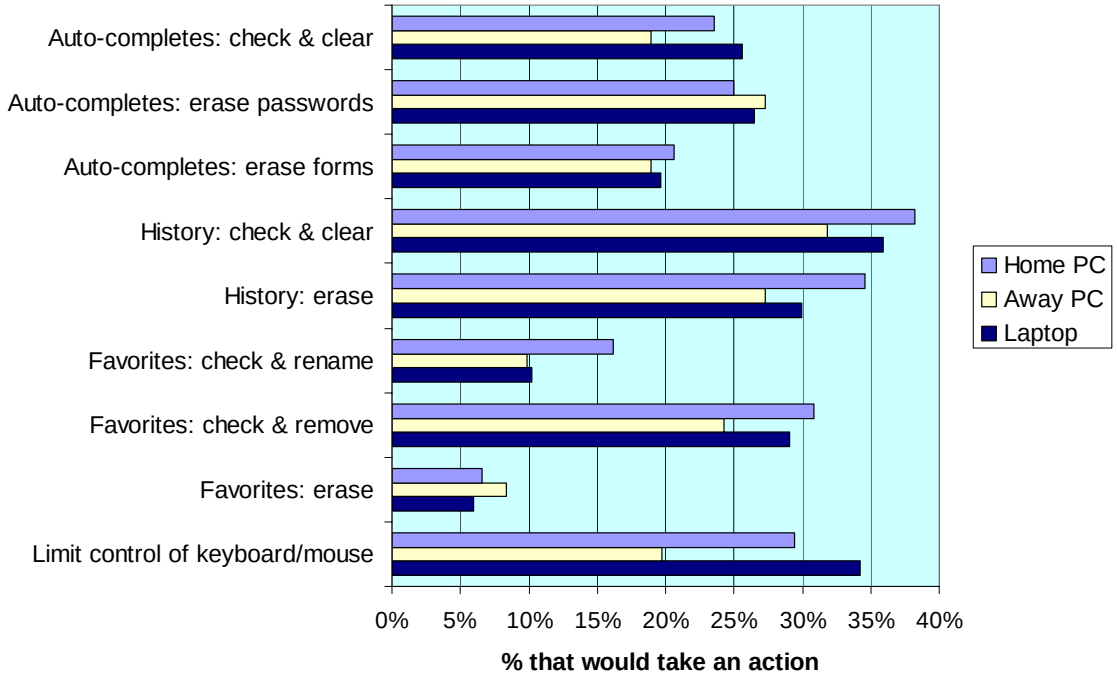


Figure 42. Percentage of participants that would take each action on their Home PC, their Away PC, and their laptop computer across all three studies.

One of the options participants reported taking was limiting their collaborator's control of the keyboard or mouse during the collaboration (29.4% home PC users, 19.7% away PC users, and 34.2% laptop users). These differences could be due to many things. Social norms may play a role; there may be more social acceptance of laptop users limiting control of what is likely to be perceived a personal device, whereas work or school PCs may be more likely perceived as non-personal. Alternatively, more personal activities may be

conducted on laptop computers which may contribute to a desire to stay in greater control over input devices.

Many participants would take some action involving their History (52.2% of home PC users, 43.9% of away PC users, and 51.3% of laptop users). As can be seen in Figure 42, slightly more people would check their history and then clear it rather than just erase it; some participants reported they might do both. In keeping with the overall trend, fewer PC users away from home report taking actions involving their History. However, actions involving the AutoComplete feature were balanced overall (39.7% of home PC users, 40.2% of away PC users, and 42.7% of laptop users), although the components varied with fewer away PC users checking and clearing the AutoCompletes. Actions involving Favorites also were slightly less likely to occur on an away PC than in other location/device combinations (home: 38.2%, away: 31.1%, and laptop: 36.8%).

Laptop users' higher likelihood of taking actions to protect their privacy may be due to participants using their laptop computers for browsing of personal nature and moving between multiple locations. Interestingly, the biggest difference seems to be with laptop users retaining control of their keyboard and mouse as a method of preserving privacy. The relatively low level of actions taken by away PC users may be due to their reduced convenience feature usage on those computers and their reduced likelihood of engaging in personal activities when away from home.

6.7 Impact of Context on Perceived Sensitivity of Traces

Sections 6.4, 6.5, and 6.6 examined what traces may be generated as a result of browsing activities, stored as a result of convenience feature settings, and allowed to remain during instances of collaboration around a display. We were also interested in how privacy concerns for the resulting potentially visible content varied between locations. We examined PG2 field study data to determine if there were differences in how the privacy levels were applied to the most frequent categories between home, school, and work. As can be seen in Figure 43, there is some variation in privacy sensitivity between locations. The graph shows the patterns of privacy application for each category with the locations stacked: home is the bottom row, school is the middle row, and work is the top row. If the patterns of application were the same between locations, we would expect to see the same general pattern repeating

in each row. As can be seen, however, many categories are quite different. For example, job search is mostly semi-public at home, all semi-public at school and mostly private at work. However, some categories are basically the same; for example News/Media was classified as mostly public in all three locations. As before, this overall picture does not account for individual differences between the users with browsing in each location.

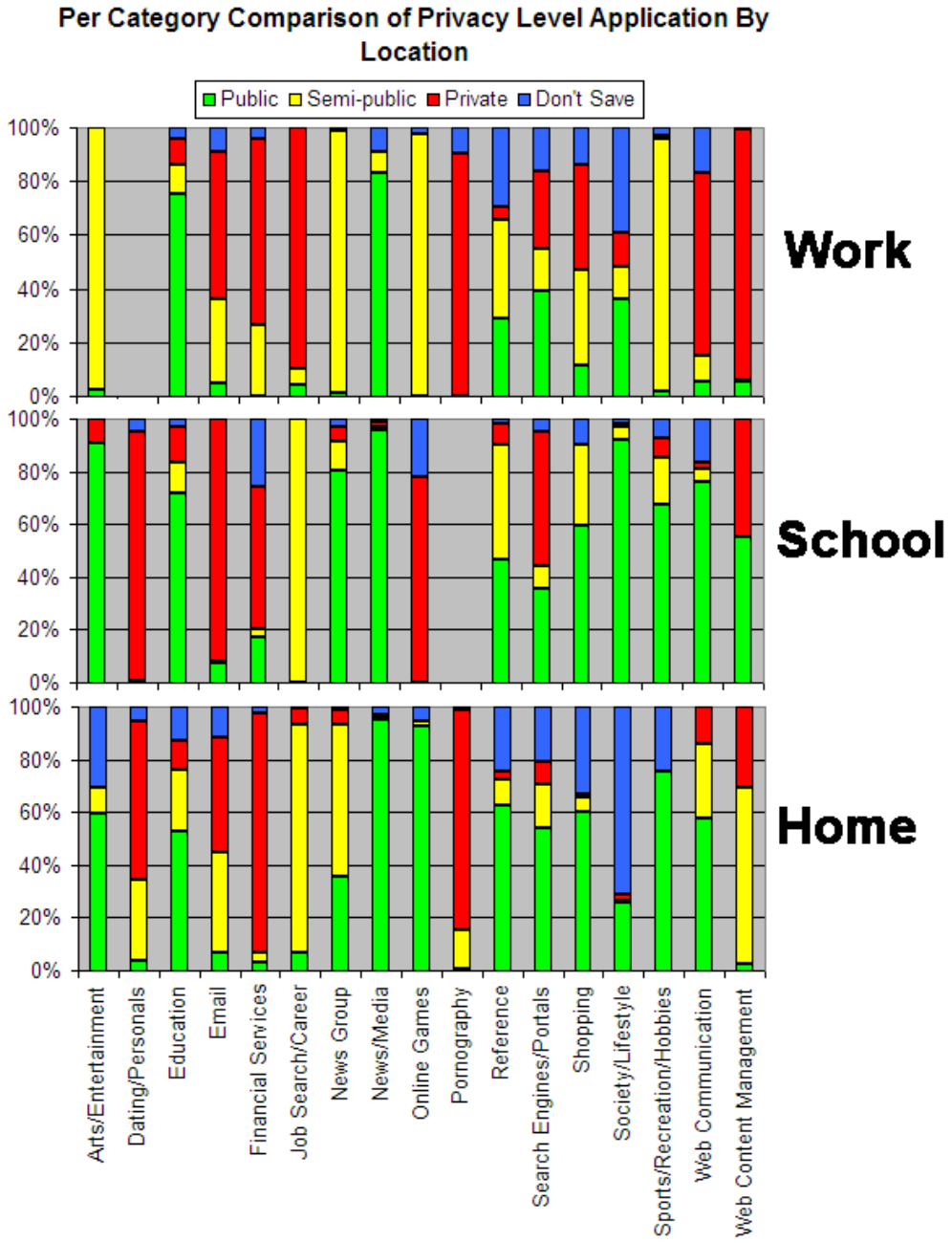


Figure 43. A per category comparison of privacy level application by location. Top layer: at work, Middle layer: at school, Bottom layer: at home.

We examined our four participants who reported browsing in multiple locations (those with more than 5% of browsing at a second location), to get some insight into how patterns change within users. For example, although a few of NTD5's most frequent categories had similar privacy levels applied between home and work (e.g. financial services, media/mp3), most were quite different (Figure 44). Most categories appeared to be considered more private at work (e.g., computers/internet, reference, search engines, shopping,, and web communication).

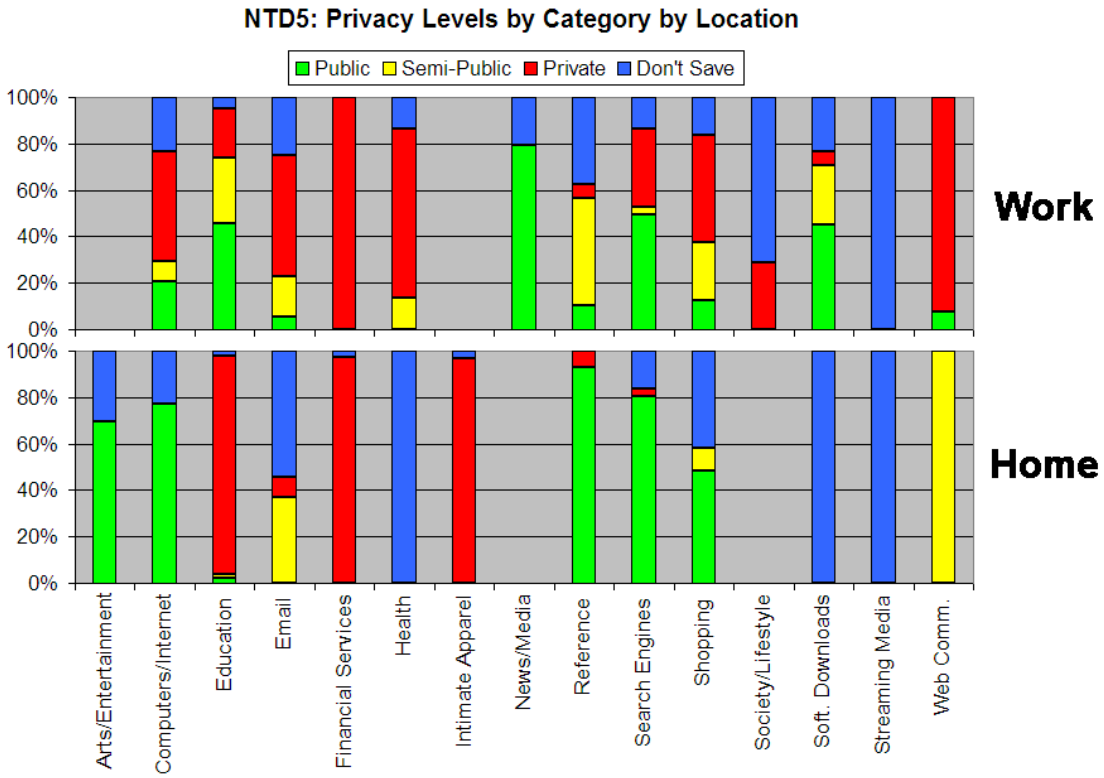


Figure 44. A per category comparison of privacy level application by location for participant NTD5.

The use of don't save, which could either mean that a page visit was irrelevant or extremely private, makes it difficult to determine the relative sensitivity of some categories. For example, the health category was 100% don't save at home and considered to be mostly private at work. One participant (TD3) was quite consistent between school and home. TD3 was not as nuanced as some other participants. This participant considered most of their browsing to be public in nature, some of it private, and little in between. The other two

participants had patterns similar to those shown for NTD5 (Figure 44) with differences in privacy level application noted for a subset of the categories.

In summary, the overall pattern of participants' applied privacy levels (perceived sensitivity of browsing) differed between locations, both when looked at across all participants and within participants. Furthermore, there appears to be some differences in the application of privacy levels within a category between locations. Three of the four participants with browsing in two locations did exhibit changes in the application of privacy levels for some categories between locations. Those with browsing split between home and work appear to have more differences than those with browsing split between home and school, but that may be due more to individual differences than location. We have noted a similar pattern during a recent evaluation of a privacy enhancing web browser (as will be discussed in Chapter 8). Staff participants considered more browsing to be sensitive than did student participants. Participants felt that non-work related browsing should not be able to be viewed by their colleagues or employer. Further studies with a greater number of participants will be required to determine if this is indeed the case.

6.8 Summary of Examination of Context

We examined the impact of context on privacy level application (section 6.3), browsing activities (section 6.4), browser settings (section 6.5), and post browsing actions (section 6.6). Table 28 summarizes the results. Implications for design of a visual privacy management system arising from these results will be discussed in Section 7.1.

6.8.1 Limitations

The PG2 field study gives us some information about the browsing activities conducted across locations and the perceived sensitivity of that browsing. However, as PG2 had a small sample size (15) and not all participants engaged in web browsing in each location, we are unable to draw many conclusions. None of the laptop users conducted more than 5% of their browsing in a second location, so we were unable to investigate differences for laptop users that move between locations. We can, however, say that the patterns observed appear to support incorporation of situational variables into the model of visual privacy concerns during web browsing.

Table 28. Summary of impact of context (location, device) on participants' application of privacy levels, browsing activities, browser settings, and post browsing actions taken.

Concept	Section	Study	Aspect of Context	Our Findings
Overall Application of Privacy Levels to Web Browsing	6.3	PG2	Location.	Differences in overall application between home and away locations and between work/school.
				Participants with browsing in multiple locations differed in patterns of application
Browsing Activity	6.4	IIP Survey	Location	For majority location = home, more browsing for personal reasons, less browsing for work reasons. Highly individual, some participants counter the trend.
			Device	Laptop users most likely to conduct personal browsing on laptop rather than a home or away desktop computer.
	6.4.1	IIP Survey	Location	More personal activities (e.g. erotica, personal improvement forums, shopping, banking, etc.) are less likely to occur at both home and away locations, more likely to be at home only.
	6.4.1	IIP Survey	Location	For majority location = home, PCL for usual browsing scenario lower than for away
	6.4.2	PG2	Location	Those with browsing in multiple locations exhibited partitioning for more than half content categories of visited pages.
Browser Settings	6.5	IIP Survey, PG1, PG2	Location & Device	Away PC: less Favorites use/modifications, less History storing than for home PC or laptop
				Laptop users more likely to use Auto Complete
				Many participants with 1+ contexts of use, used their convenience features differently: History (55%), Favorites (30%)
Post Browsing Actions	6.6	IIP Survey, PG1, PG2	Location & Device	Laptop users most likely to take some action if given warning. Home PC users least likely to take actions with History Limiting control more likely for laptop users, less likely for work PC
Perceived Sensitivity of Traces	6.7	PG2	Location	Overall and per-participant differences in per-category application of privacy levels between locations

6.9 Summary

Our mixed methodology approach allowed us to examine the privacy of incidental information both in terms of general attitudes and also based on actual behaviours. We have

integrated results from the three studies to build an initial model for user's web browsing behaviours as a result of privacy concerns in this domain. Figure 45 shows our understanding of the factors that impact this behaviour.

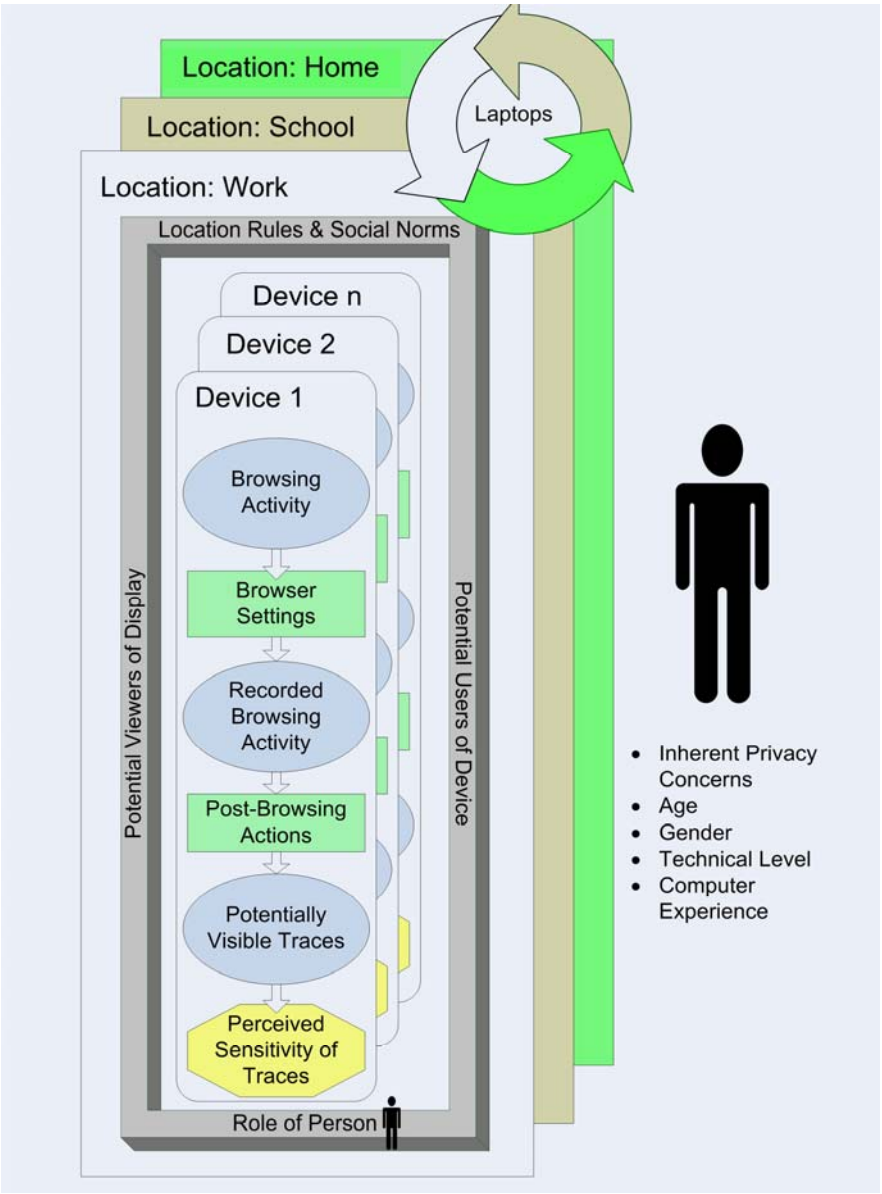


Figure 45. Model of the contextual factors that impact web browsing behaviours.

As we presented in this chapter, dispositional variables such as age, gender, technical level, and computer experience are related to a person's inherent privacy concerns. These inherent privacy concerns can be considered a person's overall privacy preferences. However, the dispositional variables combine with situational variables such as the device in use and the location. Laptop users may move between multiple locations. Within each

location, the social norms and web usage policies, role of the person, and potential viewers of the display and users of the device impact web browsing behaviours and privacy comfort in a given situation. Browsing behaviours that are impacted may include web browsing activities and browser convenience feature settings. The browser settings may reduce which traces of activity are actually recorded. Furthermore, actions taken when given advanced warning of collaboration may further limit which traces are potentially visible. Finally, the perceived sensitivity of these traces may also change depending on the situation.

Without conducting this research, we would not have been able to build such a richly operationalized model of visual privacy concerns within the context of web browsing behaviours. Our model demonstrates that not only are privacy concerns impacted by several situational factors, but also that the information generated is impacted by these factors. While prior research identified subgroups of the situational factors as impacting privacy concerns for other privacy domains, our findings are novel in that we have also demonstrated how this rich set of situational factors are impacted by dispositional factors including inherent privacy concerns. These findings may be important for other privacy domains, particularly those characterized by mobile users, changing physical contexts of use, or changing roles of users.

In Chapter 7, we will present design implications for privacy management systems based on our exploratory findings and will examine the feasibility of two automated privacy management approaches. Then in Chapter 8, we present the design, implementation, and evaluation of PrivateBits, a proof of concept browser developed to validate our design requirements.

Chapter 7
Privacy Management Approaches

Our exploratory research, as discussed in Chapters 4, 5, and 6, gave us a great deal of insight into the privacy of incidental information within web browsers. This chapter examines the feasibility of various privacy management approaches. We begin by presenting the design requirements we developed for a visual privacy management system as a result of our exploratory analysis. Then, in light of those requirements, we discuss three components of a visual privacy management system for web browsers: classification of traces of web browsing activity, filtering of that information appropriately during viewing situations, and maintenance. We then present a feasibility analysis of two automated approaches: one for classifying traces with a privacy level and one for filtering content appropriately.

7.1 Design Requirements for Visual Web Browser Privacy

We now present design requirements for managing visual privacy within a web browser. We begin with guidelines for general privacy management systems as have been identified by previous work. We then present further guidelines specific to visual privacy management within web browsers based on the results from our exploratory studies.

7.1.1 General Guidelines for Privacy Management Systems

We have identified several common themes from related work regarding designing privacy management systems (as initially discussed in section 2.4.1). Most of these are not necessarily privacy-specific, but are grounded in general HCI guidelines. These themes include increasing visualization of privacy settings and actions, working within existing behaviours, and providing opportunities for varying levels of control.

7.1.1.1 Increase Visualization of Settings and Actions

Increased visualization is a commonly proposed guideline for usable privacy and security systems. Lau et al. [90] state that privacy interfaces should make it easy to create, inspect, modify, and monitor privacy policies. Lederer et al. [91] also discuss a lack of visualization in their five pitfalls for designers of systems with personal privacy implications.

One of the pitfalls discussed is obscuring potential information flow. De Paula et al. [39] include visualization mechanisms as one of their three design principles for enhancing the usability of systems with a security and privacy component.

7.1.1.2 Configuration within the Context of Action

Beyond making it easy to visualize settings and actions, configuration of privacy preferences should be done within the context of the resulting actions so that users can more readily see the impact of changes. De Paula et al. [39] propose that configuration and action are integrated; Lederer et al. [91] propose that action should be emphasized instead of configuration.

De Paula et al. [39] also propose that an event-based architecture will help users recognize security and privacy issues as they arise. Similarly, Lau et al. [90] state that privacy policies should be applied proactively to objects as they are encountered. Some visualization of the privacy classification and the rule that triggered it may help users understand how the rules they generate are applied and may also help them notice when rules lack the intended coverage.

7.1.1.3 Provide Opportunities to Vary Granularity of Privacy Control

Lederer et al. [91] also identified a lack of coarse grained control as one of their five pitfalls for designers of systems with personal privacy implications. While fine-grained control is desirable under certain circumstances (e.g., specifying appropriate personas according to explicit situational contexts), at other times, users may best be served by more broad privacy settings. For example, allowing users to easily block the transfer of any information may be useful at times.

7.1.1.4 Work within Existing Behaviours

Another pitfall that Lederer et al. [91] identified is inhibiting existing practices. Privacy is usually a secondary consideration to the task at hand. If privacy preserving actions inhibit a user's normal interactions with the system and complete their desired task, they may be less likely to manage privacy. To be effective, a privacy management system must complement existing practices.

7.1.2 Guidelines for Visual Privacy Management within Web Browsers

We now present a set of requirements specific to the management of visual privacy within web browsers as arose from our analysis of results from the field study.

7.1.2.1 Increased Visualization of Settings

Our exploratory results confirmed the need for increased visualization of convenience settings, both for general usability and for privacy enhanced functionality, as identified in the general guidelines for privacy management systems (7.1.1.1). Developers of enhanced web browser convenience features (e.g. new History mechanisms) should provide increased visualization of the features' settings and the impact of these settings on what information is stored. Participants in our exploratory studies often were not aware of their current settings or what information was being stored that might be subsequently revealed. As was discussed in section 6.5.4, PG2 field study participants often were not only unsure of their actual settings, but incorrect in their understanding of their actual settings, particularly for the Auto Complete feature. Increased visualization of settings may be particularly important for those participants who reported using different settings depending on the location and computer in use. The feedback could indicate the current settings or which information traces are being stored. One alternative may be to provide users with the option to display visual feedback at the time of browsing, so that those users who would find increased feedback helpful could enable the feature.

7.1.2.2 Clearer Explanations of Feature Functionality

It is clear from questionnaire responses from our three exploratory studies (as presented in section 6.5), that several participants lacked awareness of the functionality of the convenience features and their configuration options. These participants expressed a desire for functionality that already existed within their browser features and may therefore not have been as effective as they desired in their convenience feature use. Clearer explanations of feature functionality, including configuration options, should be provided along with methods of easily accessing that information during web browser use.

7.1.2.3 Intelligent Default Settings

There is an opportunity to provide more intelligent defaults for convenience feature settings to fit the primary contexts of use and concerns of users. As discussed in section 6.5,

some of the differences participants' reported for their settings on their desktop computers at work and at school (e.g., increased use of accurate names in Favorites) appear to be due to fewer personal and potentially sensitive activities being conducted in these locations (section 6.4). Other differences (e.g., History saved for fewer days, less convenience feature use) indicate that participants wanted more control over what information may be revealed when they are not at home (section 5.6). For example, a person might want to maintain a more formal persona in the workplace than at home (section 5.5). A more conservative default setting would be appropriate if users indicate that they are in a workplace or another public environment.

7.1.2.4 Reduce Clutter within Convenience Features

Browser convenience features such as Favorites/Bookmarks and History, which are designed to assist with revisitation, are often under utilized [12, 80, 86]. The quantity of traces saved is one barrier to use as it can make recognizing the desired resource difficult. For example, History displays both irrelevant pages and those that are important to the user [12]. While Favorites/Bookmarks contain only those pages that were deemed to be important enough at some point to save explicitly, they may also suffer from clutter and disorganization [12, 18]. A privacy management system should help reduce the clutter by allowing more control over what traces are stored, while not interfering with the revisitation functionality of the features. This can be accomplished by providing a *more usable configuration mechanism*, a *more selective approach to deletion*, and *finer-grained mechanisms for controlling what traces are saved* at the time of browsing.

More usable configuration mechanisms are needed within the web browser to support privacy needs during collaboration. One recent change within IE and other web browsers has been the addition of a one button clear function to remove all web browsing traces from the History and Auto Complete functions, rather than requiring users to navigate to multiple locations within the menu structure. For those users that do not want traces of their activity stored, a desire was expressed to have the clearing occur automatically either at the end of the session or when the computer shut down.

Another method of reducing clutter would be to *provide a more selective approach to deletion.* Users should be able to selectively delete those items that may violate their privacy needs during collaboration. While IE History does allow users to delete specific items, that

functionality is separated from the clear function in the tool bar menu and is only visible upon a right mouse click on an entry within the History panel itself. Auto Complete does not have the functionality to inspect and selectively delete saved form data or user names and passwords.

In addition to providing opportunities for deletion after the fact, a system should also *provide fine-grained control at the time of browsing activity*. This would allow users to discard sensitive web sites immediately rather than trying to remember what sensitive browsing may have occurred in the past. Participants in our exploratory studies expressed a desire for more flexible control, with the ability to selectively toggle between saving and not saving web browsing activity in their various convenience features. This flexibility could be enabled on a per-page, per-browser window, or per-session basis.

7.1.2.5 Allow Nuanced Privacy Classifications

Some commercial privacy management tools allow users to partition their browsing into private and public activities. The underlying assumption is that the vast majority of items are public with only a small subset needing to be password protected. However, we have found the privacy of visited pages to be much more nuanced. Almost all participants in our field studies utilized all privacy categories (public, semi-public, private, don't save) when classifying their visited web pages. This use of all four privacy levels validates the need for a more *nuanced approach* than the Public/Private or Save/Don't Save approach currently used in web browser convenience features and privacy management tools. Users of COLLABCLIO also indicated a desire for a more nuanced approach than public/private for privacy classification of their shared history files [90].

7.1.2.6 Support Multi-Tasking

Privacy tools tend to allow either a public mode or a private mode, and do not support tasks of mixed sensitivity. For example, the Safari Web browser allows users to enter a private mode during which traces of their activity will not be stored [135]. This mode is applied to all open browser windows. However, experienced users often maintain several open browser windows (or tabs in the case of tabbed browsers) as a means of in-session revisitation of web pages, to help manage the search process, and for multi-tasking [12]. Users may have multiple search goals [80] and may switch between windows and tasks,

particularly when pages are slow to download [12]. A privacy management system should support concurrent windows containing content of varying privacy sensitivities.

7.1.2.7 Support Varying Privacy Concerns

A generic approach to privacy management is not appropriate given the highly individual nature of privacy concerns. Ackerman et al. [8] suggest that an individualized approach is necessary in the domain of information sharing given the large variance in privacy reactions between participants. Our exploratory results confirm the necessity of a *personalized approach* in order to ensure that a privacy management system in this particular domain is effective. Results from our survey revealed variability in overall privacy concerns. During both field studies, we observed variability both in terms of participants' browsing behaviours and the privacy classifications of their visited pages. Our investigation into the privacy levels applied to different content categories of web pages also showed that a generic approach to privacy classification was not feasible in this privacy domain. Therefore, in order to accommodate varying privacy concerns, a privacy management solution should provide for personalization or be flexible in its usage.

We found that privacy comfort levels in a given situation depend on the person's relationship to the viewer, the level of control retained over input devices, their inherent privacy concerns, and the perceived sensitivity of potential visible information. Therefore, a privacy management system should be able to adjust to changing viewing contexts with minimal effort by the user.

7.1.2.8 Reduce the Burden of Privacy Management

Determining an effective way to manage users' visual privacy of browsing activity depends heavily on users' browsing behaviours. We found during our PG1 and PG2 field studies that participants tended to visit large numbers of pages, and have rapid bursts of activity. Therefore, while a simple approach to privacy classification might be to have users manually classify each generated trace of activity, such a manual approach would be time consuming and would interfere with the flow of browsing if done as traces are generated. One key to make sure that a privacy management system is usable will be to provide some mechanisms to reduce the burden of classifying visited traces with a privacy level.

In Table 29, we summarize the guidelines we believe should be considered when developing a visual privacy management system for use in web browsers. We reiterate how results of our exploratory studies and related literature support each guideline. We also indicate which of our findings from the exploratory research will provide the context necessary to implement the guideline in a visual privacy management system for the incidental information found within web browsers. Our primary concern when considering privacy management approaches is reducing the burden on the user. As privacy is a secondary consideration to the primary task (i.e. web browsing activities), we must ensure that privacy management does not interfere with this primary task. The burden of maintaining the privacy management system must be low or users will be unlikely to adopt such a system. Next, in section 7.2, we discuss the components of a privacy management system in light of this primary consideration.

Table 29. Summary of guidelines for a visual privacy management system, including the exploratory research findings and related literature in support of each guideline.

			Support for Guideline	Context Needed for Implementation
General Guidelines for Privacy Management Systems				
7.1.1.1	Increase visualization of settings and action		[39, 90, 91]	Chapters 4, 5, 6
7.1.1.2	Configuration within the context of action		[39, 90, 91]	Chapters 4, 5, 6
7.1.1.3	Provide opportunities for varying granularities of privacy control		[91]	Chapters 5, 6
7.1.1.4	Work within existing behaviours		[91]	Chapter 4, 5, 6
Guidelines for Visual Privacy Management within Web Browsers				
7.1.2.1	Increase visualization of settings		6.5 [39, 90, 91]	
7.1.2.2	Clearer explanations of feature functionality		6.5	
7.1.2.3	Intelligent default settings for context of use		5.5, 5.6, Chapter 6	
7.1.2.4	Reduce clutter within convenience features	Provide more usable configuration mechanisms	6.5	
		Provide a more selective approach to deletion	6.5	
		Provide fine-grained control at the time of browsing activity	6.5	Chapter 4
7.1.2.5	Allow Nuanced Privacy Classification		5.4	
7.1.2.6	Support Multi-tasking		4.2	
7.1.2.7	Support varying privacy concerns(personalization)		Chapters 4, 5, 6	
7.1.2.8	Reduce the burden of privacy management		Chapter 4	Chapter 5, 6

7.2 Components of a Privacy Management System

We considered three components to a privacy management system: classifying web browsing traces with a specific privacy level, filtering the information appropriately for the current viewing context, and providing methods for users to actively maintain the system. We next give some general discussion about these components, including possible approaches to privacy management as suggested by our exploratory studies. In later sections, we will investigate the feasibility of specific approaches.

7.2.1 Classification

A privacy management system will likely need some type of (semi-) automated privacy classification in order to be manageable. One approach would be to allow users to specify classification rules, an approach suggested for History sharing within COLLABCLIO [90]. Mechanisms such as content analysis and keywords could filter what information is saved or what privacy level is applied. For example, if a page included a subset of specified keywords, visits to that site would not be saved. Other heuristics could also be used. For example, there may be increased privacy concerns for secure websites. Another simple visual privacy enhancing mechanism may be to temporarily disable the storage of text for Auto Complete when on a secure site. A difficulty with this type of approach is that users may have a difficult time determining the coverage of generated rules [90]. It would be very important with this type of approach to ensure that feedback about privacy classification is given at the time of browsing so that misconceptions about coverage can be discovered. Furthermore, the mechanisms for configuration of the privacy rules should be readily accessible so that configuration and action are closely integrated [39, 91].

An automated approach to classification of visited web pages based on content category may also be feasible. With such an approach a user could assign a privacy comfort level to each category of content and the system could classify visited web pages with the content category and then automatically assign the associated privacy comfort level. We evaluate the feasibility of this approach in section 7.2.

Another approach to semi-automate privacy classification would be to leverage browser-window based patterns. As we observed during analysis of the data collected during the PG1 and PG2 field studies, participants tended to partition their activities between

browser windows, with private browsing often occurring in a single window. Additionally, within each browser window, participants exhibited streaks of browsing at a given privacy level, with relatively few transitions between levels. Given these patterns, one approach may be to allow users to open browser windows of different privacy levels. These windows could not only filter what incidental information is displayed, but could also tag new sites visited, similar to the extensional classification described in [90]. We evaluate the feasibility of this approach in Chapter 8.

Whatever the classification mechanism, users should be able to specify which visited pages should not be saved as those pages are encountered. Many participants in our exploratory studies indicated a desire for a more fine-grained approach to managing which information is recorded in their convenience features. During our field studies, participants tended to use the "don't save" category to indicate pages that were either inconsequential or extremely private. Allowing users to stop the recording of their activity for brief periods of time will help users remove some of the most sensitive sites from their convenience features and will also reduce what data is saved.

7.2.2 Appropriately Filtering Incidental Information

Users must be provided with mechanisms to specify the current viewing context so that only contextually appropriate content is displayed. With browser windows of different privacy levels (as in the PrivateBits solution presented in Chapter 8), this can be accomplished simply by opening up a window at an appropriate privacy level so that only appropriate content is display. While some users may find a simple hierarchical scheme appropriate (e.g., public, semi-public, private, don't' save); questionnaire responses during the field study indicate that other users would want to further partition their activities (e.g., work groups).

In addition to being individual, privacy comfort levels of participants during the IIP survey were found to be highly contextual. Interrelated factors of visual privacy included the participants' inherent privacy concerns, their relationship to the potential viewers, the level of control retained over input devices, and the sensitivity of the content. Furthermore, the location of browsing and the computing device in use impacted browsing activities, convenience feature settings, and preventative actions taken. Additionally, these results were

found to be highly individual. An automated privacy management system will require personalization in order for the system to discern the user's privacy concerns given their current viewing context so that the visible information can be adjusted accordingly. In section 7.4, we make an initial attempt at developing a predictive model of user's privacy concerns in a given situation.

7.2.3 Maintenance

Regardless of the classification and filtering approaches taken, users will require methods to check the accuracy of the classified traces of web activity and to adjust those privacy levels if necessary. Visualizations will be needed so users can easily view which traces may be revealed during browser use. It may be possible to use a content classification scheme (e.g., categories, keywords, URLs) to flag traces that may be inappropriately classified. Furthermore, many of our previous study participants indicated a desire to selectively delete traces of activity when limiting the information that might be displayed.

7.3 Exploration of an Automated Approach for Classification

We next present an exploration of the feasibility of content categorization as an automated approach to classify traces of browsing activity.

7.3.1 Utilizing Automatic Content Categorization

One method of automating the privacy classification of visited web pages may be to automatically classify pages as being one of several content categories and then to apply an appropriate privacy level to each category of content. However before such a system can be designed, the relationship between the privacy of web browsing traces and their content must be understood. If people hold common views on the sensitivity of content within a category, a general privacy management solution may be feasible. If not, a personalized solution may be appropriate, allowing each user to set a default privacy level for a category. However, personalization will only work if people are consistent within each category, applying a single privacy level to visited pages. Results from the PG2 field study were used to evaluate the feasibility of these approaches.

7.3.2 Assignment of Privacy Levels to Categories of Web Browsing

In section 5.2.1, we performed a cluster analysis which grouped the content categories of web pages based on how participants applied privacy levels to visited pages within each category. Examination of the cluster centers revealed the predominant privacy levels that characterize each cluster (C1: *public/don't save*, C2: *public*, C3: *semi-public*, C4: *mixture*, and C5: *private*). However, this cluster analysis provided no information as to whether the differences in privacy levels applied to visited pages within a category are a result of participants not being in agreement with each other as to an appropriate privacy level (*between participant consistency*) or not being individually consistent in how they assigned privacy levels to pages within that category (*within category consistency*).

For our computations of consistency, we report on normalized data on a per-participant basis. Normalized data is necessary as some participants visited many more pages within a category than others. For each participant with 10 or more pages of browsing in a category, we determined the *predominant privacy level* that they applied to their browsing in that category and calculated the percentage of pages that were classified at that privacy level. We omitted instances where a participant had fewer than 10 page visits in a category; these categories were deemed to be less relevant to participants and their consistency less reliable.

7.3.2.1 Between Participants Consistency

Between participants consistency examines how much agreement there is between participants in their privacy classification of page visits in a category. We compared the predominant privacy level applied by participants within each category (see Table 30 for a breakdown of the number of participants that classified the majority of their page visits in the category with each privacy level). Complete agreement between participants with respect to which privacy level was applied was found in only 4 of the 30 categories (only categories that contained visits from at least two participants were examined).

Furthermore, over half of those categories (16/30) have a subset of participants whose predominant privacy level in that category that was not consistent with the category's cluster membership. The highlights in the privacy level cells in Table 30 represent the expected predominant privacy levels according to the cluster membership of the category. For example, for On-line Games, the overall application of privacy levels resulted in this

Table 30. Descriptive statistics of visited web pages (PG2) by content category, including overall number of pages, number of participants with page visits (total, 10+ pages), within category consistency, accuracy, predominant privacy levels applied, and cluster membership. Highlights show expected dominant privacy levels based on cluster membership.

Category	Overall page total	# part.		Consistency (%)	Accuracy (%)	Predominant privacy level				Cluster
		Total	10+ pages			Public	Semi-Public	Private	Don't Save	
Search Engines/Portals	6310	15	15	61	46	6	3	4	1	C4
Education	3315	15	14	65	57	10	3	1		C4
Email	5082	14	14	81	77	1	5	8		C5
Reference	2055	14	13	76	51	8	3		2	C4
News/Media	1320	14	7	96	95	7				C2
Shopping	770	14	10	80	38	6	3		1	C1
Arts/Entertainment	665	14	12	81	59	5	3		4	C1
Society/Lifestyle	1136	13	8	93	10	5		1	2	C1
Web Advertisement	158	12	3	71	55	2	1			C1
Computers/Internet	146	12	5	66	55	4	1			C4
Financial Services	510	11	10	90	75		1	8	1	C5
Government/ Legal	385	11	5	88	78	2	3			C2
Web Communication	660	10	6	76	32	3	1	2		C4
Sports/Recreation/Hobbies	431	10	5	91	39	3	1		1	C4
Travel	366	10	7	80	45	3	2	1	1	C4
Software Downloads	236	10	6	83	61	5			1	C2
Health	165	10	6	92	16	3	2	1		C4
News Group	1303	9	3	78	70	1	2			C3
Job Search/ Career	449	9	4	80	86		2	2		C3
Business/Economy	178	8	4	84	60	1	1	1	1	C4
Religion	127	8	2	78	44	1			1	C4
Online Games	520	7	5	90	74	2	1	1	1	C2
Streaming Media/MP3	148	7	4	76	69	2	1		1	C1
Web Content Management	598	6	4	80	--	1	2	1		--
Political /Activism/Advertising	57	6	2	95	71	2				C2
Dating/ Personals	600	5	4	88	18		1	3		C5
Internet Auction	101	5	3	92	95	1	2			C3
Humor/Jokes	77	5	1	79	73		1			C3
Restaurants/ Dining/Food	279	4	3	99	88	1	1		1	C2
Pornography	258	4	2	88	86			2		C5
Web Hosting	60	4	2	80	29		2			C3
Real Estate	147	3	1	100	99	1				C2
Brokerage/Trading	110	3	1	95	0	1				C2
Intimate Apparel/Swimsuit	94	2	1	97	95			1		C5
Other	229	13								
Empty Window	21115	15								
Total	31160	15								

category falling in Cluster 2 (public). If participants were consistent with each other, we would expect all participants to have public as their primary privacy level (hence the highlight in the Online Gaming/Public cell in Table 30). However, of the 5 participants with 10 or more online gaming page visits, an examination of their predominant privacy levels reveals that only 2 of the participants labeled most visited pages as *public*; the other 3 participants each labeled most of their visited pages with one of the other privacy levels (*semi-public*, *private*, and *don't save).*

7.3.2.2 Within Category Consistency

Within category consistency examines how consistent participants were in assigning privacy levels to pages in that category, regardless of which privacy level was applied predominantly. For each category, for each participant with 10 or more page visits, we computed the consistency in each instance as the number of pages classified at the primary privacy level divided by the total number of page visits, thus normalizing the consistency on a per-participant basis. The overall consistency for each category was obtained by averaging the per-participant results. Across all categories, the average consistency was 81% (61-100%, see Table 30 for per-category results). For many categories, participants may be able to set a default privacy level that classifies most pages accurately, but some categories (e.g., Search Engines/Portals, Education) are problematic.

7.3.2.3 Website Classification Task

During the theoretical website classification task in the PG2 study, participants assigned a single privacy classification to each of the web categories. The results are shown in Figure 46, which illustrates how differently participants felt about the sensitivity of the categories. If all participants had similar privacy concerns about content categories, we would expect to see each bar in a single colour. However, all participants used the same classification in only two categories (News/Media, Computers/Internet). It should be noted that the classification task was completed in terms of privacy of content, not relevance. Therefore, use of *don't save* may be more likely an indication that a category was considered 'extremely private' rather than 'irrelevant'. We cannot be sure of the extent to which the dual nature of this privacy level contributed to classification inaccuracies.

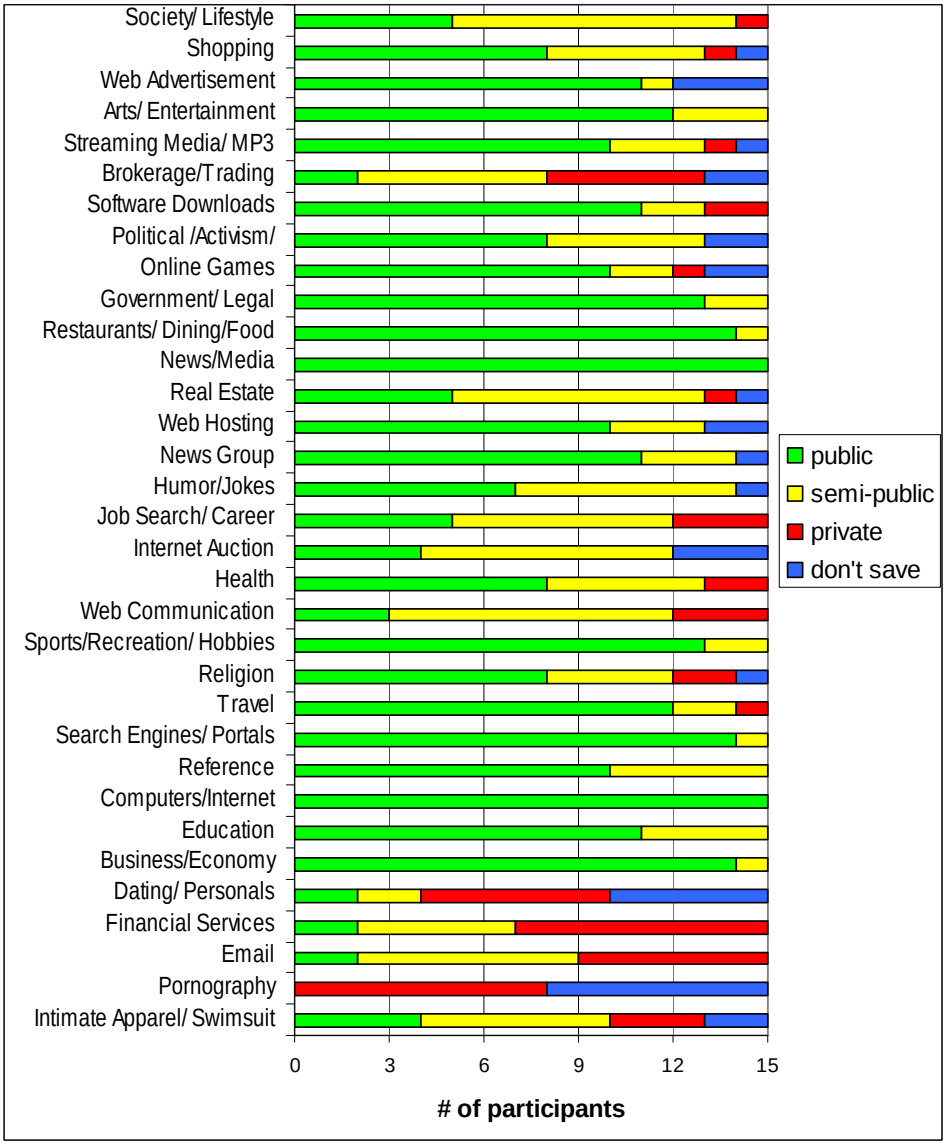


Figure 46. Results of theoretical website category privacy classification task.

7.3.2.4 Classification Accuracy

We examined how accurate the classification task was as a predictor of a participant's actual labeling of their browsing. For each participant, we computed accuracy as the number of web page visits that were labeled at the same privacy level that the category was labeled during the theoretical classification task. Overall, 57.8% of the page visits were classified accurately (see Table 30 for per category results, no accuracy results are available for web content management as it was not a category used in the classification task). Accuracy varied

greatly by category, ranging from 0% (Brokerage/Trading) to 98.6% correct (Real Estate). Accuracy also varied greatly by participant (ranging from 36% to 82%).

7.3.3 Feasibility of a General Privacy Management Scheme

For a general privacy management system (i.e. one size fits all) to be suitable, there would need to be universal agreement between users on an appropriate privacy classification for each category of web page. The results of the theoretical classification task in PG2 showed that participants differed greatly in their privacy classifications of categories; indeed only two of the categories had complete agreement between participants. Examination of the actual privacy labels applied by participants and the clusters that formed (C1: public/don't save, C2: public, C3: semi-public, C4: mixture, C5: private) revealed that some categories did exhibit basic agreement among participants. However even for those categories that were predominately labeled with one privacy level (e.g., categories in clusters C2 (public), C3 (semi-public), and C5 (private)), there were some pages that were labeled differently. Inconsistencies were found to be both between participants (with respect to the predominant privacy level) and also within participants' classifications. This was particularly true for the categories in C1 (public/don't save) and C4 (mixture) where a variety of privacy levels were applied. As these two clusters account for over 50% of the pages visited, a general privacy management scheme would not be effective.

7.3.4 Feasibility of a Personalized Privacy Management System

For a personalized privacy management system to be feasible, participants would need to be fairly consistent at their desired privacy level within each category of web browsing activity. Many categories were very consistent; 12/34 categories examined had greater than 90% consistency. However, many categories exhibited higher inconsistencies; 13 of the categories have more than 20% inconsistency between the actual labels applied and the predominant privacy level. This was most pronounced for those categories in clusters C1 and C4 (public/don't save and mixture) which tended to have lower consistency results.

Participants would also need to be able to specify the default privacy level for each category of web browsing. We examined how accurate the classification task was as a predictor of a participant's actual labeling of their browsing in a category. For each participant, we computed accuracy as the number of web page visits that were labeled at the

same privacy level that the category was labeled during the theoretical classification task. Prediction accuracy varied greatly and some participants were unable to correctly predict the majority of their labeling. Some of the inaccuracy is due to categories with low consistencies; if the pages in a category are fairly evenly divided between two or more privacy levels (e.g., categories in clusters C1 and C4), any predicted privacy level will fail to accurately classify the majority of pages.

Clearly, the consistency results must be improved for those categories with low consistency ratings as well as the participant accuracy in assigning default privacy levels in order for a personalized privacy management system to be effective. The characteristics of the web site categories that lead to inconsistent and inaccurate privacy ratings are discussed next and then recommendations are given for increasing accuracy.

7.3.4.1 Reasons for Inconsistency and Inaccuracy

Recent research (such as [8]) has been cautioning that actual behaviour with respect to privacy practices often does not follow stated privacy concerns. However, this was likely not a major source of inconsistency during this study due to the theoretical nature of both the questionnaires and participants' application of privacy levels to their web browsing. Any effects due to social desirability (i.e. participants specifying a privacy level that they feel is the socially acceptable answer) should have been mirrored in both the theoretical classification task and the classification of their actual web browsing. One cause of inaccuracy may have been that the example websites and category descriptions given in the theoretical classification task may not have adequately conveyed to participants what types of sensitive content may be visible.

Another potential cause of the inconsistency and inaccuracies within website categories may be due to the "it depends" nature of the semi-public privacy level. The uncertainty of whether visited web pages within a category should be public or private is often due to what is appropriate for the various categories of potential viewers. However, it may also be due to the variety of potential content in a given category. The potential viewing context is therefore partially resolved when a specific page is viewed. One example where this may have occurred was with the Web Communication category. This category was predominately predicted to be semi-public and in actuality, the dominant privacy level was split between public (3/6 participants), semi-public (1/6) and private (2/6).

Similarly, the dual nature of don't save (irrelevant or extremely private) causes inconsistencies related to privacy. In some cases it is applied as a fourth privacy level (extremely private) and in other cases it was applied as a mechanism for not cluttering the convenience features with irrelevant pages (i.e. those that a participant would never bother to visit again). This dual nature was intentional during the study, allowing participants to classify the end result (not having a page saved) without having to admit to extremely sensitive browsing, Much of the inconsistency (particularly for the public/don't save cluster (C1)) may be resolved if the dual nature is separated.

There were several characteristics of web page categories that led to inconsistencies and inaccuracies. Some were very *general* such that sites with very different content would appear in the same category. For example, the category News Group may be applied to forums that discuss very different topics in terms of sensitivity. The variable content being accessed at a Software Download site (e.g., free software updates, purchased products, warez) may have reduced the consistency for this category. Websites may also be very complex and are often dynamic in nature. Such sites may have varying content sensitivities depending on the content visible on a given page or at a given time. For example a News/Media site may have specific news stories that may be more sensitive than others. The content must be examined to determine the appropriate privacy sensitivity with respect to future viewing. Users may be unable to give a single default privacy level for these categories.

Further analysis of the categories with lower results revealed that many were multi-purpose (e.g., a general university site may have sub pages related to specific assignments and grades), had varying tasks associated (e.g., a travel page can be informational or a transaction such as a secure flight booking), or had sub-pages at varying content sensitivities (e.g., search results reveal more sensitive content than the search engine home page). For example, a page categorized as Brokerage/Trading may give general information or contain details about an individual's personal transactions. The Brokerage/Trading category had 0% accuracy. Examination of the data revealed that the 3 participants with browsing in this category were conducting diverse activities, from visiting informational sites (e.g., finance.yahoo.com) to logging in to conduct secure trading transactions. The large number of public pages reflects informational pages, while the secure transactions were primarily classified as private.

Often transactional web sites have an entry page that is less sensitive than the sub pages. Similarly, the categories Financial Services and Email were primarily classified as private, so clustered with categories containing socially inappropriate sites (e.g., Pornography). These categories were considered to be private due to the access of personal content, not sexually explicit material. For sites in these categories, one marker of content sensitivity appeared to be whether or not a secure transaction was taking place. Across all browsing, there were 6963 secure pages (https); categories that had a high proportion of secure pages included Email (71%), Financial Services (74%), Web Communication (46%), Search Engines/Portals (42%), Brokerage/Trading (17%), and Travel (16%). Overall, 57% of secure pages were classified as private and 13% as public. The converse was true for pages that were not secure (14% private, 52% public); the proportion of don't save and semi-public pages remained consistent. Login pages may serve as markers for the transition between more public viewing and the subsequent secure pages that may be more private in nature.

7.3.4.2 Recommendations to Increase Accuracy

To increase accuracy, we believe that two main issues must be resolved. The first is finding methods of further categorizing websites to resolve inconsistencies due to the generality, multiple task purposes and dynamic nature of sites. The second is improving participants' ability to predict the privacy levels they will apply.

As initially discussed in section 7.2.1, some heuristics exist that may help resolve some of the inconsistencies within categories. For example, for those sites that are very general or dynamic, being able to categorize the content at the sub-page level (e.g., keyword analysis) may improve accuracy. In order to distinguish between informational web sites and transactional sites, it may be necessary to identify log-in pages or secure pages (https) and modify the content accordingly. There may be other triggers that precipitate a switch between privacy levels. For example, pages that are viewed very briefly during a burst may be only used for navigation the user's quick scan of the page may indicate it is irrelevant; such pages may be candidates for the "don't save" category as they may have little relevance for future revisitation purposes.

Whatever the categorization scheme, it must be effectively communicated to users. While the classification scheme used provided both descriptions and example web sites, in

some cases it did not appear to be apparent to participants just how diverse categories were with respect to the types of pages and content that may be included. When determining an appropriate privacy level, the cost of others viewing traces of a previous web visits can only be determined if it is clear to participants what sorts of information may be visible.

7.4 Exploration of an Automated Approach for Filtering

Throughout Chapters 5 and 6, an initial *model of visual privacy* during web browsing was developed which may serve as the basis for a future intelligent systems approach. We have shown that there is a great deal of variation between individuals and that the variation transcends to their privacy comfort level; that is, no two people are alike and their privacy concerns and situations aren't alike. Our goal is to build a model of incidental information privacy that could be used by a privacy management system to control which traces of previous activity appear in a web browser.

The model could be used to classify new users of the system according to their responses to a series of questionnaires. Components of the model could include such characteristics as a person's inherent privacy concerns, their perceived sensitivity of different content types of web sites, the frequency and type of viewers/users of their display, and the actions they perform within their web browsers. Some of this information may be able to be generated as defaults given the user's privacy dispositions and usage scenarios, but the user should be able to modify the defaults.

From this information, a privacy management system could determine an appropriate default privacy comfort level. This could be an overall level according to their inherent privacy concerns, but it could also be adjusted for each viewing context (e.g., boss came into the room). Different profiles may be appropriate depending on device and location of use. The privacy comfort level generated by the system would work in conjunction with previously classified content to filter what information is visible in the web browser (e.g. History, Favorites, Auto Complete). Simplified configuration mechanisms may be possible for those participants not concerned along a particular factor (e.g., level of control).

In Chapter 6, we investigated how the different dispositional and situational variables impacted inherent privacy concerns and participants' reported browsing activities in the IIP

survey and their actual activities and privacy levels applied in the field studies. We next consider the IIP survey data to see if we can begin to develop a predictive model that might be used to determine an appropriate privacy level in a given viewing situation. Such a predictive user model would include some combination of the dispositional factors to determine user preferences and the situational factors to determine an appropriate privacy decision.

While some researchers have found that dispositional variables and inherent privacy concerns impact privacy actions in a given situation [98, 137], others have been less successful at finding a correlation [55]. For example, Hann et al. [55] examined the cost-benefit tradeoff made by undergraduate students when releasing information to websites. They found that while situational variables such monetary awards and future incentives impacted the value participants placed on their data; dispositional variables including gender, contextual knowledge (e.g. knowledge of cookies, knowledge of anonymous browsing), and trust propensity did not. Malholtra et al. [98] developed a model of Internet users's information privacy concerns with respect to consumer trust of marketers and willingness to reveal information. Validation of their model showed that inherent privacy concerns attributed for approximately 10% of the variability for behavioral intention. Sheehan [137] correlated the total score for privacy concerns across 15 different situations (total score ranging from 15-105) with 7 privacy related behaviours (e.g. providing inaccurate information when registering for web sites) and found that behaviours were impacted by gender. Although their was no significant difference between the mean overall privacy concerns, as privacy concerns increased, women were less likely to have a positive correlation with privacy preserving measures than men were (2/7 behaviours for women, 7/7 behaviours for men). We will incorporate both situational and dispositional variables in our predictive model, using our findings from Chapters 5 and 6 and the model of incidental information privacy we have developed to guide inclusion of variables.

7.4.1 Multiple Regression Analysis

Multiple regression analysis has been found to work well when investigating complex real-life questions rather than laboratory-based research questions [120]. This analysis technique can explore the relationships (and inter-relationships) between several

independent or predictor variables with a dependent variable of interest. It is appropriate for our analysis due to the multi-faceted nature of privacy concerns.

To develop our predictive models, we chose to use Standard Multiple Regression. With this technique, all the independent variables are input at one time; each is evaluated for its predictive power over and above all the other independent variables [120]. This technique is appropriate to determine how much variance a block of variables accounts for, as well as the unique variance in the dependent variable explained by each independent variable.

There are several underlying assumptions that must be satisfied when using multiple regression analysis. For generalizability, a sufficiently large sample size is required. There are various guidelines given for an appropriate sample size including 15 subjects per independent variable and a base of 50 subjects plus 8 subjects per independent variable [120]. Respectively, these guidelines would suggest that 10 or 13 independent variables would be appropriate given our 155 participants in the IIP survey. If the dependent variable is skewed, more subjects are required.

Multiple regression is also sensitive to multicollinearity and singularity [120]. Multicollinearity occurs when the independent variables are highly correlated ($r >= 0.9$). If this is the case, it is appropriate to either use a single variable or a combined score, depending on which has a greater correlation with the dependent variable. Multiple regression analysis is also very sensitive to outliers in both the independent and dependent variables. Therefore outliers should be re-coded to be high, but within range of the other values.

7.4.2 Predictive Model Results

The requirement of a large sample size is an issue for our analysis as we are unlikely to find an effective general model across all participants given the individual differences we've found. However, when we begin to break down our participants into subgroups, such as their inherent privacy concerns or other dispositional and situation attributes, it becomes less appropriate to conduct multiple regression analysis. We encounter difficulties as we are very limited in how many independent variables we may include and the dependent variable becomes more skewed in the sub-group. Nevertheless, some interesting results have emerged with our initial attempts at modeling.

In Chapters 5 and 6 we focused on participants' responses to the embarrassing and neutral scenarios to establish their inherent privacy concerns and examine the impact of various dispositional and situational variables on their inherent concerns. We now turn our attention to the privacy comfort levels participants reported for the browsing scenario that had them reflect on their *usual* recent web browsing activities. We first present a model predicting the overall privacy comfort level participants reported when reflecting on their recent web browsing activities. We then examine more contextualized models for two given viewing situations: a spouse/significant other as the viewer and a supervisor as a viewer.

7.4.2.1 General Model

We begin by generating a general predictive model for participants' average privacy comfort level for the IIP survey scenario which had them reflect on their privacy comfort level (PCL) if someone were to view their *usual browsing* (u_avg). For each participant (n=154), we computed a value for u_avg by averaging their responses across the 15 contexts queried (i.e., 3 levels of control over input devices, 5 types of viewers).

When developing the predictive model, our goal was to include those variables from our model of visual privacy concerns that may have impacted participants overall PCL for the usual browsing scenario. We identified the sensitivity of the potentially visible content and participants' inherent privacy concerns as the primary factors of interest. We did not anticipate that variables relating to level of control retained or relationship to the viewer would contribute much to the predictive model as the dependent variable (u_avg) was computed by averaging comfort levels across the 15 control/viewer situations. Similarly, as the usual browsing scenario was not situated according to location or device, we did not believe variables related to this context would be pertinent to the model.

We first examined which measures from the IIP survey would be indicative of the sensitivity of the browsing being considered. One difficulty we had was that the *usual browsing scenario* question did not have participants consider a single browsing location; instead, they considered their privacy comfort level for their browsing as whole. However, the survey questions investigating which specific browsing activities were conducted were given within the context of location (e.g., home or work/school). Therefore, we needed to determine an appropriate independent variable for use in our regression analysis. We began by examining if there was a correlation between the percentages of browsing participants reported

conducting for personal purposes and their PCL for the usual browsing scenario (u_avg); no correlation was found. We next investigated whether correlations existed between each individual activity being reported and participants' PCL. We considered a participant to have conducted each activity if they reported this activity for at least one of their locations of browsing. We found negative correlations between participants' comfort level and whether the participant reported viewing entertainment information (ent; r=-.139, p=.037) and erotica (ero: r=-.308, p=.000).

We also investigated whether a composite variable incorporating the breadth and sensitivity of browsing activities would have a stronger correlation with participants' PCL for the usual browsing scenario than the individual browsing activities. We calculated sensitivity values based on the overall percentage of participants who partitioned each activity between work and at home (section 6.4.1). Those activities that were conducted mostly at home we considered to have a higher sensitivity than those conducted in both locations. It is important to note that this judgment of sensitivity is across participants and may not necessarily be reflective for any individual participant. Sensitivity values for each activity ranged from 0.945 for erotica to 0.059 for email.

We calculated overall sensitivity values for each participant for home (s_home) and away (s_away) by summing the sensitivity values for each activity reported. If an individual reported doing all nine activities in a location, the maximum overall sensitivity value was 4.093. We also computed the differences between browsing conducted at home and away from home (s_diff) in order to gain a sense of how each individual changed their activities between home and away. Table 31 gives descriptive statistics for these composite variables including their mean, range, and correlation with the privacy comfort level for the usual browsing scenario. We found negative correlations between u_avg and the overall sensitivity

Table 31. Details of composite variables incorporating the sensitivity of browsing activities, and their correlation with participants' privacy comfort level for the usual browsing scenario.

Measure				Correlation with u_avg	
Description	**Variable**	**Mean**	**Range**	**r**	**p**
Sum of sensitivity values of activities conducted at home.	s_home	2.743	0 to 4.093	-.235	.003
Sum of sensitivity values of activities conducted away from home	s_away	1.1317	0 to 4.093	--	n.s.
Difference between sensitivity of browsing conducted at home and away from home	s_diff	1.3426	-3.148 to 4.093	-.189	.019

of browsing conducted at home and the differences in the overall sensitivity between the locations. Interestingly, there was no correlation between the sensitivity of activities conducted away from home and u_avg. For the purposes of the regression analysis, we will use s_home as it has the highest correlation with u_avg.

In terms of dispositional factors related to inherent privacy concerns, we investigated correlations between the average PCL for the usual browsing scenario and the average PCL across the neutral and embarrassing scenarios, the overall amount of contextual differences, the amount of contextual differences attributable to the scenario, computer experience, gender, and technical level. Positive correlations existed between u_avg and the average PCL for the neutral and embarrassing scenarios (ne_avg: r=.519, p=.000). There were marginally significant correlations between u_avg and the amount of contextual differences attributable to scenario (scen_diff: r=.104, p=.100), computer experience (comp_exp: r=.108, p=.091), and total devices (total_dev: r=-.132, p=.051). No other correlations were found for the variables investigated.

We examined correlations between the general situational variables of laptop use in multiple locations and total devices. No significant correlations were found. We also investigated whether the frequency with which participants had different types of viewers of their display correlated with their reported privacy comfort level for the usual browsing scenario. We summed the frequency reports for the various types of viewers (vwr_frq_sum) and users (user_frq_sum) that participants reported. In an effort to get a sense of the extent that participants' displays were viewed and computers used, we totaled the frequency reports (never: 0, rarely: 1, monthly: 2, weekly: 3, daily: 4) for all ten types of viewers/users. For any categories with missing values, the missing value was replaced with a 0 (never). We made the assumption that a viewer category was most likely skipped because it was not applicable (e.g., participant has no spouse/significant other). For the sake of our analysis, it only matters the frequency with which types of viewers are actually viewing the display or using the computer, not the reasons why. The user frequency sum was negatively correlated with u_avg (r=-.137, p=.046); however, there was no significant correlation with the viewer frequency sum.

We used standard multiple regression analysis in order to develop a predictive model for participants' average PCL as they reflected on their usual browsing (u_avg). Before

beginning analysis, outliers were rescaled for the dependent and independent variables. Our initial model included those variables with at least a marginally significant correlation to u_avg. Table 32 gives a summary of the measures considered.

Table 32. Summary of measures included in the multiple regression analysis for the general predictive model and their correlation with participant's privacy comfort level for the usual browsing scenario.

Measure					Correlation with u_avg	
Description	**Variable**	**N**	**Mean**	**SD**	**r**	**p**
Average privacy comfort level for the usual browsing scenario across all viewing/control contexts	u_avg	154	5.01	1.33	--	--
Average of participants privacy comfort level for the neutral and embarrassing scenarios across all viewing/control contexts	ne_avg	155	4.52	0.98	+.519	.000
Value of 1 if reported viewing erotica in any location	ero	155	0.42	0.50	-.308	.000
Value of 1 if reported viewing entertainment information in any location	ent	155	0.94	0.25	-.139	.037
Magnitude of difference in PCL between the neutral and embarrassing scenarios	scen_diff	155	2.54	1.40	+.104	.100
Sensitivity value of activities conducted at home	s_home	155	2.49	1.03	-.235	.003
Sum of the frequency reports for the 10 categories of users of participants computers	user_frq_sum	155	6.32	4.63	-.137	.046
Years of computer experience	comp_exp	154	12.49	5.33	+.108	.091
Total devices used across all locations	tot_dev	155	2.54	1.00	-.132	.051

The initial model with the best fit included erotica and entertainment as separate independent variables rather than our composite variable of browsing sensitivity at home (s_home). In order to avoid over fitting to the data and thereby reducing generalizability, we manually pruned the model, at each step remove those independent variables with a unique contribution to the model that was not at least marginally significant ($p < .10$) as indicated through t tests of the beta weights.

Table 33 provides the regression models predicting privacy comfort level for the usual viewing scenario, showing both the initial model and the final pruned model. An examination of the beta weights for the final model reveals that the average of the neutral and embarrassing scenarios had the largest unique contribution, followed by the negative impact of viewing erotica, the positive amount of contextual differences related to scenario and the negative impact of viewing of entertainment related sites. This model accounts for 37.6% of the variability in the privacy comfort levels for the usual browsing scenario.

Table 33. Regression model predicting privacy comfort level for the usual viewing scenario (general case).

	Initial Model			Final Model		
Measure	**B**	**Beta**	**Sig**	**B**	**Beta**	**Sig**
Intercept	2.578		.000	2.489		.000
ne_avg	+.704	+.522	.000	+.682	+.506	.000
ero	-.955	-.356	.000	-.754	-.282	.000
ent	-.828	-.154	.028	-.650	-.121	.063
scen_diff	+.160	+.169	.010	+.144	+.152	.019
s_home	+.160	+.123	.177			
user_frq_sum	-.028	-.097	.148			
comp_exp	-.018	-.071	.306			
tot_dev	+.008	+.006	.089			
Model Summary	$R=.640$, $R^2=.409$, Adj. $R^2 = .376$			$R=.627$, $R^2=.393$, Adj. $R^2 = .376$		
	$F_{8, 144} = 12.461$, $p=.000$			$F_{4, 149} = 24.077$, $p=.000$		

An examination of how well this model fits the data did reveal some problems. These may be due in part to the inclusion of participants across all privacy segmentations. For example, an examination of the residuals revealed one outlier with a high residual (3.5). This outlier was a *privacy unconcerned* participant; the model predicted an average privacy comfort level of 3.5 for this participant, but this participant's average privacy comfort level for the usual scenario was 7.0.

While our initial attempts at modeling are promising, it is clear that further refinement is required before such a model can be developed for use in an adaptive privacy management system. In future studies, it will be important to have a larger sample size and to further contextualize the viewing situations when asking participants to report on their comfort levels if their usual browsing was to be viewed. To more fully develop the model, we will need to have separate questions for each location and a better indication of the sensitivity of the content being considered.

7.4.2.2 Contextualized Model

We also investigated how the predictive model might change if we contextualized the privacy comfort level for the usual scenario for a specific type of viewer, averaging reported privacy comfort levels across the three levels of control. We developed two predictive models representing the range of privacy concerns for viewer types: spouse/significant other and supervisor.

We began with spouse/significant other; 148 of the participants reported a privacy comfort level for the usual browsing scenario for this category of viewer/user. As with the general model, we first examined the correlations between the dependent variable (u_avg_sp) and independent variables relating to browsing activities, general dispositional and situational variables, and inherent privacy concerns. Table 34 gives a summary of the measures included for the predictive model contextualized with spouse as the viewer.

Table 34. Summary of measures included in the multiple regression analysis for the predictive model contextualized for spouse as a viewer, including their correlation with participant's privacy comfort level for the usual browsing scenario.

Measure					Correlation with u_avg_sp	
Description	**Variable**	**N**	**Mean**	**SD**	**r**	**p**
Average privacy comfort level for the usual browsing scenario across all control contexts, where viewer = spouse	u_avg_sp	148	5.80	1.35	--	--
Average of participants privacy comfort level for the neutral and embarrassing scenarios across all control contexts where viewer = spouse	ne_avg_sp	151	5.55	1.26	+.596	.000
Value of 1 if reported viewing erotica in any location	ero	155	0.42	0.50	-.195	.009
Value of 1 if reported conducting online shopping in any location	shop	155	0.81	0.39	+.108	.095
Value of 1 if reported viewing medial information in any location	med	155	0.76	0.43	+.136	.050
Magnitude of difference in PCL between the neutral and embarrassing scenarios attributed to differences by viewer	view_diff	155	2.18	1.38	+.301	.000
Value of 1 for participants whose viewer concerns contributed to more than 25% of their total concerns	vwr_con	155	0.64	.482	+.286	.000
General privacy comfort level for viewer = spouse (non-contextualized for level of control or content sensitivity)	gc_vwr_sp	134	5.67	1.41	+.453	.000
General privacy comfort level for user = spouse (non-contextualized for level of control or content sensitivity	gc_usr_sp	125	5.69	1.57	+.406	.000
Frequency report for viewer = spouse	vwr_frq_sp	155	1.99	1.68	+.155	.030
Frequency report for user = spouse	user_frq_sp	155	1.67	1.66	+.126	.063
Years of computer experience	comp_exp	154	12.49	5.33	+.124	.066
Value of 1 if "away" reported as the majority location of use	maj_loc	155	0.43	0.50	+.153	.031

As before, to reduce over fitting of the model to the data, we iteratively removed from the initial model those variables that did not have at least a marginally significant unique contribution to the variable until the model stabilized and the loss of a variable

decreased the adjusted r^2 value. Table 35 provides the regression models predicting privacy comfort level for the usual viewing scenario, showing both the initial model and the final pruned model.

Table 35. Regression model predicting privacy comfort level for the usual viewing scenario (viewer=spouse).

	Initial Model			Final Model		
Measure	**B**	**Beta**	**Sig**	**B**	**Beta**	**Sig**
Intercept	.846		.139	1.079		.027
ne_avg_sp	+.477	+.446	.000	+.505	+.472	.000
ero	-.243	-.089	.223	-.313	-.115	.074
shop	+.041	+.013	.860			
med	+.225	+.065	.392			
view_diff	+.005	+.005	.962			
vwr_con	+.410	+.147	.138	+.398	+.142	.032
gc_vwr_sp	+.266	+.279	.002	+.315	+.325	.000
gc_usr_sp	+.080	+.093	.299			
vwr_frq_sp	+.026	+.032	.745			
user_frq_sp	-.032	-.039	.684			
comp_exp	-.013	-.051	.476			
maj_loc	+.266	+.098	.164			
Model Summary	$R=.714$, $R^2=.510$, Adj. $R^2 = .455$			$R=.704$, $R^2=.496$, Adj. $R^2 = .480$		
	$F_{12, 108} = 9.363$, $p=.000$			$F_{4, 128} = 31.483$, $p=.000$		

An examination of the beta weights reveals that the average of the neutral and embarrassing scenarios with spouse as the viewer had the largest unique contribution, followed by the general comfort level given for a spouse as a viewer, whether the participant was classified as viewer concerned and finally the negative impact of viewing erotica. This model accounts for 48.0% of the variability in u_avg_sp. It is interesting to note that erotica contributed less to the model for spouse than the general model across all types of viewers.

Again, examination of the residuals revealed some problems with the fit of this model to the data, which may be due in part to the inclusion of participants across all privacy segmentations. For example, the data from two participants classified as *privacy fundamentalists* had a high residual (-3.3, -3.1); the model predicted an average privacy comfort level of 5.7 and 6.1 respectively, but for these participants, their average privacy comfort level for the usual scenario when considering their spouse or significant other was 2.4 and 3.0.

We proceeded in the same fashion and examined the predictive model for the viewer category of supervisor (n=148) which had the lowest overall privacy comfort levels of the

viewer categories when participants reflected on their usual web browsing activity (u_avg_sv). We wanted to examine the extent that the model might change given the very different privacy comfort levels reported for these two types of viewers. Table 36 gives a summary of the measures included in the multiple regression analysis for the predictive model contextualized with supervisor as the viewer.

Table 36. Summary of measures included in the multiple regression analysis for the predictive model contextualized for supervisor as a viewer, including their correlation with participant's privacy comfort level for the usual browsing scenario.

Measure					Correlation with u_avg_sv	
Description	**Variable**	**N**	**Mean**	**SD**	**r**	**p**
Average privacy comfort level for the usual browsing scenario across all control contexts, where viewer = supervisor	u_avg_sv	148	4.24	1.70	--	--
Average of participants privacy comfort level for the neutral and embarrassing scenarios across all control contexts where viewer = supervisor	ne_avg_sv	150	3.76	1.33	+.519	.000
Percentage of browsing reported being conducted for personal reasons	pers_brws	155	43.9	23.8	-.136	.049
Sensitivity value of activities conducted at home	s_home	155	2.49	1.03	-.191	.010
Difference in sensitivity values of activities conducted at home and while away from home	s_diff	155	1.37	1.24	-.147	.038
Value of 1 if reported viewing erotica in any location	ero	155	0.42	0.50	-.256	.001
Value of 1 if reported viewing entertainment information in any location	ent	155	0.94	0.25	-.175	.017
Value of 1 if reported viewing medial information in any location	med	155	0.81	0.39	+.108	.095
Value of 1 for participants whose scenario concerns contributed to more than 25% of their total concerns	scen_con	155	0.68	.47	+.120	.073
General privacy comfort level for viewer = supervisor (non-contextualized for level of control or content sensitivity)	gc_vwr_sv	119	4.39	1.76	+.303	.000
General privacy comfort level for user = supervisor (non-contextualized for level of control or content sensitivity	gc_usr_sv	101	4.73	1.82	+.321	.001
Years of computer experience	comp_exp	154	12.49	5.33	+.109	.094
Technical level	tech_lvl	136	1.08	.90	+.133	.066
Value of 1 if "away" reported as the majority location of use	maj_loc	155	0.43	0.50	+.232	.002

Table 37 provides the regression models predicting privacy comfort level for the usual viewing scenario, showing both the initial model and the final pruned model. An examination of the beta weights reveals that the average of the neutral and embarrassing scenarios with supervisor as the viewer had the largest unique contribution, followed by the

negative impact of viewing erotica, whether participants indicated they performed the majority of their browsing away from home, the difference between the overall activity sensitivity values for home and away, and finally the negative impact of viewing entertainment related pages. This model accounts for 35.4% of the variability in u_avg_sv.

Table 37. Regression model predicting privacy comfort level for the usual viewing scenario (viewer=supervisor).

	Initial Model			Final Model		
Measure	**B**	**Beta**	**Sig**	**B**	**Beta**	**Sig**
Intercept	1.860		.086	2.481		.000
ne_avg_sv	+.567	+.444	.000	+.650	+.512	.000
pers_brws	+.002	+.026	.800			
s_home	+.053	+.032	.862			
s_diff	+.154	+.113	.446	+.231	+.168	.047
ero	-.995	-.291	.033	-.970	-.283	.000
ent	-.941	-.137	.167	-.899	-.131	.055
med	+.241	+.056	.613			
scen_con	+.219	+.060	.523			
gc_vwr_sv	+.071	+.074	.563			
gc_usr_sv	+.028	+.030	.812			
comp_exp	-.009	-.030	.753			
tech_lvl	+.218	+.116	.256			
maj_loc	+.520	+.152	.157	+.588	+.173	.023
Model Summary	R=.629, R^2=.396, Adj. R^2 = .294			R=.613, R^2=.376, Adj. R^2 = .354		
	$F_{13, 77}$ = 3.887, p=.000			$F_{5, 142}$ = 17.118, p=.000		

There are several differences between this predictive model and the one for spouse. While the average of the neutral and embarrassing scenarios (contextualized for the viewer) remained the strongest unique predictor, measures related to the sensitivity of browsing activity were more dominant when the supervisor was considered to be the viewer. The viewing of erotica and also entertainment (a non-work related activity) both contributed negatively to participants' reported comfort level when reflecting on their supervisor viewing their recent usual browsing activities. Furthermore, the inclusion of s_diff, a measure that captures the differences in the sensitivity of browsing activities between home and away, as a positive factor indicates that those who partition their browsing so that more sensitive activities are conducted at home are more comfortable with their supervisor viewing their web browsing traces. Also, the majority location of use remained as a unique contribution to the equation for the first time. As the majority location of use (away) contributed positively to u_avg_sv, it appears as though those doing the majority of their browsing at a location

away from home may either be acclimatized to this viewing or may be conducting less sensitive browsing overall. It is also interesting to note that whether or a not a participant was viewer concerned no longer remained a unique factor in the equation.

These differences in the predictive models for supervisor and spouse/significant other highlight the highly contextualized nature of privacy concerns. The degree to which different dispositional and situational variables impact privacy in a given situation varies. As was initially discussion in section 5.5, not only were privacy comfort levels for spouse/significant higher than for the other viewer categories, they were also less variable. This is reflected by the simpler model for spouse which primarily incorporates variables related to inherent privacy concerns (average level of privacy comfort, magnitude of change in privacy comfort level by viewer) with only a small variation attributable to the sensitivity of browsing activities (i.e. viewing of erotica). The model for supervisor was more complex, incorporating both variables related to inherent privacy concerns, as well as location of browsing, and included much more variation attributable to the sensitivity of browsing activities. The complexity of this model is also shown by the reduced amount of variance the model predicts (48.0% for spouse, 35.4% for supervisor).

7.4.3 Summary

In order to instantiate a predictive model that could be used as the basis for an adaptive system, it is clear that we would need a greater number of participants so we could develop richer profiles for each sub group of participants. Some of the issues regarding poor fit of the model for participants that are *privacy unconcerned* and *privacy fundamentalists* may be reduced if we had sufficient participants to develop individual models for each sub group. As well, we require variables more attuned to the modeling process than those collected during our survey. The primary purpose of our survey was to investigate the general factors of incidental information privacy within web browsers. In an effort to keep the survey short, we used a limited number of scenarios and did not require participants to reflect on their privacy comfort level in the context of each device and location of use. Our initial attempts at predictive modeling in this section highlight the importance of gaining more contextualized information.

While preliminary, our results do show that such predictive models have potential for use in an adaptive privacy management system to provide the basis for filtering the traces of browsing activity appropriately. Given the differing browsing activities between home and away, it would be interesting to develop models for use in those different locations, as well as developing models for the different segments of inherent privacy concerns. However, given the amount of individual difference at play in this domain, it remains to be seen if a more formalized attempt at predictive modeling can give rise to models that account for a greater amount of the variability. Such models may be best used as a baseline for an adaptive privacy management. A user's interactions with the system could then be used to continually refine the model based upon their unique situations and concerns.

7.5 Summary

In this chapter, we presented several design guidelines for a privacy management system developed to help users maintain the visual privacy of their incidental information. We also examined the theoretical feasibility of two automated approaches to privacy management. We found that automatic content categorization shows promise as a mechanism to classify traces of browsing activity with an appropriate privacy level. However, commercial classification mechanisms are not sufficiently developed to allow for real time classification of visited web pages and accuracy would need to be increased through the use of additional heuristics. Similarly, our initial attempts at developing a predictive user model show that we would need further study with more focused questions and a larger numbers of users in order to be able to develop more nuanced and contextualized models through regression.

As part of this dissertation research, we wanted to instantiate the guidelines derived from our exploratory studies by developing a proof of concept privacy management system. Given that an automated approach to classification and filtering is not currently feasible, we leave further implementation and evaluation of an intelligent system approach to future work. We elected instead to pursue a more manual approach to privacy management. As will be presented in Chapter 8, we developed a proof of concept browser application that allows a user to open up (and toggle between) browser windows of different privacy modes. These windows not only automatically tag visited pages with the privacy mode of the window, they also filter which traces of activity are shown.

Chapter 8
Proof of Concept: PrivateBits

This chapter presents the design, implementation, and evaluation of PrivateBits, an instantiation of a browser window based visual privacy management approach. As presented in Chapter 7, our exploratory research identified design requirements and proposed an approach for semi-automatically classifying the privacy of traces of browsing activity. This approach leverages browser-window based temporal patterns in the application of privacy levels during web browsing. With this approach, the onus remains with the user to manage the classification of their browsing with system support.

8.1 Design and Implementation

PrivateBits was developed in C# and utilizes an IE browser control object to handle the core web browser functionality.

PrivateBits allows users to open concurrent browser windows with different privacy modes and allows them to change the privacy mode of any window when the sensitivity of the browsing changes. Windows in the PrivateBits browser filter previous activity for the current viewing situation and enable automatic tagging of visited pages with the current privacy level.

Figure 47 shows four PrivateBits Browser windows opened concurrently in three different privacy modes. The mode can be one of three hierarchical privacy levels: public, semi-public, or private. In addition, users can toggle between recording and not recording their browsing activity at any time. These privacy levels were found to be at an appropriate granularity during our two exploratory field studies. This provides a more nuanced approach than partitioned public/private modes or the current save/don't save model in web browsers.

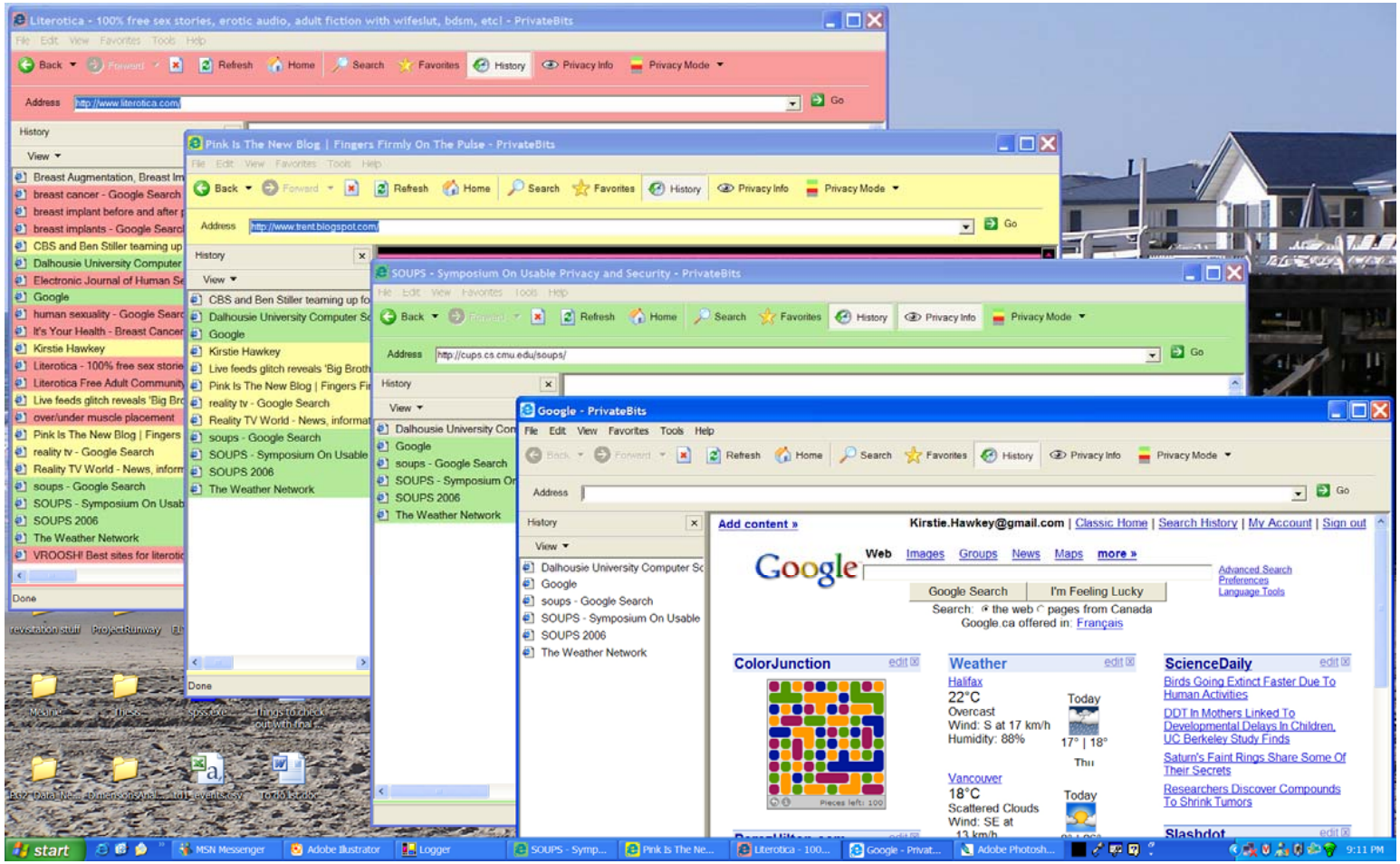


Figure 47. Screenshot of four PrivateBits browser windows, each with a different privacy mode: (from back to front) private (red), semi-public (yellow), and public (green) modes, as well as the public mode with no privacy feedback. Filtering of traces according to the privacy mode of the window can be seen within the History panel.

Traces of browsing (History, Bookmarks, Auto Completes) are automatically tagged with the window's current privacy mode. At any time, a user can easily change the privacy mode of a browser window through a simple menu option (Figure 48-a1) and any new activity that occurs in the window will be classified with the new privacy level. Additionally, users can indicate that they don't want any of their browsing activity recorded by deselecting the "record browsing" menu item accessible via the Privacy Mode button on the toolbar (Figure 48-a2)

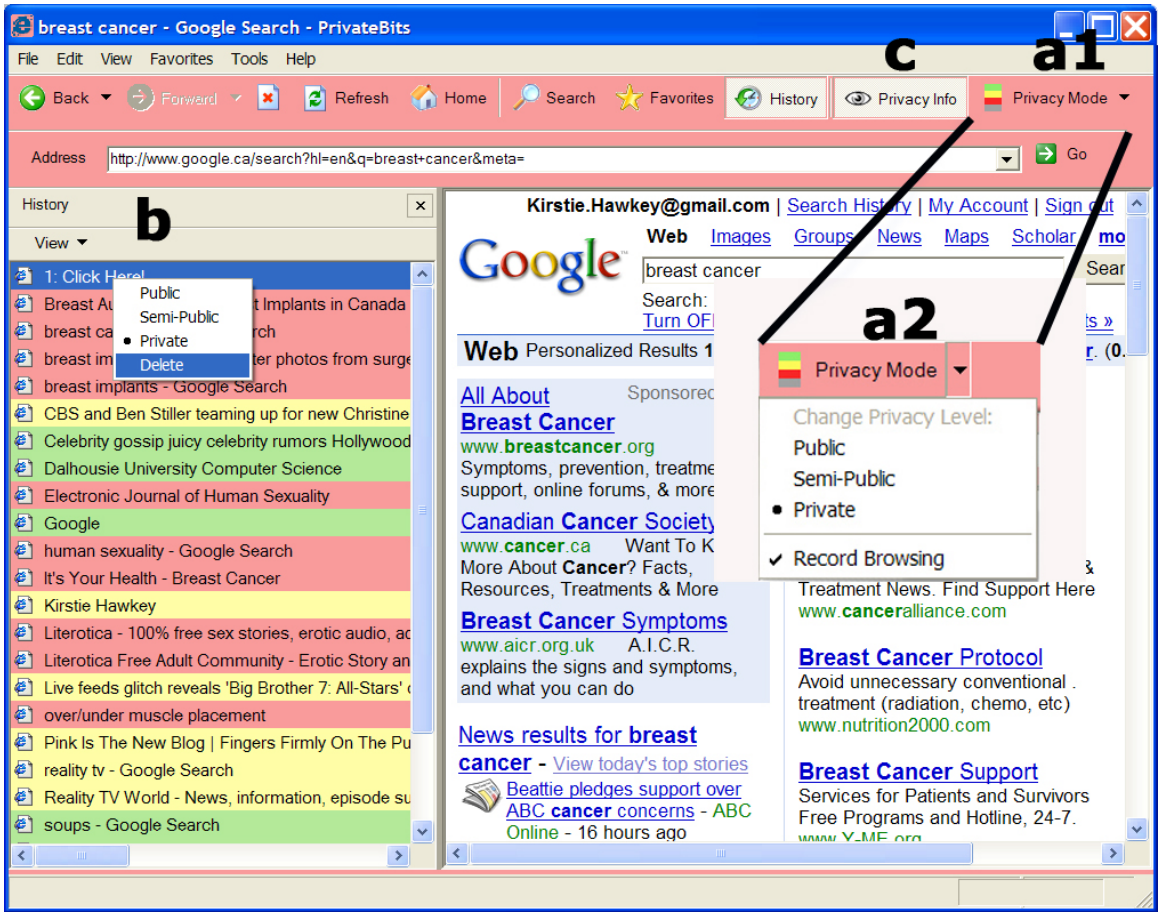


Figure 48. A PrivateBits browser window in private mode showing controls to a) change privacy mode (a2 shows the menu displayed when a1 is clicked), b) inspect and adjust the privacy level of previously classified items, and c) view/hide privacy information.

The privacy mode of a window can be made visible by clicking on the *Privacy Info* button (Figure 48-c). This button toggles between viewing and concealing visual feedback through the use of colour. The feedback indicates the current privacy level of the browser window and of previously generated traces of activity. Colours were selected using a traffic light analogy: green for public (safe), yellow for semi-public (caution), and red for private

(danger). When the visual feedback is viewed, the background colour of the toolbar panel is changed as well as the window icon on the window and in the task bar. When the visual feedback is turned off, the window appears as a normal IE window with the addition of the Privacy Info and Privacy Mode buttons (as seen in the front window shown in Figure 47.

To check the accuracy of the classified items, users can open the History or Favorites panel with visual feedback enabled. Items can also be sorted by privacy level within the panel to quickly show which items will appear in a given browser privacy mode. If any item is incorrectly categorized, users can manually correct the privacy level by right-clicking on the item and selecting a correct privacy level from the context menu; the entry can be also be deleted (Figure 48-b). Currently, modifications are done on a per-item basis; however, the ability to select multiple items will be provided in the next iteration of the browser.

To ensure that only contextually appropriate content is displayed, users simply set the privacy mode of the window according to their privacy comfort level in a given situation. In a public window, only items classified as public are visible; in a semi-public window, items classified as public *and* semi-public are visible; and in a private window, all recorded items are visible. This filtering can be seen in the History panels of the browser windows visible in Figure 47. PrivateBits currently filters History, Favorites, and Auto Complete entries (address bar and Google toolbar and web page search terms). The ability to filter the back and forward history lists and other form entries is planned for the next iteration of the browser.

8.2 Fulfillment of Design Guidelines

PrivateBits was designed to fulfill the previously described design guidelines (section 7.1). As our primary focus was on privacy management within the context of the existing web browser convenience features in IE, we omit those guidelines from section 7.1.2 that were more applicable to redesigning the features themselves (i.e., 7.1.2.2 and 7.1.2.3). Table 38 provides a summary of how PrivateBits fulfills each of our proposed design guidelines for visual privacy management.

Table 38. Summary of design guidelines, illustrating if and how PrivateBits fulfills each guideline

			Fulfilled??	Details
General Guidelines for Privacy Management Systems				
7.1.1.1 7.1.2.1	Increase visualization of settings and action		Yes	Inspection and modification of traces through History panel and visual feedback mechanisms
7.1.1.2	Configuration within the context of action		Yes	Privacy management is explicit (no policies). Classification occurs as browsing is conducted and can be inspected at time of browsing or before a viewing instance
7.1.1.3	Provide opportunities for varying granularities of privacy control		Yes	Privacy classification at granularity of task in the window, but each individual trace can be re-classified if need be. Privacy filtering is coarse grained.
7.1.1.4	Work within existing behaviours		Yes	Leverages existing web browsing patterns of activity. Flexibility in use supports varying web browsing behaviours and PIM management styles.
Guidelines for Visual Privacy Management within Web Browsers				
7.1.2.4	Reduce clutter within convenience features	Provide more usable configuration mechanisms	No	Works within existing feature functionality.
		Provide a more selective approach to deletion	Yes	Selective deletion at any time.
		Provide fine-grained control at the time of browsing activity	Yes	"don't record" mode
7.1.2.5	Allow Nuanced Privacy Classification		Yes	Public, Semi-public, Private, "Don't Record modes
7.1.2.6	Support Multi-tasking		Yes	Can have windows with different privacy modes
7.1.2.7	Support varying privacy concerns		Yes	Flexibility in use supports varying privacy concerns
7.1.2.8	Reduce the burden of privacy management		Yes	Users consider privacy at the level of task being conducted in the window. PrivateBits automatically tags all browsing activity with the privacy mode of the window.

PrivateBits helps reduce clutter in convenience features by allowing non-recording of traces at the time of browsing and easy deletion after the fact. PrivateBits provides a nuanced approach of three privacy levels for those traces that are saved. Users are supported when multi-tasking with varying content sensitivity in concurrent browser windows. PrivateBits is flexible enough to support varying privacy concerns, web browsing strategies, and personal information management styles. Users can interpret and use the privacy modes as best fits their circumstances of viewing. Users can opt to manage privacy at the time of browsing, or adjust privacy levels at a later time according to their primary usage contexts and personal information management styles. Finally, PrivateBits reduces the burden of classification by allowing users to consider privacy at the granularity of the task they are conducting in a browser window, rather than forcing individual classification of traces. However, fine-grained control at the item level is available for when it is required.

PrivateBits was also designed to address many of the requirements others have suggested for the design of privacy management systems. While PrivateBits does not create privacy policies (as in [90]), we have made it easy for users to inspect and modify the privacy classification of traces through the use of the History panel and visual feedback mechanisms. Furthermore, classification is applied proactively as traces are generated. As suggested in [39], PrivateBits provides visualization mechanisms to help users understand the current browser mode and to identify which traces will be visible in a given browser mode.

We have also integrated configuration with action by making the privacy classification explicit rather than having users create policies for classification. As suggested in [91], our approach highlights rather than obscures potential information flow though the inspection of privacy levels and emphasizes action over configuration. We also provide opportunities for both fine-grained classification of traces and coarse-grained control of what may subsequently be revealed (i.e. through non-persona based filtering). Indeed, the browser window privacy modes are conceptually similar to the precision dial which Lederer et al. [91] proposed as a method of bypassing the pitfall of relying on prior system configuration. In their case, they speculated that rather than trying to predict which level of privacy is desired for a potential situation, users could react to the situation by adjusting the level of granularity of information to be released on the dial. The position of the dial would serve as a feedback mechanism so that users could quickly observe the privacy setting. With

PrivateBits, the user can select one of three privacy modes for the browser (i.e. public, semi-public or private) which adjust the sensitivity of the traces displayed. The visual feedback of the privacy mode allows quick observation of the current privacy setting.

Furthermore, we designed PrivateBits to leverage existing web browsing patterns of activity to semi-automatically classify traces of web browsing activity as they occur. We expect that this approach will allow users to more easily accomplish the secondary task of privacy management as they conduct their primary task of web browsing.

8.3 Evaluation Study

One of the goals of this project was to evaluate our design guidelines by examining the effectiveness of PrivateBits at helping users manage the privacy of their web traces. We wanted to gather rich, qualitative data to determine whether the design and functionality of PrivateBits was appropriate for the privacy needs of participants before developing a more robust version suitable for evaluation in the field.

Laboratory studies allow researchers to observe participants in a controlled fashion. However, in privacy and security research it is particularly challenging to provide a realistic environment due to the highly personal nature of the data at stake. Participants may not be motivated to make the same effort and take the same actions in a lab study as they would if the data was their own [128, 152]. In order to address this concern, browsing scenarios used in this study were based upon actual viewing contexts identified by participants in the IIP survey during our exploratory research phase. Additionally, an online survey was administered prior to the PrivateBits evaluation session. Responses from the survey were used to personalize browsing and viewing scenarios in order to increase realism [134]. Appendix D includes the questions from the online survey as well as the scenario selection worksheet used to guide personalization of the scenarios.

Participants were asked to perform a series of personalized browsing scenarios while using PrivateBits to manage their privacy. Following this, semi-structured interviews were conducted with the participants to investigate the usability and utility of PrivateBits. Semi-structured interviews were chosen as we felt participants would be more likely to give rich information about their interaction experience verbally than if they were required to respond

in written form. We also wanted to be able to interactively probe events of interest that were observed during participants' interactions with the system.

8.3.1 Participants

Ten Internet Explorer users from the general Dalhousie university community participated in this study (see Table **39** for participant demographics). Participants were screened prior to inclusion in the study to ensure that they had regular occasions where others could view traces of previous activities on their display and that they had privacy concerns related to this viewing. Both office staff and students were recruited to help determine whether the proof of concept application worked well across a variety of usage contexts. Five participants were recruited from each group and each participant was given an honorarium of $15 for taking part in the study.

Table 39. Participant demographics and web browser usage.

	Overall	**Staff**	**Students**
N (male/female)	10 (6/4)	5 (2/3)	5 (3/2)
Average age	31	34	28
Browser Use (hrs/wk)	15-21	15-21	15-21
Technical/Non-Technical	5/5	2/3	3/2
Primary Device (laptop/PC/shared PC)	5/4/1	2/3/0	3/1/1
Primary Location of Use (home/away)	2/8	1/4	1/4
Avg. # devices in use (total: home / away)	3.3: 1.6/1.7	3.6: 1.6/2.0	3.0: 1.6/1.4
Purpose of browsing (% personal/% other)	35/65	35/65	36/64

8.3.2 Procedure

The PrivateBits evaluation session was held in an office environment at the University and lasted approximately one hour. After giving informed consent, participants were given a brief description of the visual privacy problem in web browsers and were introduced to the four levels of privacy that PrivateBits supports. Participants were then led through a demonstration of the functionality of PrivateBits and given a chance to explore its features. Once comfortable with the browser, a brief practice session consisting of a single browsing task (search for bankruptcy support group for a friend) and a single viewing

scenario (using one of their regular viewers) was completed. Appendix D includes the researcher script, participant tutorial, and practice scenarios.

Upon completion of the practice session, participants were asked to complete six personalized scenario-based browsing tasks over a 20 minute period (see Appendix D). The tasks were designed to generate traces of browsing activity across a variety of privacy sensitivities. Two tasks were designed to be fairly *private* in nature: 1) a search for information for a friend recently diagnosed with testicular or breast cancer; and 2) a search for information about reproduction, specifically the moment of conception, which was motivated by a neighbour's child needing resources for a class project. Two tasks were designed to be more *contextually private* relating to people that were regular viewers of the participant's display: 3) a search for a gift for a potential viewer, and 4) a Google search to try and determine the volunteer activities of a potential viewer. The remaining two tasks were designed to be more *neutral*: 5) a search for Madonna trivia sites in preparation for a radio station contest, and 6) a search for the most current information about a controversial political topic.

A set order was used to introduce the browsing tasks which were given both verbally and on paper. This order was intended to mimic the spontaneity of natural browsing and to provide an opportunity where multiple browser windows of varying sensitivities might be of benefit. Initially two tasks were given (cancer, local politics), after 6 minutes a third task was introduced (gift search), followed by two further tasks at the ten minute mark (Google search, Madonna trivia sites), and the final task at the sixteen minute mark (reproduction information). Participants were encouraged to use PrivateBits to manage their privacy and asked to locate and bookmark 3-4 sites for each task.

Once the browsing tasks were complete, participants were given an opportunity to inspect the traces saved in the history and adjust privacy levels as desired. This gave them the chance to take privacy preserving actions mimicking the actions participants reported taking during our exploratory studies. Then, through a series of four personalized viewing scenarios, participants were asked to evaluate how well PrivateBits was able to filter their traces of browsing activity. The viewing scenarios were personalized with names of the participants' most regular viewers, as indicated during the pre-session survey (see Appendix D). These viewers included a person that they were very comfortable with, one they were

not very comfortable with, and one with a neutral comfort level. A fourth viewer was chosen from each participant's regular viewers to provide breadth for their viewing scenarios in terms of context (personal/work/school) and equality of relationship (peer/superior/subordinate).

We also obtained user feedback about the effectiveness of PrivateBits in helping users manage their privacy. Our intent was to use this feedback to refine the interface and functionality of PrivateBits so that a future version could be deployed in a field study. Semi-structured interviews were used to enable us to gather ratings of efficacy and usability of the interface, probe for the reasons behind the ratings, and discuss opportunities for improvements. A discussion guide was used as part of the evaluation to help decrease researcher bias and maintain consistency and reliability across the evaluations (see Appendix D).

8.3.3 Data Collection

PrivateBits was implemented within an experimental framework for the purposes of user testing. Logs were created to record browser events as participants interacted with the system including button presses, web page classifications, and textual entry. We also logged which traces could be visible whenever the privacy mode changed or a browser window closed and tracked which auto complete terms were displayed with each key press of text entry. This allowed us to closely examine participants' experience with PrivateBits. Separate log files were created for each phase of the session including the system demonstration, practice session, and viewing session.

In addition to log files and observations, interview notes were made by the researcher and augmented the audio recording of the session. The audio was transcribed for analysis. The pre-session survey and the script for the semi-structured interview questions can be found in Appendix D. Participants also completed the web page category and viewer classification tasks used during the field studies (Appendix C).

8.4 Evaluation Results

As we present the results, it's important to note that although the browsing scenarios were similar, each participant visited a distinct set of pages, employed different privacy

management strategies, had different privacy concerns for the visited pages, and had different potential viewers. We reflect on the effectiveness of the interface at meeting those varying needs throughout the results, using descriptive statistics to convey the range of usage observed

8.4.1 Privacy Management during Browsing Scenarios

Participants exhibited varying browsing and privacy management strategies (see Table 40 for details). Nine of the ten participants opted to manage the privacy of their visited pages while browsing, adjusting the mode of the browser to accommodate the sensitivity of different topics and pages; the other participant (P1) used the default browser mode (public), and adjusted the privacy level of visited pages after all the browsing was completed. Of those that managed their privacy while browsing, five participants (P3, P5, P6, P7, P10) chose to not use the public mode so that it would not contain any traces of browsing. Interestingly, all of these participants were office workers. Only two of the participants (P8, P10) elected to not record some of the visited pages at the time of browsing; but five others (P1, P2, P4, P5, P9) indicated that they would anticipate using this setting when browsing in their normal environment.

Table 40. Descriptive statistics of participants' activities during the browsing scenarios.

ID	Group	Tech Level	# pages visited	# unique pages visited	# windows opened		#Google searches	#privacy mode changes	% pages with visible privacy info	# pages adjusted
					Total	By user				
P1	Student	Technical	123	57	10	1	11	0	0	11
P2	Student	Non-tech.	91	52	10	9	9	15	52	0
P3	Staff	Technical	96	47	4	3	8	4	18	1
P4	Student	Non-tech.	108	62	3	1	24	8	98	1
P5	Staff	Non-tech.	72	42	3	1	16	5	86	6
P6	Staff	Non-tech.	51	34	6	5	15	4	0	1
P7	Staff	Non-tech.	42	23	2	2	7	3	95	1
P8	Student	Technical	96	55	3	1	20	5	99	0
P9	Student	Technical	82	38	9	6	8	3	0	1
P10	Staff	Technical	105	42	4	1	8	3	0	10
		Average	86.6	45.2	5.4	3.0	12.6	5.0	44.7	3.2
		Minimum	42	23	2	1	7	0	0	0
		Maximum	123	62	10	9	24	15	99	11

Participants had varying privacy concerns for the pages they visited during the browsing scenarios. In particular, we note that staff participants, on average, considered more of the browsing to be sensitive (4.8% public, 32.0% semi-public, 57.5% private, 3.8% don't record) than the student participants did (41.1% public, 30.5% semi-public, 20.1% private, 8.3% don't record). This was also reflected in our interviews as staff indicated a concern for non-work related browsing being visible to colleagues, employees, and supervisors while only one student mentioned similar concerns. None of the browsing tasks used in this study could be considered to be work-related with the exception of the medical and sex education searches for one student participant who was also a medical doctor. Indeed, this participant was the only one to not consider any of the browsing to be private.

8.4.2 Privacy Management during Viewing Scenarios

The viewing scenarios were customized for the participants in order to represent their most regular viewers. Therefore, participants had different types of viewers. During the 40 viewing scenarios (4 scenarios x 10 participants), participants opened 31 public windows, 7 semi-public windows, and 2 private windows. The breakdown of windows opened for each type of viewer is listed in Table 41. Most participants envisioned opening their browser in a mode that would restrict the amount of trace information visible (i.e. public or semi-public mode).

Table 41. Privacy mode of windows opened during viewing scenarios (by viewer type).

		Mode of Window Opened			
Viewer Type	**# Scenarios**	**Public**	**Semi-Public**	**Private**	**% Public**
Colleagues	9	8	1	0	88.9%
Spouse/Significant other	9	5	2	2	55.6%
Supervisors	8	8	0	0	100.0%
Close friends	4	3	1	0	75.0%
Parents	4	2	2	0	50.0%
Employees	3	2	1	0	66.7%
Tech Support	2	2	0	0	100.0%
Client	1	1	0	0	100.0%
Totals	40	31	7	2	79.5%

For each of the four viewing scenarios, participants were asked to select a privacy mode for the browser and then to open the history panel so they could see the traces that might be visible. They were then asked to reflect on their privacy comfort level (PCL) using a 7-point scale (1-extremely uncomfortable, 4-neutral, 7-extremely comfortable) if the viewer

could see those traces. Participants were asked to reflect on their comfort both if they were in control of the keyboard and mouse (PrivateBits-You) and if they had left the room and the viewer was in control of the keyboard and mouse (PrivateBits-Away). If participants noted something was visible within the History that was inappropriately classified, they were asked to give their comfort level for the currently visible traces and also adjusted as if the traces had been classified as intended. Similarly, if assumptions were made about whether or not the data could be password protected, the participants were asked to give their comfort level without password protection and their adjusted comfort level if password protection was available. In the following analyses, we use the adjusted privacy comfort level values if applicable.

Figure 49 shows participants' ratings of their privacy comfort levels (PCLs) for each category of viewer. The graph contrasts participants' privacy comfort level values gathered from the pre-study survey with the privacy comfort level values obtained during viewing scenarios. It should be noted that during the pre-study survey, the comfort level was contextualized for a specific viewer, but there was no context as to the level of control retained or the sensitivity of the content that may be visible.

The viewer types in Figure 49 are positioned in ascending order by the percentage of public mode windows participants indicated they would open for that type of viewer (as

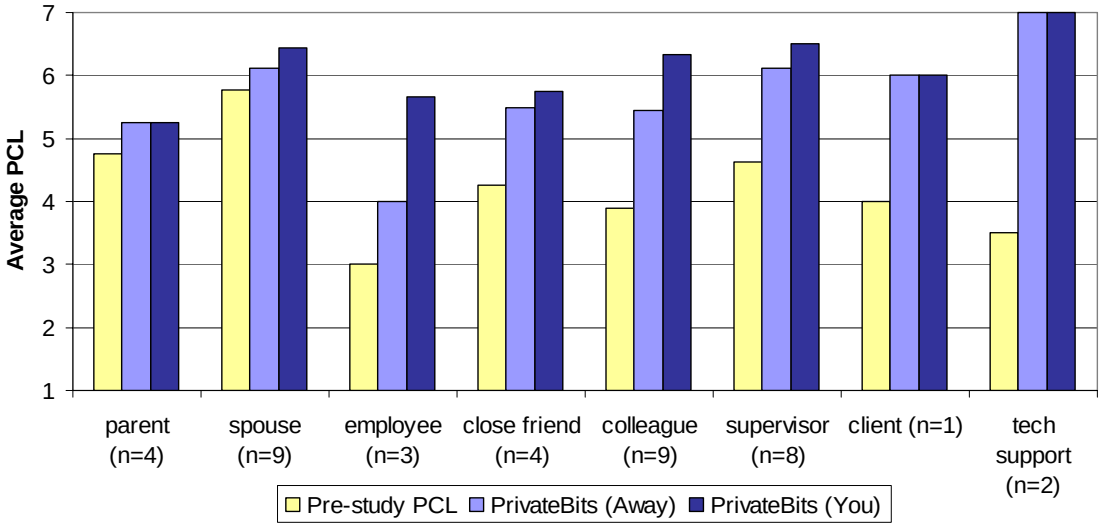


Figure 49. A comparison of participants' pre-study privacy comfort levels (PCLs) for each viewer type with their comfort for those same viewers when using PrivateBits. Viewer types are in ascending order by the percentage of public mode windows opened and magnitude of difference in PCL.

shown in Table 41) and by the magnitude of change between the pre-study privacy comfort level and that for PrivateBits-You. As can be seen, trusted viewers such as spouse and parent would be more often allowed to see sensitive traces in a semi-public or private mode window which resulted in a lower change in privacy comfort level. The greatest gains were seen in the work relationships where viewers were only permitted to see browsing classified as public. This is encouraging given that the greatest privacy concerns have been found for hierarchical relationships such as supervisor-employee (e.g., section 5.5, [93, 127]).

We also examined the privacy comfort levels by participant. Overall, participants had a high privacy comfort level when using PrivateBits (Figure 50). The average privacy comfort level across users rose from a mean of 4.5 (range 3.25 to 6.0, standard deviation 1.04) for the pre-study privacy comfort level, to a mean of 5.7 (range 2.75 to 7.0, standard deviation 1.46) for the scenario with the participant leaving the room (PrivateBits-Away), and a mean of 6.2 (range 3.75 to 7.0, standard deviation 1.05) for the scenario with the participant in control (PrivateBits-You). Eight of the ten participants (5 staff, 3 students) showed large increases in their comfort level. Interestingly, participant P1 was one of the two participants whose comfort level remained low. As mentioned previously, P1 opted to use the default public mode and made minimal adjustments to the content after browsing. Four of the participants (P1, P2, P3, P7) mentioned that they were fine with some viewers seeing potentially sensitive information (e.g., searches about conception, cancer) as long as the participant was available to give context to the viewed pages (i.e. that the browsing did not represent a personal

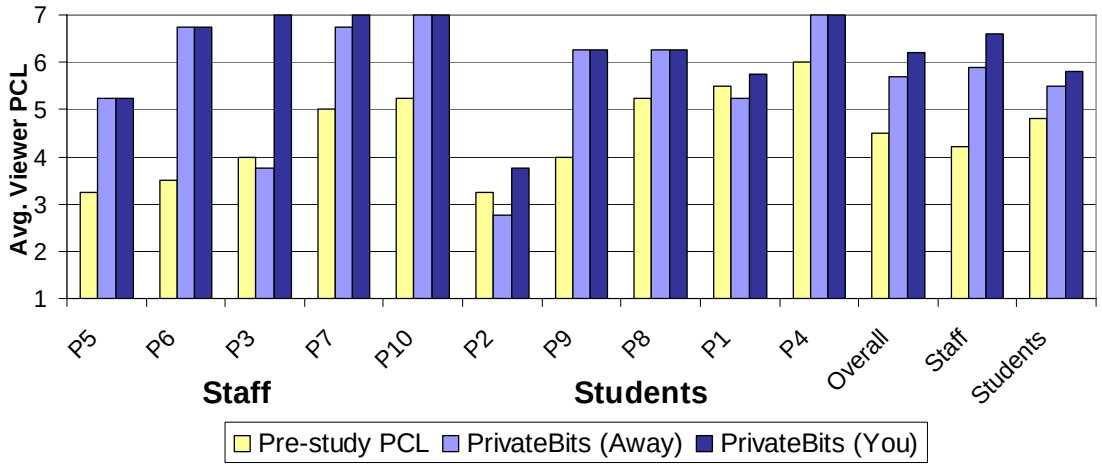


Figure 50. A comparison of each participant's pre-study privacy comfort levels (PCLs) across viewer types with their comfort for those same viewers when using PrivateBits. Participants are ordered by whether they are staff (P5, P6, P3, P7, P10) or students (P2, P9, P8, P1, P4) and by their pre-study privacy comfort level.

concern, but was for a friend). Their comfort level decreased when considering the situation of the viewer being left alone in control of the keyboard and mouse.

As can be seen in Figure 50, staff participants tended to have larger increases in their privacy comfort levels than did the student participants. The average privacy comfort level across staff users rose from a mean of 4.2 for the pre-study privacy comfort level, to a mean of 5.9 for the scenario with the participant leaving the room (PrivateBits-Away), and a mean of 6.6 for the scenario with the participant in control (PrivateBits-You). In contrast, student participants rose from a mean of 4.8 for the pre-study PCL, to 5.5 for the PrivateBits-Away scenario, and 5.8 for the PrivateBits-You scenario.

Whether or not the privacy mode could be password protected also affected participants privacy comfort level when considering the situation of the viewer being left alone in control of the computer. Three participants (P5, P8, P9) asked if the history could be locked and were told that currently it could not. They then reported on their comfort level for each case (non-password protected and password protected). In all cases, when assuming password protection was enabled, their comfort level increased to the same level as if they remained in control.

The contextually sensitive scenarios (gift search, Google-ing a viewer) emphasized that the privacy management system must be flexible enough to enable users to adapt to unforeseen circumstances. Issues arose as participants determined an appropriate browser privacy level for normally trusted viewers when there were page visits that were contextually sensitive for that person. For example, the gift buying scenario generally had the participants' spouse or significant other as the recipient of the present. While this person might normally be trusted to view pages classified as private, the existence of this secret activity needed to be hidden. The same shopping activity would be considered suitable for others to see, even if they were normally less trusted. There was also concern that if the normally trusted person realized they were restricted in their viewing, questions may arise as to what was being concealed.

8.4.3 Suitability of Privacy Levels

After completing the viewing scenarios, participants were asked to reflect on how well the four privacy levels (public, semi-public, private, don't record) fit the web pages they

had visited. Five of the participants reported that the levels fit all of the time, while the remaining five reported that they fit most of the time. Reasons for pages not fitting neatly under any one of the classifications included that it depended on the person (3/5), the location (4/5), or the context of a search (1/5).

All participants thought that the terminology used for the privacy levels was appropriate; however, 9/10 referred to semi-public as semi-private on occasion. When participants were questioned about this discrepancy in terminology, the consensus was that either term would be acceptable.

Most participants (7/10) thought that the 4-level hierarchy would be suitable in a work/school environment; two participants did not think that they would need the semi-public level; the remaining participant felt that five levels would be more appropriate (public, semi-public, private, "just me locked private", don't record). Participants indicated the same preferences for levels at home as when away, although four stated they may use them differently.

8.4.4 Usability of the Interface

Participants were questioned as to the usability of various interface elements in the PrivateBits browser. Almost all participants (9/10) found it easy to change between modes using the button's drop down menu; one participant would have preferred small buttons (one for each level). Half of the participants thought that it would also be nice to have short-cut keys enabled. Most (7/10) also found it easy to switch between recording and not recording the browsing (checking the menu option). The remaining three felt it might be more appropriate to have a separate toggle button for record/don't record, with one participant wondering if fully separating out the two functionalities of the privacy mode (filtering and classifying) would be best. All participants felt that it would be important to be able to password protect the system so that the privacy mode could not be changed by someone left alone at the computer.

8.4.4.1 Privacy Level Feedback Mechanisms

We asked participants questions about the privacy feedback information (i.e., the colour coding of the browser window, icon, and items in the History). The toggle button was felt to be easy to use by 9/10 participants, while one participant did not make the

connection between the button and the colour feedback. All participants found the colour coded items in the History and Favorites panel useful when determining if appropriate privacy levels had been set. While only 6/10 actually made use of the feedback during the browsing session, 8/10 thought it would be useful to confirm the privacy mode of the browser if switching between modes. There was some concern, however, about the visibility of the privacy mode to others that may be able to view their screen. Two participants mentioned that it might draw attention to an activity they were trying to hide; the red colour for the private mode was felt to be particularly eye catching. One participant, however, thought that it would not matter -- when others could possibly see the screen, he would not be engaging in inappropriate activities.

Whether or not participants felt a more subtle form of feedback was desirable appeared to depend in part upon participants' normal browsing situations. When browsing in a more public environment, as in an open office plan, six participants felt that more subtle feedback (i.e. something not visible from across the room) would be important. Suggestions for more subtle feedback mechanisms included having just the coloured icon in the task bar, a coloured address bar, a traffic-light icon on the tool bar that would indicate the current privacy mode, text saying the current mode, and self-selected colours.

During collaboration, it was generally felt that having no feedback at all would be appropriate, with 7/10 participants wanting the option to conceal all indications that a privacy management system was in use. Two participants did indicate that if such a privacy management systems was to become commonplace, that they would feel less of a need to conceal its use. When asked about the reverse situation, if they were the viewer and could see that somebody was using a privacy management system and may be hiding some activities from them, participants were split in their response. Three participants would be very curious or suspicious about what was being hidden, three would not care at all, and the remaining four would care more or less in different situations. Three participants felt that as an employer, they should be able to see what their employee was hiding, but would be fine with a colleague using such a system. Again, two participants mentioned that social norms would be a factor in their comfort with such a system.

8.4.4.2 Privacy Modes

One design choice that we made when developing PrivateBits was based on our opinion that people would be unlikely to proactively switch privacy modes. We assumed they would be more likely to recognize that a visited page was changed in sensitivity and then would want to retroactively switch modes. We therefore opted to have the privacy mode of the current page change when the mode of the browser changed. This design choice was not popular with participants; 9/10 would have preferred for the page to have remained at the previous level, with the remaining participant being undecided, seeing the benefits of both approaches. Participants had several suggestions to help manage a privacy mode change including providing a button to change the privacy level of the last page, to only change the privacy level if the page was refreshed, to only change upwards in privacy level automatically (public to private, never private to public), and to start fresh with a blank page once the mode was changed. The option for a blank page was mentioned by three participants who felt that starting fresh would be appropriate particularly if the mode was changed because somebody entered the room.

Another design choice we made was for the default new browser mode to be public with no privacy feedback visible. This was thought to be a suitable mode if a browser window was opened in the presence of others. Eight of the ten participants felt this was an appropriate choice. However, when another candidate default mode was described (opening the window in semi-public mode so that the public mode wouldn't be inadvertently populated with inappropriate activity), half of the participants thought that it too might be a valid option. The overall perception was that the decision of which default browser mode was best is situational and depends in part on how frequently windows would be opened in the presence of others. Seven of the participants were asked if the default mode should be configurable and all agreed that would be best. One participant felt that when a window is first opened, rather than have a default privacy level, there should be an option to select the appropriate mode.

Other suggestions for improvements to PrivateBits focused on more automated privacy support. Suggestions included alert messages if the system detected private browsing in a public window, triggers for switching privacy modes such as search terms or secure sites, and automatic deletion of traces resulting from pop-up windows.

8.4.4.3 Willingness to Adopt the Technology

Nine of the ten participants thought that, if fully developed with the features that had been discussed, they would try using PrivateBits to manage their privacy within the web browser. The remaining participant thought that it would be well suited for his wife's privacy needs. Two participants did mention that it would have to be a plug-in as they would not be willing to install a new browser.

8.5 Discussion of Results

8.5.1 In the Viewer and System We Trust

An individual's relationship to the viewer has previously been found to be a contributing factor to their privacy comfort level during viewing of web browsing traces (section 5.5); the user's trust in the viewer is an important component of that relationship (see [104] for an overview of trust sub-components). Paine et al. [117] examined the impact of trust on online disclosure of information and concluded that an increase in perceived privacy will not result in an increased disclosure of information unless there is trust in the underlying system.

During our interviews, the role of trust was frequently mentioned as participants discussed their choice of privacy modes for the various viewers and their subsequent comfort level. The choice of an appropriate privacy mode depended in part on whether or not the viewer could be trusted to understand what the activities meant and to not broadcast them to others. For those participants that did not envision the system with password protection, their privacy comfort level for the scenario where the viewer would be left alone at the computer depended upon their trust that the viewer would stay on task and not change the privacy mode of the browser. Furthermore, plausible deniability is an established practice by which people maintain privacy [91]; in their absence, participants were concerned the viewer may assume the browsing was personally motivated. Results from our IIP survey found that loss of control over input devices reduced privacy comfort in a given situation for many. Providing security is clearly important as a means of ensuring privacy and may lessen the impact of loss of control by ensuring that traces are only viewed when there is the opportunity for users to give a plausible explanation for the activity.

Similar to the institution based-trust defined by McKnight et al. [104], we also need to consider users' trust in the system. We noted more distrust from technical participants who are aware of flaws inherent in any application than from non-technical participants who seemed more willing to take the effectiveness of the system at face value. Furthermore the technical users were more aware of all the places that traces of activities may be found (e.g., cache), with one participant wanting to double check what information was visible outside of the browser before indicating his comfort level for the scenario where the viewer would be left alone at the computer.

Trust in the system is also related to willingness to let the system be proactive. The majority of our participants expressed a preference for explicitly having to change the privacy level of a page when changing the mode of the browser, rather than having the browser re-classify the page. This reluctance for the system to handle privacy decisions is consistent with results reported by Ackerman et al. [7]; their respondents indicated a reluctance for automated data transfer to websites, preferring explicit approval of the transfer. However, it is also important to note that other suggestions for improvements to PrivateBits mentioned by participants focused on more automated privacy support. Clearly automation can play a role if appropriate, particularly for identifying potentially misclassified traces.

Trust, or confidence, in the system to guard the privacy of sensitive information is necessary for adoption of a privacy management system. Camp et al. conclude (in a study of users' willingness to release personal information to web sites) that "systems designed to offer security and privacy, and thus indicating both benevolence and competence, are more likely to be accepted by users"; however, "failures in such systems are less likely to be tolerated" [23]. It will be important to implement security features such as password protection of privacy modes and encrypted system files for the privacy tagged traces to increase users' trust in the system.

8.5.2 Privacy for the Privacy System

One tenet of user interface design is to increase ease of use through visibility of options and system status. Our evaluation of PrivateBits showed, for privacy interfaces that will be used in the presence of others, there may be a conflicting need for discretion. When

people do not expect others to respect their privacy or if the value of the private space is so high that they dare not risk it being revealed, people may resort to deceit and secrecy to protect their privacy [40]. If a person can view that we are keeping secrets from them, it can undermine our relationship with them; selective sharing can be hurtful to those that are excluded [40].

Designers of privacy enhanced systems are advised to use feedback mechanisms to afford users with understanding of the system's actions and state and control mechanisms to provide users with methods of taking appropriate privacy preserving actions [15, 91]. Dourish et al. [42] discuss the usability of security systems. They describe how security systems typically act as a barrier to action (e.g., authentication mechanisms interrupt the primary task), while usability professionals try to remove barriers of use. They conclude that security enhancing technology must be highly visible and available seamlessly to the user as they conduct their primary task. Otherwise, the user may be unable to recognize and understand the security implications of their system configuration and use.

Participants expressed a desire to be able to conceal the coloured privacy feedback indicating the current privacy mode of the web browser window and the privacy level of individual traces as well as the existence of the privacy system itself (e.g., the buttons on the toolbar). Several different strategies for maintaining the tradeoff between ease of use (visibility) and privacy (concealment) were discussed. Which strategy is most appropriate depends on the situation of use and includes the frequency of viewers, the casual visibility of the display, the sensitivity of the information, and the social norms for the environment.

Whether or not participants felt they would like to conceal the privacy management system depended in part on whether such privacy systems became common and were an accepted activity. The legitimization of such a system depends on cultural and organizational acceptance of the rights of people to neutralize the unintentional surveillance of their web browsing activities (see [101] for a discussion of the social issues). Certainly, within the workplace, there may be well-defined limits on what is appropriate as well as a culture that dissuades personal activities. This was evident for those participants who remarked that, as an employer, their answers may change.

The ability to make the privacy management functions invisibly accessible, with quick access and concealment as the situation dictates may impact adoption of a final system.

Privacy has been found to be valued over convenience [8], so some effort to maintain it may be acceptable for users as long as the benefits outweigh the effort. However, a person's desire to engage in privacy preserving activities has also been found to be moderated by personal behaviours such as immediate gratification and self-control [9].

8.5.3 Incorporating Flexibility

The hierarchy of the four privacy levels (public, semi-public, private, don't record) within PrivateBits was found by participants to be flexible enough to deal with their individual privacy concerns and browsing strategies. Several opportunities for personalization were suggested including the default settings for the initial privacy mode of a new browser and the mechanisms and default settings for visual feedback of the privacy mode. Given the individual nature of privacy concerns in this domain as well as the varying environments of use and browsing behaviours, incorporating as much flexibility into the interface as possible is one of our priorities.

8.5.4 Study Limitations

While this laboratory evaluation of PrivateBits was effective at getting initial feedback about our prototype, it was not without limitations. The small population, consisting of only two groups (students and university staff), is not a representative sample and does not encompass the privacy concerns and usage environments of all potential users.

The personalized browsing and viewing scenarios did provide participants with an opportunity to use PrivateBits and to reflect upon its effectiveness for several of their regular viewers. However, the web browsing conducted did not capture all levels of privacy sensitivity. For example, the participant who marked all the browsing as either public or semi-public also thought the system should have a 'super-private' category. Presumably, he could envision browsing activities more sensitive than those undertaken during the study. Furthermore, all of the browsing was easily explainable as an activity undertaken for somebody else (e.g., a search for testicular cancer for a recently diagnosed friend). It will be important to test PrivateBits in a longitudinal evaluation in the field to ensure that it is flexible enough to accommodate a wider range of browsing and viewing scenarios over time.

8.6 Summary

PrivateBits, our proof of concept web browser, was developed as an instantiation of design guidelines for visual privacy management systems. The evaluation provides initial validation of these design requirements. Furthermore our technique using browser windows of varying privacy modes to classify and filter traces of web browsing activity was found effective by participants. However, further evaluation is required to validate this approach in a natural usage environment.

Our evaluation revealed participant concerns that are unique to privacy management systems. Concerns relating to trust in the system indicate that security mechanisms such as password protection and encrypted data must be provided. Furthermore, the need to conceal the existence of the privacy management system from others has raised questions about appropriate methods of managing the tradeoff between ease of use of the system and its privacy. Participants should be provided with several levels of visibility that they may choose between depending on their usage environment. As privacy management systems become more commonplace, the need for concealment may lessen.

Chapter 9
Suitability of Methodological Approach

In this chapter, we reflect upon the suitability of the methodological approaches taken during this research. We first discuss the suitability of the mixed methodological approach used for our exploratory studies of incidental information privacy concerns. We then reflect on the effectiveness of participant annotation of logged data as a method of studying rich natural behaviours in situ. Finally, we reflect upon the laboratory evaluation of PrivateBits.

9.1 Mixed Methodology Approach to Studying Privacy

The IIP survey and PG1 and PG2 field studies contributed to our understanding of privacy concerns in different ways. We next discuss the effectiveness of the different methodologies for examining the factors that impact privacy comfort, the browsing activities that generate the incidental information, and the feasibility of different privacy management approaches.

9.1.1 Examining the Factors that Impact Privacy Comfort

The survey was effective at examining the impact on participants' privacy comfort levels across three factors: the level of control they retained over input devices, their relationship to the viewer, and the general sensitivity of visible content. We were able to manipulate the context of each potential viewing scenario across the three factors and obtain privacy comfort responses from a large number of participants. In contrast, the field studies would not be suitable for examining the interplay of these factors given the varying content and contexts of viewing encountered during the week.

While we did not initially intend to examine participants' inherent privacy concerns through the survey, we were able to use participants' responses to the scenario questions to infer them. We examined how an individual's inherent privacy concerns varied according to the overall level of their privacy comfort level and the magnitude of change in their privacy comfort level caused by the factors of content sensitivity, viewer, and level of control. We were able segment our participants as *privacy unconcerned*, *privacy pragmatists*, or *privacy*

fundamentalists and also to sub-divide pragmatists depending on their amount of concern along the factors of control, viewer, and scenario and also on their overall level of concern. While some participants were concerned across all contexts, others had concerns along only one or two factors. Being able to segment users according to their inherent privacy concerns may let interface designers offer a simplified configuration mechanism that only presents those aspects with which a user is concerned.

One aspect of the survey that was flawed was the use of a seven point scale for privacy comfort levels. Given our subsequent analysis classifying users as being privacy concerned or unconcerned, a six point scale may have been more appropriate. This would have more clearly divided responses for each situation into either comfortable or uncomfortable. The neutral level (4 on the 7 point scale) is less meaningful as it allowed participants to indicate they were neither comfortable nor uncomfortable.

In retrospect, another aspect of the survey that was less than ideal was our choice of embarrassing, neutral, and positive to elicit the end-points and mid-point of the range of privacy comfort participants might have for traces of prior browsing activity. There was no difference in privacy comfort levels between the neutral and positive scenarios; both these scenarios were found to have a high level of comfort. Rather than attempting to craft neutral and positive scenarios, a better choice may have been moderately uncomfortable (i.e., google-ing a colleague) and comfortable (i.e. either of the neutral or positive scenarios we had used). The embarrassing scenario was found to provoke discomfort; although as discussed in section 5.4.1, it was not the lower endpoint of the discomfort scale for those participants that had a lower privacy comfort level when reflecting on their usual browsing. While the ultimate decision that a user must make in a given viewing scenario is whether traces of activity are appropriate or inappropriate for viewing, that decision is based on a judgment that includes additional factors such as the viewer and the social norms of the environment.

9.1.2 Examining In Situ Browsing Activities

The survey asked participants to reflect upon their "usual web browsing" for one scenario. While this gave us a general indication of how sensitive they feel their usual web activities are in relation to the other scenarios given, we do not know how sensitive the

specific content was. Is a low privacy comfort level the result of a single activity that is very sensitive or of several activities of a lesser sensitivity? The field studies allowed us to examine participants' perceptions of the privacy sensitivity of each page they visited.

The field studies confirmed our hypothesis that privacy concerns are fine-grained. Commercial web browser privacy tools tend to assume that most browsing is public with a small set of very private browsing (e.g., pornography). However, most (31/35) of the participants in our two field studies used all four privacy levels when classifying the privacy of their visited pages and all participants used a combination of public (i.e. suitable for anybody to see), semi-public (i.e. it depends on viewing context), and private (i.e. suitable for a close confident, or possible nobody else to see) classifications.

The survey allowed us to gather self-reported data from a large number of participants about the general types of browsing activities and the location (home, work/school) and type of computer used (desktop, laptop). While this gave us some indication as to how activities change according to location of browsing and device, the data was not specific enough to evaluate the feasibility of various methodological approaches. It is not only import to know "what content may be visible", but also to know what patterns of activity are occurring as the content is generated. For instance, commercial privacy products tend to assume that private browsing is not intermingled with other browsing (e.g., allow users to either be browsing within a "private browser" that requires password access or in their normal browser, but not both concurrently). Results from our field studies showed that while private browsing may sometimes be kept to a single window, participants generally also had other windows open and moved between the open windows.

9.1.3 Examining Feasibility of Privacy Management Systems

In addition to details about privacy levels for individual pages, one of the important contributions of the field studies was to help us understand the feasibility of different privacy management approaches. One challenge that needed to be overcome as we developed a privacy management system was the volume of visited pages and the speed at which browsing could occur. It became apparent that any management solution that required a per-page annotation of a privacy level would be overly burdensome for users. A

semi-automated or automated approach would be required to make the cost of managing this privacy more acceptable for users.

Using data from the PG2 field study, we examined the relationship between the content of the pages visited and the privacy levels applied. This allowed us to theoretically evaluate the feasibility of using automatic content categorization as a method of classifying visited pages without having to develop a prototype system. Capturing data through the field studies also allowed us to examine natural patterns in the application of privacy levels such as streaks at a given privacy level and the number of transitions between levels on a per-window basis. The feasibility of leveraging this approach to reduce the burden of classifying traces of web activity with a privacy level was examined through the proof of concept web browser, PrivateBits.

9.1.4 Summary

Privacy research is challenging, but is necessary in order to build usable tools for privacy management. Unless the various factors of privacy in a given domain are explored, including patterns of actual activity, it can be difficult to build a privacy management system that is not only effective, but also at a low cost for users in terms of configuration and on-going privacy management. A mixed methodology approach can help ground the research appropriately. Surveys can examine higher level attitudes and self-reported behaviours from many participants, while field studies can obtain the fine-grained details necessary to be able to evaluate the feasibility of potential privacy management approaches based on participants' everyday interactions.

9.2 Participant Annotation of Logged Data

When requiring participants to annotate their behaviour, there are a number of factors that impact the quality of the data, including the categorization schema provided to participants, the duration of the study, and the time of annotation (real-time vs. post hoc). In this section, we reflect on our experiences in dealing with these factors.

9.2.1 Categorization of Behaviour

Often when collecting contextual information in the field, participants are required to not only describe their actions or intent, but also to characterize their own behaviour

within a previously defined schema. In order for participants to do this, they must be trained so that they properly understand the categorization scheme. Furthermore, if participants must assign categories as was done in our studies, it is important that the categories are obvious and easily distinguishable.

Participants in our field studies were trained on the 4-level privacy gradient classification scheme through the use of a diagram as a discussion aid (Figure 4). Care was taken to not dictate what content would be considered at each of the levels. The emphasis was placed on whether or not participants would be comfortable with anybody seeing it (public), only themselves or a close confident seeing it (private) or something in between that may be suitable for some subset of viewers to see. The one content-based example given was of a job search as something that might be inappropriate for a boss to see but fine for a close friend (semi-public). Participants were provided with a reference handout to remind them of the classification scheme.

The willingness of the participants to carefully and thoughtfully annotate their data, as well as the required frequency of the annotations, must be considered when evaluating the accuracy of the annotations. Participant fatigue may cause accuracy to decline over time; however it may also improve as they become more skilled and comfortable with the categorization schema. There were some indications that participants were carefully annotating their data. In particular, one participant in the PG1 field study, who had forgotten that he could sort the data in the electronic diary, had painstakingly classified almost 50 pages which alternated between two privacy levels. These pages appear to have been a log-in page (public) and more private pages. During the PG2 field study, inspecting the URL and page title of classified pages revealed sequences that appeared to be reasonable in the privacy levels assigned.

9.2.2 Duration

In comparison to previous research that has collected web usage logs on the Web for extended periods of time (e.g., [31, 145]), there is a limit to how long participants will be willing to provide contextual information. Depending on the type of information being collected, the participant overhead may be simply too heavy to allow sustained involvement. Some of our participants expressed relief at the conclusion of the one week study as they

began to find it tiresome to annotate their web usage on a daily basis. This is unsurprising given the magnitude of pages that some participants had to classify.

There are instances of previous research that have successfully collected contextual information for extended periods of time. For example Kelly and Belkin [88] conducted a field study in which participants provided contextual information (e.g., task descriptions, measures of usefulness) on a weekly basis for 14 weeks. Therefore, participants may be willing to take part in a longer duration field study if the frequency of qualitative annotations is minimal (e.g., once a week versus once a day). However, as the annotations become more fine-grained, it becomes more important for the annotations to be provided in a timely matter. Therefore, weekly or monthly annotations may only be possible with higher level contextual information.

9.2.3 Real-time versus Post Hoc Annotation

We must also consider whether participants should provide their annotations in real-time or a post hoc basis. When the collection of participant annotations occurs in real-time, the characteristics of the activity are fresh in the participant's mind. However, the normal flow [112] of web usage may be interrupted which may impact natural behaviours. Alternatively, annotations collected at a later time are less intrusive; however, participants may not be able to accurately recall their activities. The decision of which method to use depends upon factors including the complexity of the data being collected, the distinctiveness of the activity, and the required frequency of data collection.

In our PG1 and PG2 field studies, participants provided privacy ratings at the end of each day using the electronic diary. Privacy ratings may change from one page to the next, so it would not have been feasible to interrupt the flow for each and every page to assign privacy ratings. Furthermore, privacy ratings were given based on privacy concerns for future viewing of the activity, not for concerns during the activity. It was therefore appropriate to have participants periodically reflect on future concerns using the page title and URL to remind them of the browsing activity. All participants assigned privacy ratings to all visited pages over the course of the week. During the uninstall session, participants indicated they did not find it problematic to assign their privacy ratings at the end of the day.

The electronic diary also allowed them to return to their annotations at a later time if they were unable to complete their daily classification.

In the second privacy study, location information was provided by laptop users in real-time through a browser pop-up window. We did not expect that participants would be able to accurately assign location information at the end of the day for all of their web usage, especially if they accessed the web from several locations. We were therefore willing to accept occasional interruption of flow for the benefit of more accurate location information. In order to minimize the disruption, the pop-up window appeared when a browser window was closing rather than when it opened. No participants commented that this was bothersome.

9.2.4 Data Collection

As previously discussed in Chapter 3, the choice of a data collection tool is strongly influenced by the type of data and the level of detail to be collected. The choice of the browser helper object (BHO) did limit us in several respects. The main drawback was its limited logging capabilities. We could only capture limited navigation events, such as web page URLs and document events, and not web browser interactions. In order to study participants across different locations, we needed to install the BHO on each of their computers. Our sample population was also limited to those that use IE on a Windows machine.

During the second field study, we wanted to capture windows focus events; but, due to an inability to hook into the IE browser window itself, our focus events were limited to the web document. In times of rapid browsing, not all events were captured, making analysis difficult (i.e. not all on focus events match a lost focus event). Furthermore, as documents could load in the background, it could be difficult to determine when viewing of one page ended and another began. Due to time limitations, this problem was not resolved to our satisfaction. We would like to resolve this issue in order to study how people move between different browser windows and tabs while conducting browsing activities.

One of the main reasons for selecting field studies as a methodology was to capture natural user behaviour. The focus of our research included not only an investigation of the sites they visited but also of their normal patterns of activity. The BHO was ideal in that it

did not impact participants' normal web browsing environment. In both studies participants could continue to use their usual browser (i.e. IE) and had access to all of their usual features, such as Favorites, History, and the Google toolbar. The automatic loading of the BHO with IE meant that participants did not have to remember to use the study instrument. As long as they were using IE on a computer with the BHO, their browsing data was captured.

Upon completion of the field studies, it was important for us to reflect on the perceived naturalness of our participants' behaviour. In the PG1 field study, we did not receive the page title and URL of visited sites and have no way of knowing if the browsing captured was indicative of normal behaviours. In the PG2 field study, we were able to inspect the visited pages. The proportion of participants in the PG2 field study with instances of adult content was comparable to frequency reports of erotica viewing as reported by participants in the IIP survey. This may indicate that we have captured participants' normal web usage, including those activities not considered to be socially desirable [45].

9.2.5 Data Transfer

One question that arises during field research is how to transfer the data from participant to researcher. While logged data during a laboratory experiment is typically stored directly on a research computer, when conducting research in the field we must determine where to store the data, when to transfer the data, and how to transfer the data. There are tradeoffs inherent to each approach. For instance, storing the data on the participant's machine for the duration of the study may simplify the participants' duties; however, researchers run the risk of data loss if the participant's machine crashes. If data is transferred more frequently, the participant may be inconvenienced.

We chose to build a custom application in which participants could email the researchers a daily data report after inspecting the data. This allowed us to review the data regularly to ensure that participants were properly annotating their data and to quickly spot problems with the data collection tools. For instance, in the PG1 study, we observed that the BHO was not formatting Chinese characters properly (in page titles) and were able to quickly issue a fix for the problem. The absence of data can also indicate that participants are

encountering difficulties. When we failed to see an email report from a participant for a two day period, we contacted the participant to inquire if there were any problems. During the PG1 study, 5/20 participants had problems with their software, their hardware, or their internet connections at some point during the study. While these participants did complete seven days of the study, their days were not consecutive.

While there were several advantages to participants emailing their data on a daily basis, problems did arise for some participants. In order to successfully use the custom email program, some participants had to temporarily disable their virus scanner. Less technically inclined participants sometimes failed to do so. In the PG2 field study, a few participants copied the text from the generated report and emailed that with their normal web-based email. Backups were also kept on the participants' machines for those cases when there were problems with the emailed data transfer or study software. The backups were created each time the data was accessed by the software (e.g., when opening the electronic diary). If problems were encountered with the emailed data reports, the data was recovered from the participants' computers during the uninstall sessions. This backup system ensured that no data was lost.

9.2.6 Data Analysis

One other aspect that has remained challenging for us is visualization of the data generated by the logging tools (see [69] for a framework of challenges in extracting information from logged data). Techniques are required for synchronization of various data sources and for transforming the low level captured events into meaningful instances of activity. Once transformed, techniques were needed for analysis of the data (e.g., summary statistics, pattern detection, and visualization).

Logging events can result in extremely large data sets, which can be difficult to manipulate and analyze. For instance, participants viewed a total of 36,170 web pages during the PG1 study and 31,160 web pages during the PG2 study. Therefore, it is important to be cognizant about how data transformation processes will be affected by very large data sets. Furthermore, it is not enough to rely on descriptive statistics when examining the data; patterns of activity are also important. Additionally, there are often multiple attributes related

to each visited page (e.g., content category, privacy level, secure/non-secure page, browser window, etc.) that must be tracked.

Visualization tools can be effective for understanding user behaviour, such as finding trends and patterns within the textual data logs. Figure 16 shows a visual representation of one hour of the logged data that was handcrafted during data analysis for the privacy study. Visualizations such as this can help researchers gain a better sense of which behavioural patterns should be further investigated through analysis of the logged data. In this case, the diagram helped us determine that streaks of browsing at a particular privacy level and transitions between privacy levels would be useful measures to calculate. More tools are needed that allow researchers to view combinations of logged and contextual data. While this was a topic of discussion at the recent WWW 2006 workshop "Logging Traces of Web Activity: The Mechanics of Data Collection", there do not appear to be robust solutions that are easily customized to the specifics of the data collected.

9.2.7 Summary

Studying user behaviour on the Web is a difficult area of research. It can be challenging to capture realistic behaviours when users are not studied in their natural environment, engaging in intrinsically motivated everyday activities, and using their normal tools. An understanding of these realistic behaviours is required in order to appropriately ground development of new web-based tools and techniques. It is therefore important that focused laboratory studies and attitudinal surveys are augmented with field research. In our research, we have found field study methodologies to be effective at capturing a rich set of behavioural data. In particular, we found that contextual information provided through participants' privacy annotations, coupled with logs of web usage, afforded valuable insight into our participants' privacy attitudes and web browsing behaviours.

9.3 Evaluating Privacy Management Approaches

One of the advantages of the field study data was that it let us examine naturally occurring patterns in the application of privacy patterns. This allowed us to theoretically examine the feasibility of different automated privacy management approaches (as presented in Chapter 7) and suggested that streaks of browsing at a given privacy level could be leveraged to reduce the user burden of classification in a more explicit privacy management

scheme (as presented in Chapter 8). Evaluation of the effectiveness of the privacy management approach used by PrivateBits proved to be challenging. We focus on two aspects of our evaluation: the tradeoffs between conducting a laboratory evaluation rather than studying longitudinal use in the field, and the tradeoffs between controlling the browsing and viewing scenarios rather than grounding them in participants' natural browsing and viewing contexts. For both aspects, we discuss the various approaches that may have been applicable and the evaluation methods we chose and then reflect on the suitability of our approach.

9.3.1 Laboratory Evaluation versus Usage in the Field

In order to be able to fully validate the management approach taken in PrivateBits, it will be necessary to examine long-term usage patterns to see if people find the burden of maintaining the system to be less of a cost than visual privacy violations. However, long-term field evaluations require the system to be fully developed. In our case, the software would need to be fully functional as a web browser, as well as be capable of managing privacy and logging many of the users' interactions with the system.

Robust prototypes with enhanced functionality have been implemented in the past and used in longer term evaluations. For example, SmartBack was a prototype IE web browser with enhanced functionality for revisiting pages within a session [106]. It should be noted that some of the authors were Microsoft researchers, so they may have been able to modify an existing version of IE. Kellar et al. implemented a custom version of IE for the purposes of logging browsing activity in a field study investigating browser tool usage during various information seeking tasks [86]. However, some performance issues were encountered during use of the custom web browser in the field [85], confirming the difficulty of developing research software that meets the expectations that participants have for commercial applications.

We first needed to investigate whether the basic technique of using browser windows of varying privacy modes to both filter traces and semi-automatically classify new activity would be a viable approach. We therefore decided to perform a preliminary evaluation in the lab to validate our approach before expending the effort required to develop robust custom software.

Developing PrivateBits as a piece of laboratory software allowed us to make the system just as robustly as needed to work on a single computer in the lab. We did not have to worry about various user configurations of Windows and IE. We were also able to ignore many of the browser functions. Although visually PrivateBits was a clone of IE, many functions tangential to the browsing activity (e.g. menu items such as print) were not actually implemented, thereby cutting back on development time.

9.3.1.1 Participant Sample Size

We targeted both office workers and students to evaluate if the proof of concept application worked well across a variety of usage contexts. During formative evaluations as part of an iterative design process, it is recommended to have three to five users from each population group in order to identify most of the design issues with the interface [109]. Given the individual differences inherent in privacy concerns, five participants were recruited from each group.

Participants were screened prior to inclusion in the study to ensure that they qualified as potential users of such a privacy management system. Participants must have had regular occasions where others could view traces of previous activities on their display and have had some privacy concerns related to this viewing. Not everyone will find such a privacy management system to be necessary or have viewing situations that merit expending the effort to manage privacy.

We feel that the relatively small sample size was appropriate at this stage in the research. We gathered rich data from these participants regarding their interactions with PrivateBits and the effectiveness of the per-browser window privacy mode approach to classifying traces and filtering them appropriately. However, once PrivateBits has been modified to incorporate the feedback received, future evaluation is required to validate this approach during long-term usage in the field.

9.3.2 Maintaining Control versus Encouraging Natural Behaviours

There were two main components of participants' interactions with PrivateBits that needed to be examined. First, we needed to see if participants could use the privacy modes of the windows to appropriately classify the pages they visited with a privacy level. Second, we needed to see if participants could appropriately filter the generated content when in a

collaborative situation. For both of these components, there was a tension between maintaining control of confounding variables while encouraging natural behaviours. As an additional constraint, we wanted to keep the evaluation session to an hour or less so that participants would not be overly fatigued. We next discuss the challenges for each component and the appropriateness of our solution.

9.3.2.1 Evaluating Effectiveness during Classification

One key challenge was to motivate participants to browse content with a variety of sensitivities while still encouraging natural web browsing behaviour. Although lab evaluations enable researchers to have more control over confounding variables; personal information research and usable security and privacy research requires that participants can relate to the data they are working with.

When evaluating web browser convenience feature enhancements, a common technique has been to provide participants with a narrowly focused set of navigation tasks over a set of web pages. For example, Cockburn et al. [32] evaluated different mechanisms for the back button by having users complete 19 short navigation tasks based on realistic navigation scenarios (e.g. hub and spoke navigation). The tasks were conducted within three websites: two of these websites were stripped down versions of existing sites, and one was a simple plain text site generated for the study. While this approach tightly constrained participants' browsing activities, it also only examined one small component of web browser use (i.e. the back button). Having participants focus closely on navigating between pages was appropriate in this case; however, for a study investigating attributes of web browsing at a higher level, such focused web browsing may interfere with participants' ability to relate to the pages they visit and to engage in their natural behavioural patterns.

Another technique that has been used when evaluating browser enhancements has been to populate the feature (e.g. History) a priori with data generated for the study (as in [76]). The benefit of this approach is that all participants are working with the same data, so comparative evaluations are possible. However, such data is artificial and it may be difficult for participants to envision privacy concerns for browsing activities that they had not actually undertaken [128, 152].

One technique that can enhance the reality of lab studies is to have participants come in to conduct a pre-existing task within the lab environment. For example, for a study of information seeking behaviours, one approach would be to ask participants to conduct a search for information related to a current need (e.g. class project). However, such an approach would be unlikely to have participants generate content in a range of privacy sensitivities as it is very task centric. Another approach would be to have participants conduct more free form browsing, asking them to conduct a range of their normal browsing activities. However, participants may be unwilling to conduct sensitive browsing under the researcher's scrutiny. Furthermore, without access to their convenience feature data such as Favorites and Auto Completes, participants may have difficulty accessing their "normal" sites.

We chose to provide browsing tasks for users based on six realistic scenarios. Selection of the scenarios was guided by examples of incidental information privacy concerns collected through our exploratory research. Although these scenarios may not have been personally motivated, we customized them so that they were grounded within participants' personal networks (e.g. motivating a search for information about cancer because a friend has been recently diagnosed). This strategy allowed us to ask users to search for information across a range of sensitivities. This approach is similar to that undertaken by Lederer et al. [91] who used scenarios that described a specific activity in a specific context when examining how their participants' privacy faces matched their disclosure preferences. Their participants were asked to provide two general situations they often found themselves in (e.g., shopping during the weekend) and the participants were asked to created faces for those situations for two different inquirers. The specific scenarios were chosen to be somewhat sensitive events that met the constraints of the more general situations that users had defined (e.g., buying a pint of chocolate ice cream at the grocery store on Main Street at 10pm on Saturday night).

We hoped that by keeping the search task fairly high level, that participants would not become bogged down in the details of the task. Furthermore, as each search required participants to evaluate whether or not the search results were worthwhile saving for future reference, participants may have been motivated to interact with the pages as they would during a normal search for information. In an attempt to provide opportunities for multi-

tasking, the six scenarios were introduced incrementally in four batches, with two of the batches containing two scenarios.

Our strategy was still not natural in terms of participants having their normal browsing environment. Participants were required to be regular users of Internet Explorer, so the browser was familiar; however, the convenience features did not contain any of their data. Furthermore, we restricted search engine use to Google which was the only search engine for which our form Auto Completion worked. Only one participant mentioned this to be a problem for one of the search tasks. This participant would have chosen a child friendly search engine for the reproduction search scenario that was motivated by a neighbour's child requiring the information.

As our primary concern was not to do a comparative evaluation with another privacy management technique, precise control of which sites were visited was less important to us than being able to evaluate whether PrivateBits was able to be flexible enough to meet the diverse privacy concerns of participants. We feel that our strategy was effective at motivating participants to conduct browsing across a variety of sensitivities without unduly constraining their normal browsing behaviours in terms of concurrent browser window usage and thoroughness of searching.

9.3.2.2 Evaluating Effectiveness during Filtering

In order to evaluate the effectiveness of PrivateBits at filtering content appropriately for the viewing situation, we wanted to examine how participants might use the browser to filter traces of browsing activity for their most regular viewers. We did this by using viewing scenarios personalized with the names of some of the participants' most regular viewers as specified during the pre-session survey. In order to ensure that we examined a range of privacy comfort levels, we included one viewer that the participant was most comfortable with, one they were least comfortable with, and one with a comfort level in the middle. The fourth viewer was selected to give breadth in the types of viewers if necessary by including a viewer with a hierarchical relationship (e.g. supervisor) or a personal or work relationship as applicable.

We feel that our viewing scenarios were effective at allowing us to evaluate how well participants were able to filter their viewing for a range of their regular viewers. The

downside of this approach was that each participant did not have the same viewing scenarios. However, it would be of questionable value to have participants reflect on the effectiveness of the interface at filtering their browsing appropriately for a type of viewer that never occurred (e.g. a spouse/significant other if they have none) or to have them reflect on their comfort level for an infrequent viewer type rather than one of their more regular viewers. We believe that selecting participants' regular viewers rather than maintaining consistency of viewer types across participants was appropriate given the goals of this evaluation.

One aspect of our viewing scenarios that was flawed was that we neglected to ask participants how they would feel if the browsing they had just conducted was visible to the viewer without any privacy management system in place. The omission of this question meant that in order to evaluate whether participant's comfort level increased, we could only make a comparison with their stated privacy comfort level for the viewer during the pre-session survey. This pre-session comfort level was not contextualized for the browsing that had just been completed. However, the fact that participants chose privacy modes that did not reveal all browsing reinforces our claims that their privacy comfort level was increased through use of PrivateBits.

9.3.3 Summary

In summary, our approach consisting of a laboratory evaluation using personalized browsing and viewing scenarios was effective at achieving our evaluation goals. We were able to determine the effectiveness and utility of our privacy management approach for browsing across a variety of sensitivities. Participants were able to filter the activity traces appropriately for their most regular viewers. Despite the relatively small sample size, we were able to investigate two segments of potential users (students, office workers) and received rich data that confirmed some of our design choices and showed areas where improvements are still required.

9.4 Summary of Methodological Approaches

In this chapter, we reflected upon the suitability of the methodological approaches undertaken during this research. With each approach, there were tradeoffs that needed to be made. We first discussed the suitability of the mixed methodological approach used for our

exploratory studies of incidental information privacy concerns. We then reflected on the effectiveness of participant annotation of logged data as a method of studying rich natural behaviours in situ. Finally, we reflected upon the suitability of the approach taken with the laboratory evaluation of PrivateBits. Overall, we feel the tradeoffs we made were suitable given our research questions.

Chapter 10
Conclusion

Our initial research direction was motivated by a lack of mechanisms to guard visual privacy during ad hoc collaboration. Previous privacy research in CSCW focused on those instances where collaborators were interacting at a distance and only wanted to share some aspects of their data, or when they were working closely together on specialized equipment or on devices dedicated to collaboration. There was no research investigating how to manage privacy when an individual's personal computer is used within a collaborative setting and information that is incidental to the collaborative task is visible on the display. While anecdotal evidence existed that incidental information privacy was a concern in such a scenario, it was important to validate that incidental information privacy concerns occur in the general population.

As we presented in section 5.1, results from our exploratory research validated our motivations for investigating the factors of incidental information privacy with respect to web browsing. Privacy of incidental information was indeed a problem for most participants. For the IIP survey, all 155 participants reported at least one category of viewer that could sometimes *see* their display and 145/155 participants reported at least one category of potential *users*. Trusted viewers such as spouses and close friends tended to be regular viewers; however, some of the most frequent viewers were colleagues and supervisors, both of whom tended to have lower overall privacy comfort levels. Not only did participants have incidents when others could view their displays, most were also concerned enough to take some steps to maintain the privacy of the incidental information that may be displayed. The majority of the participants indicated that they would take some action if given advance notice that someone may view their screen.

In this final chapter, we begin by providing a brief summary of the contents of this dissertation. We then itemize the main contributions of this thesis research. We discuss several opportunities for future work before giving final conclusions.

10.1 Dissertation Summary

Web browser convenience features (e.g. History, Auto Complete, Favorites) display traces of previous activity in order to assist users in refinding pages and recalling search terms. However, use of these features in a group setting can be problematic as they may display information unrelated to the task at hand. This incidental information may be inappropriate for the viewing context. Fine-grained and flexible mechanisms are required so that users can present contextually appropriate content during collaboration while still preserving the functionality of browser features.

A mixed methodology approach of a survey (155 participants) and two week-long field studies (35 participants total) was used during our exploratory research to investigate visual privacy within web browsers. The survey examined participants' privacy concerns during varying usage scenarios, while the field studies examined participants' application of a four-tier privacy gradient to their actual web browsing activity. An examination of the field study data allowed us to update understanding of general web browsing activity such as frequency, speed, and content.

This foundational research also investigated visual privacy concerns during web browsing. While prior privacy theory and research had given us indications of the individual and contextual nature of privacy, it was important that we investigate visual privacy concerns within the context of the primary task of web browsing. We identified several factors that impact a person's privacy comfort level in a given situation including their inherent privacy concerns, perceived sensitivity of potentially visible content, level of control retained over input devices, and potential viewers of the traces of web browsing activity. Beyond these factors, we also investigated how privacy concerns and browsing behaviours varied according to other dispositional and situational variables. Through this analysis, we formed an initial conceptual model of visual privacy in this domain.

Our results also led to the development of design requirements for privacy management systems in this domain. Such systems must enable easy classification of new traces of browsing activity and provide mechanisms to appropriately filter those traces during subsequent collaboration. As documented in our results, rapid bursts of activity and magnitude of pages visited during web browsing suggest that some system support will be necessary for privacy classification to be manageable.

We investigated the feasibility of an automated approach to privacy classification: categorizing the content of viewed pages and assigning privacy levels on a per- content category basis. We found that a generic approach was not feasible given the individual differences in privacy concerns, but a personalized approach may be appropriate. However, given the current low coverage of automatic content classification and differences in privacy sensitivity of pages within the broad content categories, further work will be required before an automated classification approach is feasible.

We also investigated the feasibility of providing automated support for filtering of traces through use of a predictive model of visual privacy concerns. This preliminary predictive model was developed through multiple regression analysis of the IIP survey data. Our theoretical understanding of the contextualized nature of privacy in this domain guided our selection of independent variables in the analysis. A predictive model shows promise as the basis of an intelligent systems approach for filtering content according to situational privacy concerns. However, the survey was not developed with predictive modeling in mind, so our modeling capability is limited. Furthermore, the more contextual predictive models we developed for spouse and supervisor showed how dispositional and situational variables of interest varied according to specific viewer situations. Future work will be required with a larger number of participants across all inherent privacy concern segments and more contextually focused privacy comfort data in order to develop models suitable for a range of contexts and users.

During the field studies, we identified patterns in the application of privacy levels (e.g., private browsing conducted within a single browser window, minimal changes between privacy levels within windows). To support classification of web browsing activity, our proof of concept privacy management solution capitalizes upon these patterns by providing users with browser windows of different privacy modes and allowing them to change the privacy mode of the window when the sensitivity of the browsing changes. We designed and implemented a custom web browser, PrivateBits, to evaluate the feasibility of this approach. PrivateBits was found to be flexible enough to meet varying participant concerns, privacy management strategies, and viewing contexts. Our results emphasized the need for additional security features to increase trust in the system and raised questions about how to best manage the tradeoff between ease of use and concealment of the system itself.

Finally, we reflected on the suitability of the methodologies we used during this thesis research. We discussed how the mixed methodology approach allowed us to investigate general privacy concerns, and also gave us enough data from actual browsing behaviours to investigate the feasibility of various approaches. We reflected upon the suitability of the data collection tools we developed for use during the field studies. The browser helper object allowed us to log all web browsing activity on a per window basis without visibly altering the participants' browsing environment. The electronic diary allowed participants to qualitatively annotate their visited pages with a privacy level on a daily basis. We also examined the challenges of evaluating privacy management approaches with users. Our methodology needed to evaluate our proof of concept system, PrivateBits, across users with varying privacy concerns conducting browsing activities with a range of sensitivities. Furthermore, we needed to provide an environment that encouraged natural browsing behaviours.

10.2 Thesis Contributions

This dissertation research has made several contributions to the fields of Usable Security and Privacy, Web Browsing Behaviours, Personal Information Management, and Human Computer Interaction. Some of these contributions have been presented, in whole or in part, in earlier publications. A detailed list of the publications and presentations generated as a result of this dissertation research can be found in Appendix E. As we discuss the main thesis contributions this section, we will identify pertinent publications.

10.2.1 Updating General Web Browsing Behaviours

Although web browsing behaviour was studied in detail in the mid-to-late 1990s, few recent results have been reported. The nature of web browsing has changed extensively since these early studies, both in the profile of the typical web user and in the context of their browsing (e.g. location, connection speed, web browser features). Before developing a privacy management system for use during web browsing, it was important to understand web usage patterns that might impact system design.

As presented in Chapter 4, the results from the two field studies (PG1 and PG2) clearly demonstrated that variability and magnitude of browsing behaviours complicate the development of any tool or technique for web browsing. The sheer number of pages that

people visit while browsing means that manual tools, that operate on a per-page level, will be overly arduous and therefore impractical. Beyond just the number of pages visited, the speed with which users browsed was at times staggering. The high volume of web sites visited and the rapid browsing indicate the need for seamless interactions between users and their web browser tools. Participants' behaviours varied considerably in terms of the number of pages visited, number of separate windows in use, and the session length and speed of browsing. Participants also varied in the relative frequency with which they visited various categories of pages. Furthermore, there can be great variability both across users and within the browsing of a single user. This variability makes it difficult to arrive at standard solutions for web browsing tools and techniques. Web browsing tools and techniques must be sensitive to the changing needs and behaviours of users and allow users flexibility in their interactions with the system.

Our results from the PG1 field study were presented in a poster at CHI 2005 (Appendix E, E2), while analyses from the PG2 field pertaining to the categories of visited pages was presented in a long paper at WWW 2006 (Appendix E, E4).

10.2.2 Modeling Incidental Information Privacy Concerns

While there has been much research investigating privacy in various domains, little has directly examined visual privacy issues of incidental information. Furthermore, most prior research has focused on a subset of factors relating to privacy concerns. We were unclear the extent to which these factors applied to incidental information privacy concerns within the context of web browsing. Before designing a privacy management system, it was therefore important that we determine which factors of privacy apply to this domain (and their inter-relationships) so that our solutions were grounded appropriately. The IIP survey and two field studies combined to give us a rich picture of incidental information privacy concerns.

As presented in Chapter 5, our results showed us that, overall, the privacy comfort level in a given situation depended upon the perceived sensitivity of potentially visible content, the relationship to the viewer, the level of control retained over input devices, and the person's inherent privacy concerns. Furthermore, we found there was variability in the

importance of the different factors according to participants' inherent privacy concerns; not all users were concerned across all factors.

We were able to segment our IIP survey participants into privacy classifications. These segments were determined by participants' level of overall privacy concerns and the magnitude of contextual differences in those privacy concerns across the different viewing contexts (i.e., viewer, level of control, content sensitivity). Privacy fundamentalists are those participants with little differences according to context and low overall privacy comfort levels. Privacy unconcerned participants are those with little differences according to context and high overall privacy comfort levels. Privacy pragmatists are those participants with high contextual differences. Privacy pragmatists can be further subdivided according to their overall privacy comfort level (wary, circumspect) or according to factors of privacy that impact their concerns (i.e., viewer, level of control, content sensitivity). These classifications could be used to determine suitable default settings for a privacy management system based upon a person's responses to a questionnaire during system initialization. The examination of the overall factors of incidental information privacy and an initial attempt at segmenting participants according to their inherent privacy concerns was presented as a long paper at CHI 2006 (Appendix E, E1).

As presented in Chapter 6, we also examined the impact of dispositional and situational variables on an individual's inherent privacy concerns. We then extended our initial model of incidental information privacy concerns to include the impact of these variables on contextualized privacy concerns in a given viewing situation. In addition to examining responses from the IIP survey, we found support for our model in the data collected during the PG2 field study. Furthermore, we examined how situational variables impacted web browsing behaviours including the types of web sites visited, the browser convenience feature settings, and the actions participants reported taking to preserve their privacy. These behaviours combine to produce the incidental information which may become visible within web browser convenience features during collaboration.

Our model of incidental information privacy, which includes both dispositional and situational variables, provides a theoretical contribution. This rich model is unique in its incorporation of multiple factors as well as its coverage of both privacy concerns and the activities that generate the information to be protected. This model can be used as a guide

for future study of visual privacy concerns both of incidental information within web browsers and also for other personal information management systems which may give rise to similar incidental information privacy concerns. This model may also be of benefit to researchers investigating other privacy domains, particularly those with mobile users, changing contexts, or changing user roles. Furthermore, as shown by our attempts at developing a predictive model in Chapter 7, the model can be used to inform practical privacy management solutions for visual privacy domains.

10.2.3 Examining Patterns in Privacy Application

One of the concerns about developing a system to enhance privacy management is that the act of managing privacy should not interfere with users' normal behaviours for their primary task. It was therefore important that we be able to not only examine participants' general web browsing behaviours, but also the privacy concerns that they had within the context of their normal web browsing activities.

The PG1 and PG2 field studies allowed us to gather participants' qualitative annotations giving us their privacy concerns for the actual web browsing activity they engaged in over the course of a week. This allowed us to analyze the data for patterns in the application of privacy levels (as presented in section 5.2). Analysis of the data from both field studies showed temporal patterns in privacy application on a per window basis. We determined that most browsing occurred in streaks at a given privacy level and that there were minimal transitions between privacy levels within a window. Furthermore, private browsing activities were often partitioned to a single window. These underlying patterns were leveraged when we designed and developed PrivateBits, our visual privacy enhanced proof of concept web browser.

The additional contextual data collected during the PG2 field study also allowed us to examine how the application of privacy levels changed depending on the category of web site being visited and the location in which the browsing occurred. The analysis of privacy levels as applied to content categories was used to determine the feasibility of an automated privacy classification approach. Patterns in the application of privacy levels related to content and location were also important as we developed our model of visual privacy within web browsers.

The patterns found in our PG1 field study were presented as a short talk at CHI 2005 (Appendix E, E2). The examination of patterns according to content category of visited pages and the feasibility of content categorization as a privacy management approach was presented as a long paper at WWW 2006 (Appendix E, E4).

10.2.4 Developing Design Guidelines

As was presented in section 7.1, our exploratory research combined to provide several design guidelines. Some of these guidelines are applicable to the design of web browsing tools including increased visualization of settings, clearer explanation of feature functionality, and more intelligent default settings according to the contexts of use. Others are applicable to the design of tools to help users manage the visual privacy of the incidental information which may be visible within web browsers. These include reducing the clutter within convenience features, allowing nuanced privacy classifications, supporting multi-tasking, supporting diverse privacy concerns, and reducing the burden of privacy management. These guidelines provide a practical contribution to designers and developers of enhanced web browser features and visual privacy management tools. The guidelines will be presented as a long paper at Graphics Interface 2007 (Appendix E, E10); preliminary requirements were presented as a poster at SOUPS 2006 (Appendix E, E12).

10.2.5 Evaluating Privacy Management Approaches

A variety of privacy management approaches were suggested as a result of our exploratory research. Our design guidelines further shaped our vision of what techniques might be appropriate. We focused our investigation of privacy management approaches on techniques that would lessen the burden on users of classifying their browsing activities with a privacy level and subsequently filtering that content appropriately. We conducted two theoretical evaluations of automated approaches: content categorization for classification and a predictive model for automated filtering of content. Our proof of concept privacy enhanced web browser, PrivateBits instantiated and validated our design guidelines and a privacy management technique that semi-automatically classifies traces of activity. This technique leveraged the privacy patterns we observed during the field studies (i.e. streaks of browsing at a given level, minimal transitions, partitioning of sensitive activities). Our

investigation of these various privacy management approaches provides a practical contribution for other researchers and practitioners.

10.2.5.1 Examining Content Categorization for Classification

In section 7.3, we presented our theoretical evaluation of content categorization as a method of automatically classifying traces of browsing activity with a privacy level. This evaluation highlighted the need for a personalized or flexible approach to privacy management. Participants in the PG2 field study exhibited a lack of agreement about an appropriate privacy level for each category of content during both the theoretical classification task and in their application of privacy levels during real browsing. This approach is not currently feasible given the lack of coverage of content categorization and a need for more fine-grained mechanisms to further specify a privacy level within some categories. We provided several suggestions for how to achieve better classification accuracy which may guide future work. This evaluation was presented in a long paper at WWW 2006 (Appendix E, E4).

10.2.5.2 Examining a Predictive Model for Use during Filtering

In section 7.4, we used the privacy model we developed as the theoretical basis for multiple regression analyses. We developed preliminary predictive models of privacy comfort in a given situation. Such models could be used in an automated approach to content filtering during times of collaboration. The models developed included a general model for overall privacy comfort and two models developed for specific viewer types (spouse/significant other, supervisor). The differences in the variables included in each of the models further demonstrated the highly contextualized nature of privacy concerns.

While this approach shows promise; future research will be required in order to develop and validate models specific to varying contexts of use and customized according to the privacy segmentation of users. Our initial attempts were limited as our sample size was insufficient to develop models for each of the privacy segmentations of our participants. In addition, we had limited data in terms of which prior activities may have been considered by participants when they reflected on their privacy comfort level for their recent web browsing activities. Furthermore, the privacy comfort data from our survey was not contextualized in

terms of location or device. We were therefore unable to fully explore how all the situational aspects of visual privacy concerns contributed to the predictive model.

10.2.5.3 Examining a Browser Window Privacy Mode Approach

Given the limitations of the automated approaches we considered, we chose to use a more explicit approach to classification and filtering when instantiating the design requirements in a proof of concept application. We leveraged the underlying privacy patterns we observed during the field studies. We presented the design and preliminary evaluation of PrivateBits, a custom web browser that allows users to manage traces of previous activities that may be visible within browser convenience features (Chapter 8). Users can open browser windows of varying privacy modes. They can select a privacy mode that is appropriate for the browsing task for which the window has been opened. Each window tags browsing activity with the current privacy mode and filters which traces of prior activity are displayed. This approach allows users to consider the privacy sensitivity at the task level rather than on a per-page basis.

The prototype was found to be effective by participants for classifying the visited pages with an appropriate privacy level and for filtering traces appropriately during viewing scenarios. The evaluation provides initial support for the design requirements as well as for our semi-automated privacy classification approach. However, given that our initial evaluation of the prototype is limited by the small sample size and lab environment, further study will be required to validate the requirements and privacy management approach in a natural usage environment.

Our evaluation also revealed participant concerns that are unique to privacy management systems. Concerns relating to trust in the system indicate that security mechanisms such as password protection and encrypted data must be provided. Furthermore, the need to conceal the existence of the privacy management system from others has raised questions about appropriate methods of managing the tradeoff between ease of use of the system and its privacy. Users should be provided with several levels of visibility that they may choose from between depending on their usage environment. As privacy management systems become more commonplace, the need for concealment may lessen.

The design and evaluation of PrivateBits will be presented in a long paper at Graphics Interface 2007 (Appendix E, E10). The interface for PrivateBits was presented in a poster at SOUPS 2006 (Appendix E, E12) and PrivateBits was demonstrated at CSCW 2006 (Appendix E, E11).

10.2.6 Methodological Contributions

As we designed our studies throughout this dissertation research, we encountered several methodological challenges. The choice of an appropriate approach in this area is challenging as evidenced by several recent workshops we have attended with a methodological focus. Our reflections upon the appropriateness of the data collection tools we developed and the research methodologies we employed (as presented in Chapter 9) provide contributions to various research areas.

10.2.6.1 Data Collection Tools

We developed a browser helper object for use during our field studies to record participants' web browsing activities in a non-intrusive manner. We also developed an electronic diary that allowed participants to qualitatively annotate their visited pages with a privacy level on a daily basis. We provided privacy protection for our participants' data in order to encourage natural browsing behaviours throughout the studies. For the PG1 field study, we did not receive the title and URL of visited pages; for the PG2 field study, we allowed users to blind the title and URL of page visits they did not want to share. We found our data collection methods to be effective at capturing natural web browsing activities while still gathering rich qualitative data.

Our research methodologies have been of interest to both the HCI and the WWW communities. We first presented our field study methodologies, including our technique for allowing participants to annotate their data at the CHI 2005 workshop on Usage Analysis: Combining Logging and Qualitative Methods (Appendix E, E6). The appropriateness of a browser helper object for capturing natural user behaviour on the web was further discussed at the WWW 2006 workshop on Logging Traces of Web Activity: the Mechanics of Data Collection that we organized (Appendix E, E7-E8). A journal article will soon appear in the International Journal of Human Computer Interaction as part of their special issue In Use, In Situ: Extending Field Research Methods (Appendix E, E5). This article contrasts the data

collection and annotation techniques we used with those used by our colleague Melanie Kellar during her dissertation research.

10.2.6.2 Privacy Research

Privacy research is also a challenging area in and of itself. It can be difficult to elicit participants' privacy concerns in a research setting. Questionnaire responses may be a better indicator of privacy preferences than of the actions that participants would actually take during normal system usage. We found our mixed methodology approach of a survey and two field studies to be effective at studying privacy in this domain. The survey represented users' self-reported perceptions of their concerns; however, it was important to build a more complete picture grounded in actual behaviours by combining survey results with those from the field studies. Our mixed methodology approach was presented at the CHI 2006 workshop entitled Privacy and HCI: Methodologies for Studying Privacy Issues (Appendix E, E9).

10.3 Future Work

We next present some areas of future work that are suggested by the results presented in this dissertation. We begin with future work we would like to conduct in order to further develop and evaluate a privacy management approach to the incidental information visible within web browsers. We then discuss our plans for future work beyond the web browser as we consider other visual privacy management issues.

10.3.1 Visual Privacy Management within the Web Browser

10.3.1.1 PrivateBits

Our first steps are to refine the interface of PrivateBits, incorporating the feedback received during our evaluation. It is clear that in order to be flexible enough to accommodate users with varying privacy concerns, web activities, and viewing environments, we must allow for personalization of the interface and functionality. Our goal is to implement the next version of PrivateBits as a toolbar extension with versions for both IE and FireFox and to also support tabbed browsing.

Once PrivateBits has been refined, we will examine the tradeoffs for users with respect to system concealment. Participants expressed the desire to conceal the visual

feedback mechanisms of the privacy management system as well as the existence of the system itself when their display was visible to others. The ability to make privacy management functions invisibly accessible, with quick access and concealment as the situation dictates may impact adoption of these systems. A basic tenet of user interface design is to increase ease of use through visibility of options and system status. However, with privacy interfaces that will be used in the presence of others, there may be a conflicting need for discretion. An appropriate strategy for maintaining the tradeoff between ease of use (visibility) and privacy (concealment) depends on the contexts of use, including the frequency of viewers, the casual visibility of the display, the sensitivity of the information, and the social norms of the environment.

We will examine concealment mechanisms using our PrivateBits browser as a test bed. A laboratory evaluation will investigate the effectiveness of a range of feedback mechanisms of varying subtlety. We will also evaluate whether users can remain effective at managing their privacy using few visible affordances. Once we have validated the interface of PrivateBits in the lab, we will conduct a longitudinal evaluation in the field. This will allow us to determine whether participants find its privacy management approach tractable for daily use.

10.3.1.2 Automated Approaches

During our theoretical evaluation of content categorization as a method of automatically classifying visited pages with a privacy level, we theoretically evaluated a single classification scheme (i.e. the Cerberian content categories [1] as used in Zone Alarm). However, content filtering and classification of web pages is an active research area and there may be other approaches that are more suitable for our needs. For example, we may be able to determine which attributes of a web page provoke privacy concerns. Our results revealed that some heuristics such as keywords, secure pages, and logins showed promise as mechanisms to more finely adjust the privacy classification. It is possible that those heuristics combined with other attributes of web pages could provide better accuracy when determining the sensitivity of a specific page than our two step process (i.e. categorizing the content of the page and then applying a privacy level for that content category). Further study is required to determine a set of potential attributes. Given the variability we have seen in privacy classifications of pages, it is likely that users will draw on different attributes to

determine their privacy concern or that they will weight the same attributes differently. We will also need to build a corpus of pages classified with privacy levels by several participants. We can then evaluate whether some subset or weighting of the attributes can be used for each participant to effectively classify the pages with a privacy level according to their concerns.

We would also like to refine our model of incidental information privacy. Before we can refine the predictive model, we will need to gather contextualized data about privacy concerns. For example, we need to investigate privacy concerns related to device use in a given location. When designing the study questionnaires, we will use our model of incidental information privacy as a theoretical basis. This should help ensure that questions are sufficiently focused to elicit privacy concerns across participants' specific situations of use. We also need to gather more information about the specific content that participants are considering as they reflect on their privacy concerns for their recent web browsing activities. Once the models have been refined, we will need to validate them against actual privacy concerns gathered in the field.

10.3.1.3 Development of a Blended System

One of the problems with automated approaches to privacy management is that users may not trust the system to act appropriately. Combining an intelligent systems approach within a user controlled environment may allow users to better identify concerns as they arise and reduce the burden of maintaining a privacy management system. We would like to augment PrivateBits with some automated approaches to support users in their privacy management. For example, an automated privacy classification scheme might be used to flag content that may have been inappropriately classified by the mode of the browser window. The privacy models could be used to determine more appropriate default settings for a user's specific environments of use. The system could use the predictive models to recommend appropriate levels of filtering given an upcoming situation of use. The models could also be adjusted in response to the user's actual actions taken so that the system becomes more attuned to the user's privacy actions rather than being primarily based on their privacy preferences as garnered through initialization questionnaires.

Once the individual components of such a system have been evaluated in a controlled fashion, the system will need to be validated in the field. It will be important to

evaluate whether the automated approaches provide a useful benefit to the user, actually assisting them in managing their visual privacy, or whether they are viewed as being too intrusive or annoying. In order to determine system performance and feature use over time, a longitudinal evaluation will be required.

10.3.2 Extending Privacy Management beyond the Browser

10.3.2.1 Other Personal Information Management Systems

While our focus has been on developing a privacy management system for web browsers, lessons that we have learned may be applicable to visual privacy issues in other personal information management (PIM) systems. Rather than building privacy management systems to fix the privacy problems that arise from existing applications, it would be better to address privacy concerns during development of the applications. Affordances should be provided for those users who have visual privacy concerns due to working in close collaboration with others. We presented this perspective at the PIM 2006 workshop (Appendix E, E13).

New paradigms for storing and accessing information are enhancing PIM systems, but they also increase opportunities for visual privacy violations. For instance, search allows users to find information without remembering precisely how the information was generated or saved; however, users may be less sure of which information will be revealed within search results than if they had navigated a known hierarchy. This problem can be exacerbated in systems that incorporate results across tasks or applications or provide enhanced visualization of the results through thumbnails or snippets of text (as in [43]).

Tagging is gaining increasing acceptance as a mechanism for assigning multiple attributes to personal information. It may emerge as a useful method for classifying the privacy level of items in the personal information space and for filtering results appropriately. One recent paper about the PIM system Phlat [37] gives an example using a 'personal' tag to organize and filter items, although the privacy of items was not a focus of the paper. However, many of the test users of Phlat did not make use of tags and consistent management of tags can be overly burdensome for users. The authors of Phlat note that tags should be able to be applied during the workflow as information is encountered and also when decisions are being made about saving information items. Our research has examined

different approaches to managing the privacy classification of visited web pages at the time of browsing which may be applicable in other PIM systems. It may also be possible to automatically associate privacy tags with other tags being applied, such as tags for people, task types, and content types. For example, information tagged as being related to one task may have a different privacy association than for information associated with another task. With such a scheme, users would not have to additionally consider the privacy for each item encountered.

10.3.2.2 Managing Visual Privacy Across Applications

In addition to personal information management systems, notification systems (e.g., email alerts) can raise similar privacy concerns. My dissertation research investigated visual privacy concerns for a single application. The next step is to investigate the problem more globally, examining aspects of the desktop, including the file system, applications, and notification systems. It is critical that solutions we develop work across a breadth of applications.

Many applications are beginning to consider privacy concerns, allowing users to configure how information is presented. However, it can be difficult during times of collaboration to individually change application configurations to maximize privacy by minimizing the information revealed. We propose to enable privacy management at the desktop level, and allow users to specify overarching privacy modes which can adjust the visibility of files, notification events, and application behaviours accordingly. Semi-automated approaches to classification in such a privacy management system will be complicated by the tendency of users to multi-task. Recent advances in activity based computing [13] may provide a suitable level of abstraction, allowing users to associate a privacy level with an activity so that the set of services and data associated with the activity are modified appropriately when the privacy mode of the desktop is changed. We will investigate privacy concerns and the feasibility of potential solutions through a mixed methodology approach. Once design requirements have been developed, a prototype system will be implemented and our approach evaluated.

10.4 Conclusions

Visual privacy issues can occur when people collaborate around someone's personal computer. This dissertation provided an examination of visual privacy concerns within web browsers resulting in an initial model of incidental information privacy. Through our exploratory research, we developed guidelines for visual privacy management within web browsers. Our exploratory data also allowed us to examine the feasibility of three privacy management approaches. Given the current limitations of an automated approach, we pursued the more explicit approach suggested by the underlying privacy patterns we observed in web browsing. PrivateBits, a proof of concept privacy enhancing web browser, was developed as an instantiation of our design guidelines. Our initial evaluation showed that PrivateBits was effective at allowing users with varying privacy concerns and browsing behaviours to manage the privacy of their web browsing for their most regular viewing scenarios.

While our focus was on the incidental information found within web browser convenience features, our results are likely applicable for other personal information management systems. Personal information management is a growing research area and recent efforts to assist users in re-finding and managing the information they've encountered may also increase the chance that traces of previous activity are visible to others as well. Computers continue to become more ubiquitous in our personal lives as well as our work activities. We anticipate that visual privacy concerns of incidental information will increase as people continue to become more mobile with their devices, moving between various contexts of use. This dissertation has provided several contributions of a theoretical, practical, and methodological nature that may be of use to researchers and practitioners.

Bibliography

1. Cerberian Web Filter Categories. www.webrootdisp.net/audit/rating-descriptions.htm. Accessed September 10, 2004.

2. Internet use from Any Location by Individuals Age Three and Older. Online at: http://www.infoplease.com/ipa/A0901651.html. Accessed: Nov 2, 2004.

3. Platform for Privacy Preferences (P3P) Project. www.w3.org/P3P/.

4. WebRoot Software | Window Washer. www.webroot.com/consumer/products/windowwasher.

5. Yahoo! Directory. http://dir.yahoo.com.

6. ZoneAlarm Internet Security Suite Datasheet. http://download.zonelabs.com/bin/media/pdf/zaiss60_datasheet.pdf. 2005.

7. Ackerman, M. and Cranor, L. (1999). Privacy Critics: UI Components to Safeguard Users' Privacy. In *Proc. of CHI '99*, Pittsburgh, PA, 258-259.

8. Ackerman, M., Cranor, L. and Reagle, J. (1999). Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. In *Proc. of EC '99*, Denver, CO, 1-8.

9. Acquisti, A. (2004). Privacy in Electronic Commerce and the Economics of Immediate Gratification. In *Proc. of EC '04*, New York, New York, 21-29.

10. Acquisti, A. and Grossklags, J. (2005). Privacy and Rationality in Individual Decision Making. *IEEE Security & Privacy Magazine* 3(1): 26-33.

11. Adams, A. (2000). Multimedia information changes the whole privacy ballgame. In *Proc. of Proceedings of the conference on computers, freedom, and privacy (CFP 2000)*, Toronto, Canada, 25-32.

12. Aula, A., Jhaveri, N. and Kaki, M. (2005). Information Search and Re-access Strategies of Experienced Web Users. In *Proc. of WWW 2005*, Chiba, Japan, 583-592.

13. Bardram, J., Bunde-Pedersen, J. and Soegaard, M. (2006). Support for activity-based computing in a personal computing operating system. In *Proc. of CHI 2006*, Montreal, Quebec, Canada, 211-220.

14. Begole, J., Tang, J. C. and Hill, R. (2003). Rhythm Modeling, Visualizations and Applications. In *Proc. of UIST 2003*, Vancouver, Canada, 11-20.

15. Bellotti, V. and Sellen, A. (1993). Design for Privacy in Ubiquitous Computing Environments. In *Proc. of ECSCW '93*, Milano, Italy, 77-92.

16. Berry, L., Bartram, L. and Booth, K. S. (2005). Role-Based Policies to Control Shared Application Views. In *Proc. of UIST*, Seattle, WA, 23-32.

17. Blandford, A. and Stelmaszewska, H. (2001). Shooting the information rapids. In *Proc. of IHM-HCI2001, Vol. II*, 51-54.

18. Boardman, R. and Sasse, M. A. (2004). "Stuff Goes Into the Computer and Doesn't Come Out": A Cross-tool Study of Personal Information Management. In *Proc. of CHI '04*, Vienna, Austria, 583-590.

19. Boyle, M. and Greenberg, S. (2005). The Language of Privacy: Learning from Video Media Space Analysis and Design. *ACM Transactions on Computer-Human Interaction* 12(2): 328-370.

20. Butler, J. K. (1991). Toward Understanding and Measuring Conditions of Trust: Evolution of a Conditions of Trust Inventory. *Journal of Management* 17: 643-663.

21. Byrne, M., John, B., Wehrle, N. and Crow, D. (1999). The Tangled Web We Wove: A Taskonomy of WWW Use. In *Proc. of CHI '99*, Pittsburgh, PA, 544-551.

22. Cadiz, J. and Gupta, A. (2001). Privacy Interfaces for Collaboration. Microsoft Research, Redmond, WA. Technical Report No. MSR-TR-2001-82.

23. Camp, L. J., McGrath, C. and Genkina, A. (2006). Security and Morality: A Tale of User Deceit. In *Proc. of Models of Trust for the Web (MTW '06)*.

24. Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M. and Ouimet, J. A. (2003). College Student Responses to Web and Paper Surveys: Does Mode Matter? *Research in Higher Education* 44(1): 1-19.

25. Catledge, L. and Pitkow, J. (1995). Characterizing Browsing Strategies in the World-Wide Web. In *Proc. of WWW 1995*, Darmstadt, Germany, 1065 - 1073.

26. Chatterjee, P., Hoffman, D. L. and Novak, T. P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts. *Marketing Science* 22(4): 520-541.

27. Choo, C. W., Detlor, B. and Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday* 5(2): http://firstmonday.org/issues/issue5_2/choo/index.html.

28. Clark, L., Ting, I.-H., Kimble, C., Wright, P. and Kudenko, D. (2006). Combining ethnographic and clickstream data to identify user Web browsing strategies. *Information Research* 11(2): paper 249 [Available at http://InformationR.net/ir/211-242/paper249.html].

29. Clarke, R. (2002). Statement for Panel on Information Privacy in a Globally Networked Society: Implications for I.S. Research. ICIS 2002, http://www.anu.edu.au/people/Roger.Clarke/DV/ICIS2002.html.

30. ClickZStatsStaff (2002). Internet Usage Stats. Online at: www.clickz.com/stats/big_picture/traffic_patterns/article.php/960101.

31. Cockburn, A. and McKenzie, B. (2001). What do web users do? An empirical analysis of web use. *Int. J. Human-Computer Studies* 54: 903-922.

32. Cockburn, A., McKenzie, B. and JasonSmith, M. (2002). Pushing Back: Evaluating a New Behavior for the Back and Forward Buttons in Web Browsers. *International Journal of Human-Computer Studies* 57: 397-414.

33. Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J. and Powledge, P. (2005). Location Disclosure to Social Relations: Why, When, & What People Want to Share. In *Proc. of CHI '05*, Portland, Oregon, 81-90.

34. Consumer Reports (2005). Filtering software: Better, but still fallible. www.consumerreports.org/cro/electronics-computers/internet-filtering-software-605/overview.htm.

35. Cothey, V. (2002). A Longitudinal Study of World Wide Web Users' Information-Searching Behavior. *JASIST* 53(2): 67-78.

36. Curry, A. (2002). What are Public Library Users Viewing on the Internet?: An Analysis of the Transaction Logs of Burnaby, Brantford, Calgary, Winnipeg, and Halifax Public Libraries., National Library and Archives Canada Virtual Collection of Monographs and Periodicals, http://tinyurl.com/8v7qc.

37. Cutrell, E., Robbins, D. C., Dumais, S. T. and Saran, R. (2006). Fast, Flexible Filtering with Phlat - Personal Search and Organization Made Easy. In *Proc. of CHI 2005*, Montreal, Quebec, Canada, 261-270.

38. Cvrcek, D., Kumpost, M., Matyas, V. and Danezis, G. (2006). A Study on the Value of Location Privacy. In *Proc. of WPES'06*, Alexandria, Virginia, USA, 109-118.

39. de Paula, R., Ding, X., Dourish, P., Nies, K., Pillet, B., Redmiles, D., Ren, J., Rode, J. and Filho, R. S. (2005). Two Experiences Designing for Effective Security. In *Proc. of Symposium On Usable Privacy and Security (SOUPS)*, Pittsburgh, PA, 25-34.

40. DePaulo, B. M., Wetzel, C., Sternglanz, R. W. and Wilson, M. J. W. (2003). Verbal and Nonverbal Dynamics of Privacy, Secrecy, and Deceit. *J. Of Social Issues* 59(2): 391-410.

41. Dillman, D. A. (2000). Mail and web-based survey: the tailored design method. New York, John Wiley & Sons.

42. Dourish, P., Grinter, R. E., Delgado de la Flor, J. and Joseph, M. (2004). Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing* 8: 391-401.

43. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R. and Robbins, D. (2003). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In *Proc. of SIGIR 2003*, Toronto, Canada, 72-79.

44. Edmonds, K. A. A., Hawkey, K., Kellar, M. and Turnbull, D. (2006). Logging Traces of Web Activity: The Mechanics of Data Collection. Workshop held at WWW 2006, May 23, 2006.

45. Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research* 20: 303-315.

46. Flanagin, A. J. and Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society* 9(2): 319-342.

47. Fowler, F. J., Jr. (1995). Improving Survey Questions: Design and Evaluation. Thousand Oaks, CA, Sage Publications.

48. Garson, G. D. (2007). Cluster Analysis. http://www2.chass.ncsu.edu/garson/PA765/cluster.htm.

49. Goffman, E. (1959). The Presentation of Self in Everyday Life. Garden City, New York, Doubleday Anchor Books.

50. Google (2004). Google Corporate Information: Timeline. Online at: http://www.google.ca/corporate/timeline.html.

51. Grace-Martin, M. and Gay, G. (2001). Web Browsing, Mobile Computing and Academic Performance. *Educational Technology & Society* 4(3): 95-107.

52. Greenberg, S. (2001). The Notification Collage: Posting Information to Public and Personal Displays. In *Proc. of CHI 2001*, Seattle, WA, 514-521.

53. GVUOnlineSurvey (1997). GVU's 8th WWW User Survey. Online at: http://www.cc.gatech.edu/gvu/user_surveys/survey-1997-10.

54. Gwizdka, J. (2004). Email Task Management Styles: The Cleaners and the Keepers. In *Proc. of CHI 2004*, Vienna, Austria, 1235-1238.

55. Hann, I.-H., Hui, K.-L., Lee, T. S. and Png, I. P. L. (2002). Online Information Privacy: Measuring the Cost-Benefit Trade-Off. In *Proc. of 23rd Int. Conf. on Information Systems*, 1-10.

56. Hawkey, K. (2005). Privacy Management of Incidental Information During Collaboration: Data Analysis and Evaluation Challenges. Workshop on Usage Analysis: Combining Logging and Qualitative Methods, CHI '05: 1-4.

57. Hawkey, K. (2006). Mission Impossible? Capturing Rich Yet Natural User Behaviour on the Web. Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, WWW 2006.

58. Hawkey, K. (2006). Privacy Research: A Mixed Methodology Approach. In *Proc. of Workshop on Privacy and HCI: Methodologies for Studying Privacy Issues, CHI '06*, Montreal, PQ.

59. Hawkey, K. and Inkpen, K. (2005). Privacy Gradients: Exploring ways to manage incidental information during co-located collaboration. Ext. Abstracts CHI 2005, ACM Press: 1431-1434.

60. Hawkey, K. and Inkpen, K. (2005). Web Browsing Today: The impact of changing contexts on user activity. Ext. Abstracts CHI 2005. Portland, Oregon, ACM Press: 1443-1446.

61. Hawkey, K. and Inkpen, K. M. (2006). Examining the Content and Privacy of Web Browsing Incidental Information In *Proc. of WWW 2006*, Edinburgh, Scotland, 123-132.

62. Hawkey, K. and Inkpen, K. M. (2006). Incidental Information Privacy and PIM. Presentation at Personal Information Management: Now that we're talking, what are we learning? (PIM 2006), SIGIR 2006 2-Day Workshop.: 67-70.

63. Hawkey, K. and Inkpen, K. M. (2006). Keeping Up Appearances: Understanding the Dimensions of Incidental Information Privacy. In *Proc. of CHI 2006*, Montreal, Quebec, Canada, 821-830.

64. Hawkey, K. and Inkpen, K. M. (2006). PrivateBits: Managing Visual Privacy within Web Browsers. Poster presentation at SOUPS 2006.

65. Hawkey, K. and Inkpen, K. M. (2006). PrivateBits: Managing Visual Privacy within Web Browsers. Demonstration at CSCW 2006.

66. Hawkey, K. and Inkpen, K. M. (2007). PrivateBits: Managing Visual Privacy within Web Browsers. In *Proc. of Graphics Interface 2007 (to appear)*, Montreal, Quebec.

67. Hawkey, K. and Kellar, M. (2004). Recommendations for reporting context in studies of web browsing behaviour. Dalhousie University, Halifax, NS. Technical Report No. CS-2004-16.

68. Herder, E. and Juvina, I. (2004). Discovery of Individual User Navigation Styles. In *Proc. of Workshop on Individual Differences, AH2004*.

69. Hilbert, D. and Redmiles, D. (2000). Extracting Usability Information from User Interface Events. *ACM Computing Surveys* 32(4): 384-421.

70. Holscher, C. and Strube, G. (2000). Web Search Behavior of Internet Experts and Newbies. In *Proc. of WWW 2000*, Amsterdam, The Netherlands, 337-346.

71. Huang, E. M. and Mynatt, E. D. (2003). Semi-Public Displays for Small, Co-located Groups. In *Proc. of CHI '03*, Ft. Lauderdale, FL, 49-56.

72. Hunter, C. D. (2000). Social Impacts: Internet Filter Effectiveness Testing: Over- and Underinclusive Blocking Decisions of Four Popular Web Filters. *Social Science Computer Review* 18(2): 214-222.

73. Hutchings, H. M. and Pierce, J. S. (2006). Understanding the whethers, hows, and whys of divisible interfaces. In *Proc. of AVI 2006*, Venezia, Italy, 274-277.

74. Infoplease (2000-2004). Internet Timeline. Online at: http://www.infoplease.com/ipa/A0908398.html.

75. Jackson, L. A., Eye, A. v., Barbatsis, G., Biocca, F., Zhao, Y. and Fitzgerald, H. E. (2003). Internet attitudes and Internet use: some surprising findings from the HomeNetToo project. *Int. J. Human-Computer Studies* 59: 355-382.

76. JasonSmith, M. and Cockburn, A. (2003). Get a Way Back: Evaluating Retrieval from History Lists. In *Proc. of AUIC2003*, Adelaide, Australia, 33-38.

77. Jensen, C., Potts, C. and Jensen, C. (2005). Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies* 63: 203-227.

78. Joinson, A. N., Paine, C., Reips, U.-D. and Buchanan, T. (2006). Privacy and Trust: The role of situational and dispositional variables in online disclosure. In *Proc. of Privacy, Trust, and Identity Issues for Ambient Intelligence Workshop, Pervasive 2006*, Dublin, Ireland, 1-6.

79. Jones, W. and Bruce, H. (2005). A Report on the NSF-Sponsored Workshop on Personal Information Management, Seattle, WA. Technical Report No.

80. Jones, W., Bruce, H. and Dumais, S. (2001). Keeping Found Things Found on the Web. In *Proc. of Proceedings of the 2001 ACM CIKM International Conference on Information and Knowledge Management (CIKM 2001)*, Atlanta, GA, 119-126.

81. Kaasten, S. and Greenberg, S. (2001). Integrating Back, History and Bookmarks in Web Browsers. In *Proc. of CHI 2001*, Seattle, WA, 379-380.

82. Kaasten, S., Greenberg, S. and Edwards, C. (2002). How People Recognize Previously Seen WWW Pages from Titles, URLs and Thumbnails. In *Proc. of Human Computer Interaction 2002*, 247-265.

83. Kandogan, E. and Shneiderman, B. (1997). Elastic Windows: A Hierarchical Multi-Window World-Wide Web Browser. In *Proc. of UIST*, Banff, AB, 169-177.

84. Kehoe, C. M., Pitkow, J., Sutton, K., Aggarwal, G. and Rogers, J. D. (1999). Results of GVU's Tenth World Wide Web User Survey, Online at: http://www.cc.gtech.edu/gvu/user_surveys/survey-1998-10/tenthreport.html.

85. Kellar, M., Hawkey, K., Inkpen, K. M. and Watters, C. (In Press). Challenges of Capturing Natural Web-based User Behaviours. *In Use, In Situ: Extending Field Research Methods, Special issue of the International Journal of Human Computer Interaction. (Accepted April 12, 2006).*

86. Kellar, M., Watters, C. and Shepherd, M. (2006). The Impact of Task on the Usage of Web Browser Navigation Tools. In *Proc. of Graphics Interface*, Quebec City, Canada, 235-242.

87. Kellar, M., Watters, C. and Shepherd, M. (In Press). A Field Study Characterizing Web-based Information Seeking Tasks. *Journal of the American Society for Information Science and Technology*. Currently available as Dalhousie Computer Science Technical Report CS-2005-20.

88. Kelly, D. and Belkin, N. (2004). Display Time as Implicit Feedback: Understanding Task Effects. In *Proc. of SIGIR 2004*, Sheffield, UK, 377-384.

89. Kerner, S. M. (2004). More Broadband Usage Means More Online Spending. Online at: www.clickz.com/stats/markets/broadband/article.php/3419281.

90. Lau, T., Etzioni, O. and Weld, D. S. (1999). Privacy Interfaces for Information Management. *Communications of the ACM* 42(10): 89-94.

91. Lederer, S., Hong, J. I., Dey, A. K. and Landay, J. A. (2004). Personal privacy through understanding and action: five pitfalls for designers. *Pers Ubiquit Comput* 8: 440-454.

92. Lederer, S., Mankoff, J. and Dey, A. K. (2003). Towards a Deconstruction of the Privacy Space. Workshop on Ubicomp Communities: Privacy as Boundary Negotiation, UBICOMP 2003, http://guir.berkeley.edu/pubs/ubicomp2003/privacyworkshop/papers/lederer-privacyspace.pdf

93. Lederer, S., Mankoff, J. and Dey, A. K. (2003). Who Wants to Know What When? Privacy Preference Determinants in Ubiquitous Computing. Ext. Abstracts CHI 2003. Ft. Lauderdale, Florida, USA, ACM Press: 724-725.

94. Lee-Tiernan, S., Farnham, S. and Cheng, L. (2003). Two Methods for Auto-Organizing Personal Web History, Ext. Abstracts of CHI 2003: 814-815.

95. Liu, C., Marchewka, J. T., Lu, J. and Yu, C.-S. (2004). Beyond concern: a privacy-trust-behavioral intention model of electronic commerce. *Information & Management* 42: 127-142.

96. Loeber, S. C. and Cristea, A. (2003). A WWW Information Seeking Process Model. *Educational Technology & Society* 6(3): 43-52.

97. Lycos (1999). The Lycos 50 Daily Report. Online at: http://50.lycos.com/083099.html.

98. Malhotra, N. K., Kim, S. S. and Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15(4): 336-355.

99. Margulis, S. T. (2003). Privacy as a Social Issue and Behavioral Concept. *J. Of Social Issues* 2003(59): 2.

100. Marsh, D. (2003). History of the Internet. Online at: http://www.internetvalley.com/archives/mirrors/davemarsh-timeline-1.htm.

101. Marx, G. T. (2003). A Tack in the Shoe: Neutralizing and Resisting the New Surveillance. *J. Of Social Issues* 59(2): 369-390.

102. McCullagh, D. (2006). AOL's disturbing glimpse into user's lives. CNET News.com. http://news.com.com/2100-1030_3-6103098.html.

103. McGrath, J. E. (1995). Methodology matters: doing research in the behavioral and social sciences. Human-computer interaction: toward the year 2000. J. G. R. Baeker, W. Buxton, and S. Greenberg: 152-169.

104. McKnight, D. H., Choudhury, V. and Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research* 13(3): 334-361.

105. Metzger, M. J., Flanagin, A. J. and Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education* 41: 271-290.

106. Milic-Frayling, N., Jones, R., Rodden, K., Smyth, G., Blackwell, A. and Sommerer, R. (2004). SmartBack: Supporting Users in Back Navigation. In *Proc. of WWW 2004*, New York, NY, 63-71.

107. Moor, J. H. (1997). Towards a theory of privacy in the information age. *ACM SIGCAS Computers and Society* 27(3): 27-32.

108. Nie, N. H. and Erbring, L. (2000). Internet and Society: A Preliminary Report, Stanford Institute for the Quantitative Study of Society, http://www.stanford.edu/group/siqss/Press_Release/Preliminary_Report.pdf.

109. Nielsen, J. (1993). Usability Engineering, Elsevier Science & Technology Books.

110. Nielsen/NetRatings United States: Average Web Usage, Month of September 2004, Home Panel. Online at: www.nielsen-netratings.com.

111. Nielsen/NetRatings United States: Average Web Usage, Month of September 2004, Work Panel. Online at: www.nielsen-netratings.com.

112. Novak, T. P. and Hoffman, D. L. (1997). Measuring the flow experience among web users. *Interval Research Corporation* 31.

113. O'Neil, D. (2001). Analysis of Internet Users' Level of Online Privacy Concerns. *Social Science Computer Review* 19(1): 17-31.

114. Olson-Buchanan, J. B. and Boswell, W. R. (2006). Blurring boundaries: Correlates of integration and segmentation between work and nonwork. *Journal of Vocational Behavior* 68: 432-445.

115. Olson, J. S., Grudin, J. and Horvitz, E. (2005). A Study of Preferences for Sharing and Privacy. Ext. Abstracts of CHI '05. Portland, Oregon, ACM Press: 1985-1988.

116. P&AB (2003). Consumer Privacy Attitudes: A Major Shift Since 2000 and Why. *Privacy & American Business Newsletter* 10(6): 1,3-5.

117. Paine, C., Joinson, A. N., Buchanan, T. and Reips, U.-D. (2006). Privacy and self-disclosure online. In *Proc. of CHI '06 (Ext. Abstracts)*, Montreal, Quebec, Canada, 1187-1192.

118. Palen, L. (1999). Social, individual and technological issues for groupware calendar systems. In *Proc. of Proceedings of the CHI'99 conference on human factors in computing systems*, Pittsburgh, PA, 17-24.

119. Palen, L. and Dourish, P. (2003). Unpacking "Privacy" for a Networked World. In *Proc. of CHI '03*, Ft. Lauderdale, FL, 129-136.

120. Pallant, J. (2005). SPSS Survival Manual. Berkshire, UK, Open University Press.

121. Pastore, M. (1998). Microsoft Leads Browser Race. Online at: www.clickz.com/stats/big_picture/hardware/article.php/151351.

122. Pastore, M. (1998). Online Users Need Speed. Online at: www.clickz.com/stats/markets/broadband/article.php/151701.

123. Pastore, M. (2000). E-Commerce, Mobile Access Drawing Interest from Net Users. Online at: www.clickz.com/stats/big_picture/geographics/article.php/5911_494701.

124. Pastore, M. (2000). Internet Usage Stats. Online at: www.clickz.com/stats/big_picture/traffic_patterns/article.php/291211.

125. Pastore, M. (2000). Slow Modems Still Dominate Home Internet Scene. Online at: www.clickz.com/stats/big_picture/hardware/article.php/277191.

126. Pastore, M. (2001). Online Consumers Now the Average Consumer. Online at: www.clickz.com/stats/big_picture/demographics/article.php/5901_800201.

127. Patil, S. and Kobsa, A. (2005). Uncovering Privacy Attitudes and Practices in Instant Messaging. In *Proc. of GROUP '05*, Sanibel Island, Florida, USA, 109-112.

128. Patil, S. and Lai, J. (2005). Who Gets to Know What When: Configuring Privacy Permissions in an Awareness Application. In *Proc. of CHI '05*, Portland, Oregon, 101-110.

129. Phillips, D. J. (2002). Context, identity, and privacy in ubiquitous computing environments. In *Proc. of Workshop on socially-informed design of privacy-enhancing solutions, 3rd international confererence on ubiquitous computing (Ubicomp 2002).* Goteborg, Sweden.

130. Phillips, D. J. (2004). Privacy policy and PETS: The influence of policy regimes on the development and social implications of privacy enhancing technologies. *new media and society* 6(6): 691-706.

131. Pirolli, P., Pitkow, J. and Rao, R. (1996). Silk from a Sow's Ear: Extracting Usable Structures from the Web. In *Proc. of CHI '96*, Vancouver, Canada, 118 - 125.

132. Pitkow, J. and Kehoe, C. M. (1996). Emerging Trends in the WWW User Population. *Communications of the ACM* 39(6): 106-108.

133. Pitkow, J. and Recker, M. M. (1994). Using the Web as a Survey Tool: Results from the Second WWW Survey. *Computer Networks and ISDN Systems* 27(6): 809-822.

134. Reilly, D., Dearman, D., Ha, V., Smith, I. and Inkpen, K. (2006). "Need to Know": Examining Information Need in Location Disclosure. In *Proc. of Pervasive 2006*, Dublin, Ireland, 33-49.

135. Safari 2.0 Help (2007). Protecting private information on shared computers, http://docs.info.apple.com/article.html?path=Safari/2.0/en/ibr1069.html.

136. Sellen, A. J., Murphy, R. and Shaw, K. L. (2002). How Knowledge Workers Use the Web. In *Proc. of CHI '02*, Minneapolis, MN, 227-234.

137. Sheehan, K. B. (2002). Toward a Typology of Internet Users and Online Privacy Concerns. *The Information Society* 18: 21-32.

138. Shoemaker, G. B. D. and Inkpen, K. M. (2001). Single Display Privacyware: Augmenting Public Displays with Private Information. In *Proc. of CHI '01*, Seattle, WA, 522-529.

139. Spiekermann, S., Grossklags, J. and Berendt, B. (2001). E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior. In *Proc. of EC '01*, Tampa, Florida, USA, 38-47.

140. Statistics Canada (2003). Household Internet use at home by Internet activity. http://www40.statcan.ca/l01/cst01/comm09a.htm, CANSIM.

141. Statistics Canada (2003). Household Internet use by location of access. http://www40.statcan.ca/l01/cst01/comm12a.htm, CANSIM.

142. Statistics Canada (2005). Characteristics of individuals using the Internet. http://www40.statcan.ca/101/cst01/comm15.htm, CANSIM.

143. Tan, D. S. and Czerwinski, M. (2003). Information Voyeurism: Social Impact of Physically Large Displays on Information Privacy. In *Proc. of CHI 2003*, Ft. Lauderdale, FL, 748-749.

144. Tarasewich, P., Gong, J. and Conlan, R. (2006). Protecting private data in public. In *Proc. of CHI 2006 (Ext. Abstracts)*, Montreal, Canada, 1409-1414.

145. Tauscher, L. and Greenberg, S. (1997). How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *Int. J. Human-Computer Studies* 47: 97-137.

146. Tauscher, L. and Greenberg, S. (1997). Revisitation patterns in World Wide Web navigation. In *Proc. of CHI '97*, Atlanta, GA.

147. Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H. and Sonenstein, F. L. (1998). Adolescent sexual behaviour, drug use, & violence: Increased reporting with computer survey technology. *Science* 280: 867-873.

148. Weinreich, H., Obendorf, H., Herder, E. and Mayer, M. (2006). Off the Beaten tracks: Exploring Three Aspects of Web Navigation. In *Proc. of World Wide Web Conference 2006 (WWW 2006)*, Edinburgh, Scotland, 133-142.

149. Weisband, S. P. and Reinig, B. A. (1995). Managing User Perceptions of Email Privacy. *Communications of the ACM* 38(12): 40-47.

150. Westin, A. F. (2003). Social and Political Dimensions of Privacy. *Journal of Social Issues* 59(2): 431-453.

151. Westin, A. F. and Maurici, D. (1998). E-Commerce and Privacy: What Net Users Want, Privacy and American Business.

152. Whalen, T. and Inkpen, K. M. (2005). Gathering evidence: use of visual security cues in web browsers. In *Proc. of Graphic Interface 2005*, Victoria, British Columbia, 137-145.

153. Whittaker, S. and Sidner, C. (1996). Email overload: exploring personal management of email. In *Proc. of CHI '96*, 276-283.

154. Whittaker, S., Terveen, L. and Nardi, B. A. (2000). Let's Stop Pushing the Envelope and Start Addressing It: A Reference Task Agenda for HCI. *Human Computer Interaction* 15: 75-106.

155. Whitten, I. H. and Frank, E. (2000). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco, CA, Academic Press.

Appendix A: The Evolving Web Browsing Environment

The World Wide Web is relatively young and has been continually evolving since its inception in 1991. Technological innovations have changed the way that people access the Web: as the state of hardware (Unix boxes, desktop PCs, laptop computers, handheld, cellular phones) and software (web browsers and search engines) progresses, the experience of the end user changes. With penetration of the market expanding from technologists to general home and business users, the World Wide Web's user population has become more diverse. Internet access is no longer restricted to those with a high income and level of education. Web browsing behaviours have changed over time as a result of the changing technological environment and user population. This temporal context is important to consider when interpreting the applicability of the seminal research to the web browsing environment of today.

In this appendix (an excerpt from [67]), we present snapshots of the state of the World Wide Web and its users at the times the seminal research about web browsing behaviours was conducted. This timeline was created in the fall of 2004. We first give a brief description of each of the seminal works.

Seminal Works

One of the first studies examining user behaviour on the Web was conducted by Catledge and Pitkow [25] for three weeks in 1994. Participant behaviour was logged for three weeks while they browsed the Web using a modified version of XMosaic that collected browsing activity. Two dominant methods of navigation were revealed by the participants: hyperlinks and the back button. Navigation strategies were categorized according to frequency.

Pirolli, Pitkow and Rao [131] used trace logs of web usage from March through May of 1995, along with topology and textual similarity between nodes, to extract structures of websites. This work was one of the early applications of web usage logs.

Similar to Catledge & Pitkow, Tauscher and Greenberg [145] observed user behaviour with a modified version of XMosaic, in order to study revisitation patterns of

users. Over a six week period in 1995, they observed that 58% of page visits were revisits and the back button was used in 30% of navigations.

Byrne et al. [21] conducted a task analysis of user web behaviour through a 1998 study. Participants were video taped in their offices, for a day, as they used the Web. Participants spent the majority of the time on the Web reading and the most common navigation method was hyperlinks, followed by the back button.

Choo, Detlor, and Turnbull [27] investigated information seeking behaviour on the Web in a two week study conducted circa 1998. Participants' web behaviour in the workplace was logged client-side during the course of the study. Through the analysis of user's clickstream data, interviews and questionnaires, four modes of information seeking behaviour were defined.

Cockburn and McKenzie [31] conducted a four month retrospective observational study, from October 1999 to January 2000, of history and bookmark files retrieved from server backups. The authors found an average revisitation rate of 81%. Analysis of the bookmark files found that participants were either heavy or light users of bookmarks.

Sellen, Murphy & Shaw [136] studied the activities and characteristics of knowledge workers on the Web. Participants were interviewed circa 2001 in front of their history lists and described the web activities they had recently completed. Knowledge workers engaged in six types of activity on the Web: finding, information gathering, browsing, transacting, communicating, and housekeeping..

Timeline of Web Browsing Environmental Changes

We next present snapshots of the changing web browsing environment. It must be noted that the figures reported have been selected from a variety of sources with varying methodologies, populations, and metrics. Therefore, direct comparisons are not always appropriate. These snapshots have been provided to illustrate the changing nature of user behaviour on the Web that gives the context for the seminal papers in the area. In each of the snapshots, we have indicated the dates of the studies appearing in the seminal papers. If a study date was not available, we note the likely date based upon the submission deadlines for the publication.

In the Beginning

- Catledge and Pitkow [25]: August 1994
- Pirolli, Pitkow, and Rao [131]: March-May 1995
- Tauscher and Greenberg [145]: October-December 1995

Fall/94: The typical user is a 31 year-old educated male who works with computers and has authored about 30 web documents [133]. He uses a Mosaic browser 1-4 times a day for about 5 hours per week [133]. Netscape has just been released [133]. He uses the Web to browse, for entertainment, for work or business, and for research [133]. He has a choice of about 10,000 websites [100]

Fall/95 [132]: Worldwide web traffic has surpassed ftp data and search engines are now available [100]. Users are shifting towards "early adopters/seekers of technology" instead of the "technology developers/pioneers" of a year before [132] with the start of commercial internet providers such as Compuserve, AOL and Prodigy [74]. Women now account for about 30% of web users and there has been some increase in the number of younger and older users [132]. Most users have 14.4 or 28.8 kbs modems [132].

Home Users and Browser Wars

- Byrne, John, Wehrle, and Crow [21]: circa 1998
- Choo, Detlor, and Turnbull [27]: circa 1998

Fall/98: Women now account for almost 40% of web users [53]. About a third of users have a 56K modem [84] and 84% are interested in high speed internet access [122]. Microsoft IE wins the browser wars, just surpassing last year's dominant browser, Netscape Navigator, to capture 50% of the market [121]. More than 40% of the people between the ages of 9 and 49 now have on-line access [2]; their average age is 38 [84]. Almost a third of users shop on-line [123]. Google arrives 10,000 searches are performed per day [50].

Work and Home: The Need for Speed

- Cockburn and McKenzie [31]: Oct. 1999-Jan. 2000

Fall/99: The year 2000 is looming and the 150 million web users [74] worldwide are looking for information about Y2K as the Lycos 50 listing of the top searches debuts (although Pokemon and the Blair Witch Project top the list) [97]. Google performs 3 million searches per day [50]. Napster allows swapping of music and 'E-Commerce' is the new buzz word [74]. The 6% of users with high speed internet access view 130% more pages and surf the Web 83% more often than the 45% of users that still have a 28.8/33.6 K modem [125]. According to Nielsen//NetRatings the average web user had 17 x 29-minute sessions per month, viewing an average of 32 pages per session [124].

In the Mainstream: Just Google it

- Sellen, Murphy and Shaw [136]: circa 2001

Fall/01: Google has become a verb: with over 3 billion web documents [50] available to be searched and the Google toolbar to help them do it, users over the world are telling each other to Google it. Napster has lost its court case [74] but other file sharing applications are quick to fill the void. The demographic structure of the population on-line is much closer to that from census data than in previous years [126]. There is an equal split of male/female users, but household incomes for web users are still higher than for the general population ($49, 800 vs. $40, 800) and the web user population is still younger (75% of adults 18-49 are on-line vs. 63% of the population, 24% of adults 50+ are on-line vs. 37% of population) [126]. Our average web user now has 33 x 33-minute sessions per month, viewing an average of 36 pages per session [30]. 72% of the population are now using the Internet (58% at home, 73% at school, 51% at work) [89].

A Daily Tool:

Fall/04: The Internet has become a daily tool: 56% of those with access to the Internet go on-line daily, 48% send email, 27% get news, and 19% do research for a job. Google has added Gmail and Desktop Search [50] and the division between on-line and off-line blurs. Our average web user now has 31 web sessions per month at home during the

almost 26 hours of PC use [110] and 65 sessions at work during the 76 hours of work PC use [111].

Summary: Importance of Temporal Context

As we have presented, the state of the Web has changed quickly and drastically since its inception. It is important that seminal works are acknowledged, but given the ever-changing state of the Web, there is a concern that data that is no longer relevant is being used to support current research. Care must be taken to ensure the context in which the data was recorded does not differ significantly from the current context with respect to the aspects of web browsing behaviour under study. These seminal works do however provide us with a baseline from which we can measure the changes in user behaviour through the evolution of the Web.

One example of changing user patterns can be shown with research about the Back button. Catledge & Pitkow [25] reported the Back button was used in 41% of all navigation, while one year later Tauscher [145] reported the Back button was used in only 30% of all navigation. In the two studies reported in the Smartback paper [106] (dates unknown, approximately 12 months apart, and published in 2004), back button usage was down to 22% (exploratory study) and approximately 8% (back button and Smartback button equivalent, evaluation study) of all navigation. However, each of these studies had a relatively small number of participants and there may be individual differences or population differences that account for the decrease in usage in addition to the increased in navigation aids such as auto-complete of URLs and enhanced History and Favorites features.

This does not imply that all results from different contexts are not relevant, but the relevance has to be challenged by evaluating the context of the state of the Web, the web browsing environment, and the characteristics of the user population studied. There are aspects of web browsing behaviour that may be relatively stable. For example, during the construction of the timeline it was noted that the page view time has remained fairly constant at about 55-60 seconds per page, with no large variations in the 1999-2004 monthly reports from Nielsen//Net ratings.

Appendix B: IIP Survey Questionnaire

Survey of Privacy Issues Arising From Collaboration Around a Computer

Submission of the completed the survey will indicate that you have read the explanation about this study, have been given the opportunity to discuss it and questions have been answered to your satisfaction, and that you consent to take part in the study. Your participation is voluntary and you are free to withdraw from the study at any time.

If you have any questions, please contact the Principal Investigator:

Mrs. Kirstie Hawkey
PhD Candidate
Faculty of Computer Science
Dalhousie University
Email: hawkey@cs.dal.ca
Cell: 452-4675

Note: The survey will only work properly if your browser/proxy server/firewall is not blocking the HTTP REFERER and allows META REFRESH. If you do not see the screening questions after entering your PIN or can not progress past page 1 of the survey, please check your settings and contact Kirstie Hawkey (hawkey.dal.ca, 452-4675) for a new PIN number.

Please Enter Your PIN: _____

Begin Survey

Survey of Privacy Issues Arising From Collaboration Around a Computer

- What is the primary location of your web browsing activity?
  - ○ At work
  - ○ At school
  - ○ At home
- What is the primary computer that you use in this location?
  - ○ A laptop/notebook computer
  - ○ An assigned or personal computer
  - ○ A shared computer

Continue

Survey of Privacy Issues Arising From Collaboration Around a Computer

01. What is your sex? ○ Male ○ Female

02. What is your age (in years)? _____

03. What is your highest level of completed education?
○ Less than high school
○ Completed high school
○ Completed technical school (specify field)_____
○ Some university (specify field)_____
○ Completed university degree (specify field)_____
○ Some graduate work (specify field)_____
○ A graduate degree (specify field)_____

04. What is your occupation? _____

05. How many years have you been in this occupation? (if student, leave blank) _____ (in years)

06. How many years have you been using a computer on a regular basis? _____ (in years)

07. How many hours in an average week do you use a computer?
○ < 7 hours
○ 8-14 hours
○ 15-21 hours
○ 22-28 hours
○ 29-35 hours
○ 36+ hours

08. How many hours in an average week do you use a web browser?

- ○ < 7 hours
- ○ 8-14 hours
- ○ 15-21 hours
- ○ 22-28 hours
- ○ 29-35 hours
- ○ 36+ hours

Percentage of that time spent for personal reasons? _____

Percentage of that time spent for work-related reasons? _____

Percentage of that time spent for educational reasons? _____

09. When using a web browser, where is your computer located? For example, if you search for entertainment information, do you use your home, work, or school computer? For each type of web browser activity, please check all settings that apply (check Never if not applicable)

Reason	Setting (check all that apply)			
	Home	Work	School	Never
Searching for medical information	☐	☐	☐	☐
Searching for entertainment information	☐	☐	☐	☐
Searching for erotic material	☐	☐	☐	☐
Visiting personal improvement support forums	☐	☐	☐	☐
Visiting technical support forums	☐	☐	☐	☐
Banking on-line	☐	☐	☐	☐
Shopping on-line	☐	☐	☐	☐
Playing games on-line	☐	☐	☐	☐
Email	☐	☐	☐	☐
Other regular activity (specify)_____	☐	☐	☐	☐

Next Page

Survey of Privacy Issues Arising From Collaboration Around a Computer

Please think for a moment only about using your computer at home for web browsing and then answer the following questions. If you have no computer at home, skip to question 12

10. On average, how many hours per week do you use a web browser at home?
- ○ < 7 hours
- ○ 8-14 hours
- ○ 15-21 hours
- ○ 22-28 hours
- ○ 29-35 hours
- ○ 36+ hours

11. What computer do you use at home when using a web browser? (check all that apply)
- ☐ Desktop computer that only you use
- ☐ Desktop computer shared with family members
  Do you have separate login accounts? ○ Yes ○ No
- ☐ Laptop computer that only you use
- ☐ Other (please specify):_____

Please think for a moment only about using your computer at work or school for web browsing and then answer the following questions. If you do not work or go to school, or don't use a computer at work or school, skip to question 14 (next page)

12. On average, how many hours per week do you use a web browser at work/school?(check next to answer)
- ○ < 7 hours
- ○ 8-14 hours
- ○ 15-21 hours
- ○ 22-28 hours
- ○ 29-35 hours
- ○ 36+ hours

13. What computer do you use at work/school when using the web browser? (check all that apply)

☐ Desktop computer that only you use

☐ Desktop computer shared by members of a small group

Do you have separate login accounts? ○ Yes ○ No

☐ Desktop computer for general use

Do you have separate login accounts? ○ Yes ○ No

☐ Laptop computer that only you use

☐ Other (please specify): _____

Several of the questions on the next pages will ask you to think about how comfortable some situation makes you feel in terms of privacy. When answering those questions, please use the following scale to indicate your comfort level. This scale will be included in any questions that ask you to use the privacy comfort scale.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Next Page

Survey of Privacy Issues Arising From Collaboration Around a Computer

14. Please think about when you are using computers at home, work, and/or school. On an average day, please think about *who can clearly **see** the contents of your screen as you are using it* (this could be somebody working with you at your computer or sitting close by and able to see your entire screen), approximately *how often* they may be in this situation, and, in general, *how comfortable this makes you feel in terms of your privacy*. Select information for all that apply and use the privacy comfort scale when rating your comfort level.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Potential Viewer	How often they may see the screen					Your comfort in terms of privacy (use privacy comfort scale)						
Parents	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Spouse / Significant Other	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Close friends	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Acquaintances	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Colleagues / fellow students	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Clients	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Supervisor	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Employees / Students (that you supervise)	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Audience at a presentation you give	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Technical support staff	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7

15. Now think about *who may **use** your computer after you use it* at home, work, and/or school, approximately *how often they may use it*, and, in general, how *comfortable this makes you feel in terms of your privacy*. Select information for all that apply and use the privacy comfort scale when rating your comfort level.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Potential User	**How often they may use the computer**					**Your comfort in terms of privacy (use privacy comfort scale)**						
Parents	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Spouse / Significant Other	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Close friends	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Acquaintances	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Colleagues / fellow students	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Clients	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Supervisor	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Employees / Students (that you supervise)	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Technical support staff	○ Daily	○ Weekly	○ Monthly	○ Rarely	○ Never	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7

Next Page

Survey of Privacy Issues Arising From Collaboration Around a Computer

The next set of questions concern laptop use. If you do not use a laptop, skip to question 20

16. Do you use your laptop for web browsing in multiple locations?
- ○ Yes
- ○ No

17. Do you use a laptop in addition to other computers?
- ○ Yes
- ○ No (if no, skip to question 20:)

18. Please specify where the other machines are located (check all that apply)
- ☐ At home
- ☐ At work/school
- ☐ Other (specify) _____

19. Which machine(s) do you use when doing web browsing of a private/personal nature? (check all that apply):
- ☐ Laptop
- ☐ Work/school computer
- ☐ Home computer
- ☐ Other (specify) _____

The next set of questions concerns use of publicly accessible computers (such as those found at libraries or Community Access (C) sites). If you do not use such computers, skip to question 24

20. Do you use multiple public access sites for web browsing?
- ○ Yes
- ○ No

21. Do you use other computers in addition to those you use at public access sites?
- ○ Yes
- ○ No (if no, skip to question 24)

22. Please specify where the other machines are located (check all that apply):
- ☐ At home
- ☐ At work/school
- ☐ Other (specify) _____

23. Which machine(s) do you use when doing web browsing of a private/personal nature? (check all that apply):
- ☐ Laptop
- ☐ Work/school computer
- ☐ Home computer
- ☐ Other (specify)_____

Next Page

Survey of Privacy Issues Arising From Collaboration Around a Computer

24. Please read the following scenario carefully and think about the situation. Then answer the following questions.

Your car is starting to make funny noises so you have decided to start looking for a new model. You use your web browser to search for such topics as "low gas mileage" and "4-door sedan" and have visited such web pages as www.honda.com and www.carcomparisons.com (which you add to your favorites for future reference).

Later, you are in a situation where you must use your web browser with others gathered in front of your computer screen as you collaborate on a project. There is a chance that the URL's to the pages you've been viewing recently and the search terms you've been using will be displayed as you use the browser for the task at hand. Please indicate how you would feel in this situation, considering *who is the person viewing and who is in control of the web browser* using the privacy comfort scale to rate your comfort level.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Person viewing	Person in control of the web browser	Your comfort in terms of privacy (use privacy comfort scale)
Close friend	You	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Close friend, with you sitting right there	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Close friend, and you leave the room	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
Supervisor	You	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Supervisor, with you sitting right there	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Supervisor, and you leave the room	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
Parent	You	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Parent, with you sitting right there	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Parent, and you leave the room	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
Spouse / significant other	You	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Spouse / significant other, with you sitting right there	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Spouse / significant other, and you leave the room	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
Colleague / fellow student	You	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Colleague / fellow student, with you sitting right there	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7
	Your Colleague / fellow student, and you leave the room	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

Next Page

Survey of Privacy Issues Arising From Collaboration Around a Computer

25. Please read the following scenario carefully and think for a moment about the situation. Then answer the following questions.

You have been experiencing itching and pain in your groin area. You go see the doctor who unfortunately diagnosed you with shingles on the genitals. Shingles can occur in people who have previously had chicken pox. It is a very painful disease. You have been experiencing uncomfortable symptoms and have been looking for relief. You use your web browser to search for such topics as "burning genitals" and "itching groin" and have visited such web pages as www.yoursexualhealth.com/stoptheburning.html and www.genitalhealthcare.com/topics/infectiousdiseases (which you add to your favorites for future reference).

Later, you are in a situation where you must use your web browser with others gathered in front of your computer screen as you collaborate on a project. There is a chance that the URL's to the pages you've been viewing recently and the search terms you've been using will be displayed as you use the browser for the task at hand. Please indicate how you would feel in this situation, considering *who is the person viewing and who is in control of the web browser* using the privacy comfort scale to rate your comfort level.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Person viewing	Person in control of the web browser	Your comfort in terms of privacy (use privacy comfort scale)						
Close friend	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Close friend, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Close friend, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Supervisor	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Supervisor, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Supervisor, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Parent	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Parent, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Parent, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Spouse / significant other	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Spouse / significant other, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Spouse / significant other, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Colleague / fellow student	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Colleague / fellow student, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Colleague / fellow student, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7

Next Page

Survey of Privacy Issues Arising From Collaboration Around a Computer

26. Please read the following scenario carefully and think about the situation. Then answer the following questions.

You are the one-millionth person to shop at your local grocery store and you unexpectedly win a prize for an all expense paid trip to Barcelona for you and 20 of your closest friends. You are delighted! Excitedly you start researching the prize. You use your web browser to search for such topics as "Barcelona tourism" and "Barcelona historic sites" and have visited such web pages as thing-to-do.barcelona.com and www.spanishfiestas.com (which you add to your favorites for future reference).

Later, you are in a situation where you must use your web browser with others gathered in front of your computer screen as you collaborate on a project. There is a chance that the URL's to the pages you've been viewing recently and the search terms you've been using will be displayed as you use the browser for the task at hand. Please indicate how you would feel in this situation, considering *who is the person viewing and who is in control of the web browser* using the privacy comfort scale to rate your comfort level.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Person viewing	Person in control of the web browser	Your comfort in terms of privacy (use privacy comfort scale)
Close friend	You	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Close friend, with you sitting right there	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Close friend, and you leave the room	○1 ○2 ○3 ○4 ○5 ○6 ○7
Supervisor	You	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Supervisor, with you sitting right there	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Supervisor, and you leave the room	○1 ○2 ○3 ○4 ○5 ○6 ○7
Parent	You	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Parent, with you sitting right there	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Parent, and you leave the room	○1 ○2 ○3 ○4 ○5 ○6 ○7
Spouse / significant other	You	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Spouse / significant other, with you sitting right there	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Spouse / significant other, and you leave the room	○1 ○2 ○3 ○4 ○5 ○6 ○7
Colleague / fellow student	You	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Colleague / fellow student, with you sitting right there	○1 ○2 ○3 ○4 ○5 ○6 ○7
	Your Colleague / fellow student, and you leave the room	○1 ○2 ○3 ○4 ○5 ○6 ○7

Next Page

Survey of Privacy Issues Arising From Collaboration Around a Computer

27. Please now think for a moment about *all **your** usual personal web browsing activities*. Then answer the following.

Later, you are in a situation where you must use your web browser with others gathered in front of your computer screen as you collaborate on a project. There is a chance that the URL's to the pages you've been viewing recently and the search terms you've been using will be displayed as you use the browser for the task at hand. Please indicate how you would feel in this situation, considering *who is the person viewing and who is in control of the web browser* using the privacy comfort scale to rate your comfort level.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Person viewing	Person in control of the web browser	Your comfort in terms of privacy (use privacy comfort scale)						
Close friend	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Close friend, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Close friend, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Supervisor	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Supervisor, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Supervisor, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Parent	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Parent, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Parent, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Spouse / significant other	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Spouse / significant other, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Spouse / significant other, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
Colleague / fellow student	You	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Colleague / fellow student, with you sitting right there	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7
	Your Colleague / fellow student, and you leave the room	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7

Next Page

Survey of Privacy Issues Arising From Collaboration Around a Computer

Web browsers offer various convenience features such as Favorites/Bookmarks, History, and Auto-completion to allow for easier web browsing; but these features may also display material that can be inappropriate. Please think about how you handle the tradeoff between convenience and privacy.

28. The History feature allows you to keep a record of URL's visited. How is this feature set on the various computers you use? (check one setting for each applicable computer)

Setting	Home Computer	Work/School Computer	Laptop
Unsure	○	○	○
Default Setting	○	○	○
Set for 0 days history to be stored	○	○	○
Set for some number of days history to be stored	○	○	○
- Specify number of days (if known)			

Please briefly describe any other ways you would like to be able to handle this feature.

29. The AutoComplete feature stores previous entries and lists possible matches from entries you've typed before. How do you currently have this feature set? (check all settings in use for each applicable computer).

Setting	Home Computer	Work/School Computer	Laptop
Unsure	☐	☐	☐
Default Setting	☐	☐	☐
Use Auto Complete for web addresses	☐	☐	☐
Use Auto Complete for forms	☐	☐	☐
Use Auto Complete for user names and passwords on forms	☐	☐	☐
Don't use	☐	☐	☐

Please briefly describe any other ways you would like to be able to handle this feature.

30. The Favorites/Bookmarks feature allows you save the title and web address of web pages that you would like to re-visit. How do you use this feature? (check one setting for each applicable computer)

Setting	Home Computer	Work/School Computer	Laptop
Use it to save web addresses with default/accurate names	○	○	○
Use it to save web addresses, but rename some to conceal the identity	○	○	○
Don't use	○	○	○

Please briefly describe any other ways you would like to be able to handle this feature.

31. If you had advance warning that somebody else would be working closely with you as you used your web browser and could see all areas of your screen, what actions would you take to conceal potentially sensitive information? (check all actions you would take for each applicable computer)

Setting	Home Computer	Work/School Computer	Laptop
No actions	☐	☐	☐
Retain control of the keyboard/mouse and limit functionality	☐	☐	☐
Check Favorites/Bookmarks and remove any inappropriate web pages	☐	☐	☐
Check Favorites/Bookmarks and rename any inappropriate web pages	☐	☐	☐
Check History and clear if any inappropriate entries	☐	☐	☐
Check Auto-completions and clear if any inappropriate entries	☐	☐	☐
Erase all Favorites/Bookmarks	☐	☐	☐
Erase all History records	☐	☐	☐
Erase all passwords in Auto complete	☐	☐	☐
Erase all forms in Auto complete	☐	☐	☐

Please briefly describe any other actions you would like to be able to take in this situation.

32. (OPTIONAL) Please describe an occasion when you felt uncomfortable that traces of your computer activity were viewed by others.

Next Page

You are finished! Thank you for your participation.

If you have any comments or questions, please contact Kirstie Hawkey at hawkey@cs.dal.ca or (902) 452-4675.

Appendix C: Field Study (PG1 & PG2) Questionnaires

PG1 Field study:

Install Session:

Demographic Questionnaire

Privacy Background Questionnaire (Laptop version)

Future Viewers Classification Task

Website Classification Task

Uninstall Session:

Privacy Gradient Questionnaire

Privacy Background Questionnaire (Laptop version)

Future Viewers Classification Task

Website Classification Task

PG2 Field Study:

Install Session:

Demographic Questionnaire

Privacy Background Questionnaire (Desktop or Laptop version)

Uninstall Session:

Privacy Gradient Questionnaire

Future Viewers Classification Task

Website Classification Task

Demographic Questionnaire Participant #_____

Please answer all questions honestly. Your responses will be kept confidential.

1. What is your sex? (check next to answer) □ Male □ Female

2. What is your age? _____ (years)

3. What is your highest level of completed education? (check next to answer)
   - □ Less than high school
   - □ Completed high school
   - □ Completed technical school (specify field) _____
   - □ Some university (specify field) _____
   - □ Completed university degree (specify field) _____
   - □ Some graduate work (specify field) _____
   - □ A graduate degree (specify field) _____

4. What is your occupation? _____

5. How many years have you been in this occupation? (if student, leave blank) _____ (years)

6. How many years have you been using a computer on a regular basis? _____ (years)

7. How many hours in an *average week* do you use a computer? (check next to answer)
   □ < 7 hours □ 8-14 hours □ 15-21 hours □ 22-28 hours □ 29-35 hours □ 36+ hours

8. How many hours in an *average week* do you use a web browser? (check next to answer)
   □ < 7 hours □ 8-14 hours □ 15-21 hours □ 22-28 hours □ 29-35 hours □ 36+ hours
   Percentage of that time spent for personal reasons? _____%
   Percentage of that time spent for work-related reasons? _____%
   Percentage of that time spent for educational reasons? _____%

Privacy Background Questionnaire (Laptop Version) Participant #_____

Several of the questions on the next pages will ask you to think about *how comfortable* some situation *makes you feel in terms of privacy.* When answering those questions, please use the following scale to indicate your comfort level. This scale will be included at the bottom of any page that contains questions that use the privacy comfort scale.

Privacy comfort scale						
extremely uncomfortable	**very uncomfortable**	**somewhat uncomfortable**	**Neutral**	**somewhat comfortable**	**very comfortable**	extremely comfortable
1	2	3	4	5	6	7

1. Please think about when you are using your laptop (at home, work, and/or school). On an average day, please think about *who can clearly see the contents of your screen as you are using it* (this could be somebody working with you at your computer or sitting close by and able to see your entire screen), approximately *how often* they may be in this situation, and, in general, how *comfortable this makes you feel in terms of your privacy.* Circle information for all that apply and use the privacy comfort scale below when rating your comfort level.

Potential viewer	**How often they may see the screen**					**Your comfort in terms of privacy (use privacy comfort scale below)**						
Parents	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Spouse / Significant Other	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Close friends	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Acquaintances	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Colleagues/fellow students	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Clients	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Supervisor	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Employees/Students (that you supervise)	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Audience at a presentation you give	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Technical support staff	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7

2. Now think about *who may use your computer after you use it* at home, work, and/or, approximately *how often they may use it*, and, in general, how *comfortable this makes you feel in terms of your privacy.* Circle information for all that apply and use the privacy comfort scale below when rating your comfort level.

Potential user	**How often they may use the computer**					**Your comfort in terms of privacy (use privacy comfort scale below)**						
Parents	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Spouse / Significant Other	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Close friends	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Acquaintances	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Colleagues/fellow students	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Clients	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Supervisor	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Employees/Students (that you supervise)	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Technical support staff	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7

Privacy comfort scale						
extremely uncomfortable	**very uncomfortable**	**somewhat uncomfortable**	**Neutral**	**somewhat comfortable**	**very comfortable**	extremely comfortable
1	2	3	4	5	6	7

Please think for a moment about your personal web browsing activities. Then answer the following.

3. You are in a situation where you must use your web browser with others gathered in front of your computer screen as you collaborate on a project. There is a chance that the URL's to the pages you've been viewing recently and the search terms you've been using will be displayed as you use the browser for the task at hand. Please indicate how you would feel in this situation, considering *who is the person viewing* and *who is in control of the web browser* using the privacy comfort scale below to rate your comfort level.

Person viewing	**Person in control of web browser**	**Your comfort in terms of privacy (use privacy comfort scale below)**						
Close friend	You	1	2	3	4	5	6	7
	Your close friend, with you sitting right there	1	2	3	4	5	6	7
	Your close friend, and you leave the room	1	2	3	4	5	6	7
Supervisor	You	1	2	3	4	5	6	7
	Your supervisor, with you sitting right there	1	2	3	4	5	6	7
	Your supervisor, and you leave the room	1	2	3	4	5	6	7
Parent	You	1	2	3	4	5	6	7
	Your parent, with you sitting right there	1	2	3	4	5	6	7
	Your parent, and you leave the room	1	2	3	4	5	6	7
Spouse/ significant other	You	1	2	3	4	5	6	7
	Your spouse/significant other, with you sitting right there	1	2	3	4	5	6	7
	Your spouse/significant other, and you leave the room	1	2	3	4	5	6	7
Colleague/ fellow student	You	1	2	3	4	5	6	7
	Your colleague/fellow student, with you sitting right there	1	2	3	4	5	6	7
	Your colleague/fellow student, and you leave the room	1	2	3	4	5	6	7

Privacy comfort scale

extremely uncomfortable	**very uncomfortable**	**somewhat uncomfortable**	**Neutral**	**somewhat comfortable**	**very comfortable**	extremely comfortable
1	2	3	4	5	6	7

Web browsers offer various convenience features such as Favorites/Bookmarks, History, and Auto-completion to allow for easier web browsing; but these features may also display material that can be inappropriate. Please think about how you handle the tradeoff between convenience and privacy.

4. The History feature allows you to keep a record of URL's visited. How is this feature set on your laptop?

Setting	Laptop
Unsure	□
Default setting	□
Set for 0 days history to be stored	□
Set for some number of days history to be stored	□
Specify number of days (if known)	□

How would you like to be able to manage this feature?

_____

5. The AutoComplete feature stores previous entries and lists possible matches from entries you've typed before. How do you currently have this feature set? (check all settings in use on your laptop).

Setting	Laptop
Unsure	□
Default setting	□
Use Auto Complete for web addresses	□
Use Auto Complete for forms	□
Use Auto Complete for user names and passwords on forms	□
Don't use Auto Complete	□

How would you like to be able to manage this feature?

_____

6. The Favorites/Bookmarks feature allows you save the title and web address of web pages that you would like to re-visit. How do you use this feature? (check one setting)

Setting	Laptop
Use it to save web addresses with default/accurate names	□
Use it to save web addresses, but rename some to conceal the identity	□
Don't use	□

How would you like to be able to manage this feature?

_____

7. If you had advance warning that somebody else would be working closely with you as you used your web browser and could see all areas of your screen, what actions would you take to conceal potentially sensitive information? (check all actions you would take for each applicable computer)

Setting	**Laptop**
No actions	□
Retain control of the keyboard/mouse and limit functionality	□
Check Favorites/Bookmarks and remove any inappropriate web pages	□
Check Favorites/Bookmarks and rename any inappropriate web pages	□
Check History and clear if any inappropriate entries	□
Check Auto-completions and clear if any inappropriate entries	□
Erase all Favorites/Bookmarks	□
Erase all History records	□
Erase all passwords in Auto complete	□
Erase all forms in Auto complete	□

Privacy Background Questionnaire (Desktop Version) Participant #_____

Several of the questions on the next pages will ask you to think about *how comfortable* some situation *makes you feel in terms of privacy.* When answering those questions, please use the following scale to indicate your comfort level. This scale will be included at the bottom of any page that contains questions that use the privacy comfort scale.

Privacy comfort scale						
extremely uncomfortable	**very uncomfortable**	**somewhat uncomfortable**	**Neutral**	**somewhat comfortable**	**very comfortable**	extremely comfortable
1	2	3	4	5	6	7

1. Please think about when you are using your computer(s) (at home, work, and/or school). On an average day, please think about *who can clearly see the contents of your screen as you are using it* (this could be somebody working with you at your computer or sitting close by and able to see your entire screen), approximately *how often* they may be in this situation, and, in general, how *comfortable this makes you feel in terms of your privacy.* Circle information for all that apply and use the privacy comfort scale below when rating your comfort level.

Potential viewer	**How often they may see the screen**					**Your comfort in terms of privacy (use privacy comfort scale below)**						
Parents	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Spouse / Significant Other	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Close friends	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Acquaintances	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Colleagues/fellow students	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Clients	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Supervisor	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Employees/Students (that you supervise)	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Audience at a presentation you give	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Technical support staff	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7

2. Now think about *who may use your computer(s) after you use it* at home, work, and/or, approximately *how often they may use it*, and, in general, how *comfortable this makes you feel in terms of your privacy.* Circle information for all that apply and use the privacy comfort scale below when rating your comfort level.

Potential user	**How often they may use the computer**					**Your comfort in terms of privacy (use privacy comfort scale below)**						
Parents	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Spouse / Significant Other	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Close friends	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Acquaintances	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Colleagues/fellow students	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Clients	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Supervisor	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Employees/Students (that you supervise)	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7
Technical support staff	Daily	Weekly	Monthly	Rarely	Never	1	2	3	4	5	6	7

Privacy comfort scale						
extremely uncomfortable	**very uncomfortable**	**somewhat uncomfortable**	**Neutral**	**somewhat comfortable**	**very comfortable**	extremely comfortable
1	2	3	4	5	6	7

Please think for a moment about your personal web browsing activities. Then answer the following.

3. You are in a situation where you must use your web browser with others gathered in front of your computer screen as you collaborate on a project. There is a chance that the URL's to the pages you've been viewing recently and the search terms you've been using will be displayed as you use the browser for the task at hand. Please indicate how you would feel in this situation, considering *who is the person viewing* and *who is in control of the web browser* using the privacy comfort scale below to rate your comfort level.

Person viewing	**Person in control of web browser**	**Your comfort in terms of privacy (use privacy comfort scale below)**						
Close friend	You	1	2	3	4	5	6	7
	Your close friend, with you sitting right there	1	2	3	4	5	6	7
	Your close friend, and you leave the room	1	2	3	4	5	6	7
Supervisor	You	1	2	3	4	5	6	7
	Your supervisor, with you sitting right there	1	2	3	4	5	6	7
	Your supervisor, and you leave the room	1	2	3	4	5	6	7
Parent	You	1	2	3	4	5	6	7
	Your parent, with you sitting right there	1	2	3	4	5	6	7
	Your parent, and you leave the room	1	2	3	4	5	6	7
Spouse/ significant other	You	1	2	3	4	5	6	7
	Your spouse/significant other, with you sitting right there	1	2	3	4	5	6	7
	Your spouse/significant other, and you leave the room	1	2	3	4	5	6	7
Colleague/ fellow student	You	1	2	3	4	5	6	7
	Your colleague/fellow student, with you sitting right there	1	2	3	4	5	6	7
	Your colleague/fellow student, and you leave the room	1	2	3	4	5	6	7

Privacy comfort scale

extremely uncomfortable	**very uncomfortable**	**somewhat uncomfortable**	**Neutral**	**somewhat comfortable**	**very comfortable**	extremely comfortable
1	2	3	4	5	6	7

Web browsers offer various convenience features such as Favorites/Bookmarks, History, and Auto-completion to allow for easier web browsing; but these features may also display material that can be inappropriate. Please think about how you handle the tradeoff between convenience and privacy.

4. The History feature allows you to keep a record of URL's visited. How is this feature set on each of your computers?

Setting	_____	_____
Unsure	□	□
Default setting	□	□
Set for 0 days history to be stored	□	□
Set for some number of days history to be stored	□	□
Specify number of days (if known)	□	□

How would you like to be able to manage this feature?

_____

5. The AutoComplete feature stores previous entries and lists possible matches from entries you've typed before. How do you currently have this feature set? (check all settings in use on each applicable computer).

Setting	_____	_____
Unsure	□	
Default setting	□	
Use Auto Complete for web addresses	□	
Use Auto Complete for forms	□	
Use Auto Complete for user names and passwords on forms	□	
Don't use Auto Complete	□	

How would you like to be able to manage this feature?

_____

6. The Favorites/Bookmarks feature allows you save the title and web address of web pages that you would like to re-visit. How do you use this feature? (check one setting for each PC)

Setting	_____	_____
Use it to save web addresses with default/accurate names	□	
Use it to save web addresses, but rename some to conceal the identity	□	
Don't use	□	

How would you like to be able to manage this feature?

_____

7. If you had advance warning that somebody else would be working closely with you as you used your web browser and could see all areas of your screen, what actions would you take to conceal potentially sensitive information? (check all actions you would take for each applicable computer)

Setting	_____	_____
No actions	□	
Retain control of the keyboard/mouse and limit functionality	□	
Check Favorites/Bookmarks and remove any inappropriate web pages	□	
Check Favorites/Bookmarks and rename any inappropriate web pages	□	
Check History and clear if any inappropriate entries	□	
Check Auto-completions and clear if any inappropriate entries	□	
Erase all Favorites/Bookmarks	□	
Erase all History records	□	
Erase all passwords in Auto complete	□	
Erase all forms in Auto complete	□	

Future Viewers Classification Task Participant #_____

Give a classification for each of these types of viewers based on how you would feel if these viewers saw that you'd visited sites (either accidentally or on purpose) of the various types. Classify the person as "public", if you would only like them to be able to view sites you have classified as public, "semi-public" if you wouldn't mind them viewing sites you have classified as semi-public or those sites you have classified as public, "private" if you don't mind them seeing any site that you have bothered to save.

Person viewing your screen	Privacy Classification
Parents	__ Public __ Semi-public __ Private
Spouse / Significant Other	__ Public __ Semi-public __ Private
Close friends	__ Public __ Semi-public __ Private
Acquaintances	__ Public __ Semi-public __ Private
Colleagues/fellow students	__ Public __ Semi-public __ Private
Clients	__ Public __ Semi-public __ Private
Supervisor	__ Public __ Semi-public __ Private
Employees/Students (that you supervise)	__ Public __ Semi-public __ Private
Audience at a presentation you give	__ Public __ Semi-public __ Private
Technical support staff	__ Public __ Semi-public __ Private

Web Site Classification Task Participant #_____

Give a classification for each of these types of websites based on whether or not you would mind if others saw that you visited a site of this type (either accidentally or on purpose). Classify it as "public", if you wouldn't mind anybody seeing it, "semi-public" if you wouldn't mind some subset of people seeing it, "private" if you would like to restrict most others from seeing it but still want to have access to it yourself, and "don't save" if you would not want a site of this type saved by your web browser.

Category Name / Examples	**Description**	**Classification**
Adult/Mature Content www.humorbomb.org www.steakandcheese.com	Sites that contain material of adult nature that does not necessarily contain excessive violence, sexual content, or nudity. These sites include very profane or vulgar content and sites that are not appropriate for children.	__ Public __ Semi-public __ Private __ Don't Save
Pornography www.playboy.com www.whitehouse.com	Sites that contain sexually explicit material for the purpose of arousing a sexual or prurient interest.	__ Public __ Semi-public __ Private __ Don't Save
Sex Education www.viagra.com www.sexuality.org	Sites that provide graphic information (sometimes graphic) on reproduction, safe sex practices, sexuality, birth control, and sexual development. Also includes sites that offer tips for better sex as well as products used for sexual enhancement.	__ Public __ Semi-public __ Private __ Don't Save
Intimate Apparel/Swimsuit www.victoriassecret.com www.fredericks.com	Sites that contain images or offer the sale of swimsuits or intimate apparel or other types of suggestive clothing. Does not include sites selling undergarments as a subsection of other products offered.	__ Public __ Semi-public __ Private __ Don't Save
Nudity www.bodyscapes.com www.nudistnews.com	Sites containing nude or seminude depictions of the human body. These depictions are not necessarily sexual in intent or effect, but may include sites containing nude paintings or photo galleries of artistic nature. This category also includes nudist or naturist sites that contain pictures of nude individuals.	__ Public __ Semi-public __ Private __ Don't Save
Alcohol/Tobacco www.budweiser.com www.cigar.com	Sites that promote or offer for the sale alcohol/tobacco products, or provide the means to create them. Also includes sites that glorify, tout, or otherwise encourage the consumption of alcohol/tobacco. Does not include sites that sell alcohol or tobacco as a subset of other products.	__ Public __ Semi-public __ Private __ Don't Save
Illegal/Questionable www.oppapers.com www.antiessays.com	Sites that advocate or give advice on performing illegal acts such as service theft, evading law enforcement, fraud, burglary techniques and plagiarism. Includes sites that provide questionable educational materials (e.g. term papers).	__ Public __ Semi-public __ Private __ Don't Save

Gambling www.gambling.com www.casino.com	Sites where a user can place a bet or participate in a betting pool (including lotteries) online. Includes sites that provide information, assistance, or training on placing bets or participating in games of chance. Does not include sites that sell gambling related products . Also does not include sites for offline casinos and hotels (as long as those sites do not meet one of the above requirements).	__ Public __ Semi-public __ Private __ Don't Save
Violence/Hate/Racism www.whitepower.com www.bumfights.com	Sites that depict extreme physical harm to people or property, or that advocate or provide instructions on how to cause such harm. Also includes sites that advocate, depict hostility or aggression toward, or denigrate an individual or group on the basis of race, religion, gender, nationality, ethnic origin, or other involuntary characteristics.	__ Public __ Semi-public __ Private __ Don't Save
Weapons www.weapons.com www.shooters.com	Sites that sell, review, or describe weapons such as guns, knives or martial arts devices, or provide information on their use, accessories, or other modifications. Does not include sites that promote collecting weapons, or groups that either support or oppose weapons use.	__ Public __ Semi-public __ Private __ Don't Save
Abortion www.gynpages.com www.abortionfacts.com	Sites that provide information or arguments in favor of or against abortion, describe abortion procedures, offer help in obtaining or avoiding abortion, or provide information on the effects, or lack thereof, of abortion.	__ Public __ Semi-public __ Private __ Don't Save
Arts/Entertainment www.eonline.com www.moviephone.com	Sites that promote and provide information about motion pictures, videos, television, music and programming guides, books, comics, movie theatres, galleries, artists or reviews on entertainment.	__ Public __ Semi-public __ Private __ Don't Save
Business/Economy www.ge.com www.sunbeam.com	Sites devoted to business firms, business information, economics, marketing, business management and entrepreneurship. This does not include sites that perform services that are defined in another category.	__ Public __ Semi-public __ Private __ Don't Save
Cult/Occult www.satannet.com www.churchofsatan.com	Sites that promote or offer methods, means of instruction, or other resources to affect or influence real events through the use of spells, curses, magic powers, satanic or supernatural beings.	__ Public __ Semi-public __ Private __ Don't Save
Illegal Drugs www.marijuana.org www.hightimes.com	Sites that promote, offer, sell, supply, encourage or otherwise advocate the illegal use, cultivation, manufacture, or distribution of drugs, pharmaceuticals, intoxicating plants or chemicals and their related paraphernalia.	__ Public __ Semi-public __ Private __ Don't Save
Education www.ed.gov www.dal.ca	Sites that offer educational information, distance learning and trade school information or programs. Also includes sites that are sponsored by schools, educational facilities, faculty, or alumni groups.	__ Public __ Semi-public __ Private __ Don't Save

Cultural Institutions www.childmuseum.org www.scouting.org	Sites sponsored by cultural institutions, or provide information about museums, galleries, theatres (not movie theaters). Includes groups such as 4H and the Boy Scouts of America.	__ Public __ Semi-public __ Private __ Don't Save
Financial Services www.td-canadtrust.com www.paypal.com	Sites that provide or advertise banking services (online or offline) or other types of financial information, such as loans. Does not include sites that offer market information, brokerage or trading services.	__ Public __ Semi-public __ Private __ Don't Save
Brokerage/Trading www.progressive.com www.etrade.com	Sites that provide or advertise trading of securities and management of investment assets (online or offline). Also includes insurance sites, as well as sites that offer financial investment strategies, quotes, and news.	__ Public __ Semi-public __ Private __ Don't Save
Games www.nintendo.com www.gamespot.com	Sites that provide information and support game playing or downloading, videogames, computer games, electronic games, tips, and advice on games or how to obtain cheat codes. Also includes sites dedicated to selling board games as well as journals and magazines dedicated to game playing. Includes sites that support/host online sweepstakes/giveaways.	__ Public __ Semi-public __ Private __ Don't Save
Government/Legal www.whitehouse.gov www.federalreserve.gov	Sites sponsored by or which provide information on government, government agencies and government services such as taxation and emergency services. Also includes sites that discuss or explain laws of various governmental entities.	__ Public __ Semi-public __ Private __ Don't Save
Military www.army.mil www.navy.mil	Sites that promote or provide information on military branches or armed services.	__ Public __ Semi-public __ Private __ Don't Save
Political/Activist Groups www.texasgop.org www.aclu.org	Sites sponsored by or which provide information on political parties, special interest groups, or any organization that promotes change or reform in public policy, public opinion, social practice, or economic activities.	__ Public __ Semi-public __ Private __ Don't Save
Health www.cvs.com www.webmd.com	Sites that provide advice and information on general health such as fitness and well-being, personal health or medical services, drugs, alternative therapies, medical information about ailments, dentistry, optometry, general psychiatry, self-help, and support organizations dedicated to a disease or condition.	__ Public __ Semi-public __ Private __ Don't Save
Computers/Internet www.microsoft.com www.javaworld.com	Sites that sponsor or provide information on computers, technology, the Internet and technology-related organizations and companies.	__ Public __ Semi-public __ Private __ Don't Save
Hacking/Proxy Avoidance www.anonymizer.com www.phreak.com	Sites providing information on illegal or questionable access to or the use of communications equipment/software, or provide information on how to bypass proxy server features or gain access to URLs in any way that bypasses the proxy server.	__ Public __ Semi-public __ Private __ Don't Save

Search Engines/Portals www.google.com www.yahoo.com	Sites that support searching the Internet, indices, and directories.	__ Public __ Semi-public __ Private __ Don't Save
Web Communications www.hotmail.com www.aim.com	Sites that allow or offer Web-based communication via e-mail, chat, instant messaging, message boards, etc.	__ Public __ Semi-public __ Private __ Don't Save
Job Search/Careers www.hotjobs.com www.monster.com	Sites that provide assistance in finding employment, and tools for locating prospective employers.	__ Public __ Semi-public __ Private __ Don't Save
News/Media www.cnn.com www.msnbc.com	Sites that primarily report information or comments on current events or contemporary issues of the day. Also includes radio stations and magazines. Does not include sites that can be rated in other categories.	__ Public __ Semi-public __ Private __ Don't Save
Personals/Dating www.singlelinks.com www.lovesites.com	Sites that promote interpersonal relationships.	__ Public __ Semi-public __ Private __ Don't Save
Reference www.dictionary.com www.encarta.com	Sites containing personal, professional, or educational reference, including online dictionaries, maps, census, almanacs, library catalogues, genealogy-related sites and scientific information.	__ Public __ Semi-public __ Private __ Don't Save
Chat/Instant Messaging www.aim.com www.messenger.msn.com	Sites that provide chat or instant messaging capabilities or client downloads.	__ Public __ Semi-public __ Private __ Don't Save
Email www.email.com www.hotmail.com	Sites offering web-based email services, such as online email reading, e-cards, and mailing list services.	__ Public __ Semi-public __ Private __ Don't Save
Newsgroups www.slashdot.org www.newsville.com/news/groups	Sites that offer access to Usenet news groups or other messaging or bulletin board systems.	__ Public __ Semi-public __ Private __ Don't Save
Religion www.catholic.net www.gospel.com	Sites that promote and provide information on conventional or unconventional religious or quasi-religious subjects, churches or other houses of worship. Does not include sites containing alternative religions such as Wicca (Cult/Occult) or atheist beliefs (Political/Activist Groups).	__ Public __ Semi-public __ Private __ Don't Save
Shopping www.amazon.com www.tigerdirect.com	Sites that provide or advertise the means to obtain goods or services. Does not include sites that can be classified in other categories (such as vehicles or weapons).	__ Public __ Semi-public __ Private __ Don't Save
Auctions www.bidfind.com www.ebay.com	Sites that support the offering and purchasing of goods between individuals. Does not include classified advertisements.	__ Public __ Semi-public __ Private __ Don't Save

Real Estate www.century21.com www.realtor.com	Sites that provide information on renting, buying, or selling real estate or properties.	__ Public __ Semi-public __ Private __ Don't Save
Society/Lifestyle www.sheafox.com www.style.com	Sites providing information on matters of daily life. This does not include sites relating to entertainment, sports, jobs, sex or sites promoting alternative lifestyles such as homosexuality. Also, personal homepages fall within this category if they cannot be classified in another category.	__ Public __ Semi-public __ Private __ Don't Save
Gay/Lesbian www.gay.com www.waf.org	Sites that provide information, promote, or cater to gay and lesbian lifestyles. Does not include sites that are sexually oriented.	__ Public __ Semi-public __ Private __ Don't Save
Restaurants/Dining/Food www.foodtv.com www.zagats.com	Sites that list, review, discuss, advertise and promote food, catering, dining services, cooking and recipes.	__ Public __ Semi-public __ Private __ Don't Save
Sports/Recreation/Hobbies www.espn.com www.nba.com	Sites that promote or provides information about spectator sports, recreational activities, or hobbies. Includes sites that discuss or promote camping, gardening, and collecting.	__ Public __ Semi-public __ Private __ Don't Save
Travel www.travelocity.com www.hertz.com	Sites that promote or provide opportunity for travel planning, including finding and making travel reservations, vehicle rentals, descriptions of travel destinations, or promotions for hotels or casinos.	__ Public __ Semi-public __ Private __ Don't Save
Vehicles www.autotrader.com www.boattrader.com	Sites that provide information on or promote vehicles, boats, or aircraft, including sites that support online purchase of vehicles or parts.	__ Public __ Semi-public __ Private __ Don't Save
Humor/Jokes www.comedycentral.com www.the-jokes.com	Sites that primarily focus on comedy, jokes, fun, etc. May include sites containing jokes of adult or mature nature. Sites containing humorous Adult/Mature content also have an Adult/Mature category rating.	__ Public __ Semi-public __ Private __ Don't Save
StreamingMedia/MP3 www.mp3.com www.windowsmedia.com	Sites that sell, deliver, or stream music or video content in any format, including sites that provide downloads for such viewers.	__ Public __ Semi-public __ Private __ Don't Save
SoftwareDownloads www.download.com www.tucows.com	Sites that are dedicated to the electronic download of software packages, whether for payment or at no charge.	__ Public __ Semi-public __ Private __ Don't Save
Pay to Surf www.bestfreemoneyonline.com www.mypoints.com	Sites that pay users in the form of cash or prizes, for clicking on or reading specific links, email, or web pages.	__ Public __ Semi-public __ Private __ Don't Save

For Kids www.yahooligans.com www.playhousedisney.com	Sites designed specifically for children.	__ Public __ Semi-public __ Private __ Don't Save
Web Advertisements rd.companion.yahoo.com adserver.inetzone.com	Sites that provide online advertisements or banners. Does not include advertising servers that serve adult-oriented advertisements.	__ Public __ Semi-public __ Private __ Don't Save
WebHosting www.geocities.com groups.yahoo.com	Sites of organizations that provide top-level domain pages, as well as web communities or hosting services.	__ Public __ Semi-public __ Private __ Don't Save

Privacy Gradient Questionnaire Participant #_____

Please answer all questions honestly. Your responses will be kept confidential.

1. How well did the 4-levels of privacy fit the web sites you visited?
   Please circle one.

1	2	3	4	5
Didn't fit at all	Somewhat of a bad fit	Fit some of the time	Fit most of the time	Fit all of the time

2. Were there any web sites that didn't fit neatly under any one of the classifications (public, semi-public, private, don't save)?
   _____ YES , approximately _____% didn't neatly fit under any one of the classifications

   _____ NO (skip to question 5)

3. If you answered YES to Question 2, please give examples of web sites you found hard to classify.

   _____

   _____

   _____

   _____

4. If you answered YES to Question 2, please indicate what made these web sites hard to classify:

   _____ Depends on the person looking at it

   _____ Depends on the location where it was being viewed

   _____ The website has multiple purposes

   _____ Other (please specify)_____

5. When you were applying the privacy levels in the diary, how did you sort the data? (check all that apply)

   _____ By Window ID

   _____ By Date/Time

   _____ By URL

   _____ By Page Title

   _____ By Privacy Level

   _____ Default Ordering

6. Please describe the strategy you used when applying the Privacy labels:

7. Please indicate the appropriateness of the terminology used for the privacy levels and provide alternative labels for each level.

Name of the privacy level	Is this name appropriate?	Suggested alternative
Public	YES / NO	
Semi-public	YES / NO	
Private	YES / NO	
Don't save	YES / NO	

8. Was a 4-level scale appropriate for indicating your privacy concerns?

_____ YES (skip to question 9

_____ NO

9. If you answered NO to question 8 please outline the levels that you would find most appropriate

Appendix D: PrivateBits Evaluation Materials

Pre-Session:

On-line Questionnaire

Scenario Selection Worksheet

Evaluation Session:

Researcher Script

Participant Tutorial/Reference Sheet

Practice Scenarios

Browsing Scenarios

Viewing Scenarios and Questions

Semi-Structured Interview

Final Questionnaires:

Future Viewers Classification Task (see Appendix C)

Website Classification Task (see Appendix C)

Pre-Session: On-Line Questionnaire

The on-line questionnaire shown in Appendix B was used with the following changes:

Questions 1-13 same

Questions 14-15 changed to Questions 18-19 with additional context about a specific viewer/user (see below)

Questions 16-19 now Questions 14-17

Omitted Questions 20-23 (re: public computer use)

Questions 24-25 now Questions 20-21

Omitted Question 26 omitted (positive browsing scenario)

Questions 27-32 no Questions 22-27

18. Please think about when you are using computers at home, work, and/or school. On an average day, please think about *who can clearly **see** the contents of your screen as you are using it* (this could be somebody working with you at your computer or sitting close by and able to see your entire screen), approximately *how often* they may be in this situation, and, in general, *how comfortable this makes you feel in terms of your privacy*. Select information for all that apply and use the privacy comfort scale when rating your comfort level.

For each type of person that may view your display, please give the first name of the person you are thinking of and describe a reason or activity that you might do with this person that would lead to them viewing your display.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Potential Viewer	First name (if applicable)	How often they may see the screen	Your comfort in terms of privacy (use privacy comfort scale)	Typical reason for viewing screen
Parents		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Spouse / Significant Other		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Close friends		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Acquaintances		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Colleagues / fellow students		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Clients		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Supervisor		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Employees / Students (that you supervise)		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Audience at a presentation you give		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	
Technical support staff		○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	

19. Now think about *who may **use** your computer after you use it* at home, work, and/or school, approximately *how often they may use it*, and, in general, how *comfortable this makes you feel in terms of your privacy*. Select information for all that apply and use the privacy comfort scale when rating your comfort level.

For each type of person that may view your display, please give the first name of the person you are thinking of and describe a reason or activity that would lead them to use your computer.

Privacy Comfort Scale						
extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Potential User	First name (if applicable)	How often they may use the computer	Your comfort in terms of privacy (use privacy comfort scale)	Typical reason for using computer
Parents	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____
Spouse / Significant Other	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____
Close friends	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____
Acquaintances	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____
Colleagues / fellow students	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____
Clients	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____
Supervisor	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____
Employees / Students (that you supervise)	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____
Technical support staff	_____	○ Daily ○ Weekly ○ Monthly ○ Rarely ○ Never	○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7	_____

Next Page

Pre-Session: Scenario Selection Worksheet

Selecting viewing scenarios:

1. Most regular viewer with the highest comfort level
2. Most regular viewer with the lowest comfort level
3. Most regular viewer with a moderate comfort level
4. A person that is a viewer at least sometimes and provides some contrast not existing in 1-3:
    a. Personal/work
    b. Superior/peer/underling
    c. Preferably either a: parent, spouse, close friend, colleague, or supervisor

- Principles:
    - Most common
    - Range of privacy concerns (trusted/untrusted)
    - Range of relationship equality (peer/underling/superior)
    - Those viewers that can be related to survey responses

Checklist:

Viewer	Comfort Level (1..7)	Personal or work relationship P/W	Peer status S/P/U	One of 5 viewer categories in survey (Y/N)
1.				
2.				
3.				
4.				

Viewer category	activity	PCL	Freq.	Name
Close friend				
Client				
Spouse				
Acquaintance				
Audience				
Colleague				
Employee				
Parent				
Supervisor				
Tech support				

Evaluation Session: Researcher Script

But PrivateBits Study

Before start:

- Personalize browsing and viewing scenarios with survey data
- Create folder with participant study materials
- Clear History in IE
- Create Participant folder on desktop

Evaluation Session

Consent forms (2 minutes)

Intro (2 minutes)

This research is examining privacy issues arising from the incidental viewing of private information on a computer display during co-located collaboration. For example, imagine we were working on a term paper together and gather around your computer. If you launch Internet Explorer so we could find a reference, I might be able to see traces of your previous web browsing. If you type in a url, the auto-completion may show different web sites you've visited lately. If you use the History to re-visit a web site, other web sites you've visited will be visible. While one option is to turn off these convenience features, they can be very useful when trying to revisit a site.

This study will evaluate PrivateBits, a custom web browser that helps people manage the privacy of the traces of web browsing activity that are visible within the browser convenience features. Before I introduce PrivateBits to you, I want to introduce the 4-level privacy scheme that PrivateBits uses.

When using PrivateBits, visited web sites will be classified as belonging to one of four privacy levels. The four privacy levels that partition websites (each page belongs in only 1 of the categories) are:

- **don't record browsing** (these could be irrelevant (first 17 pages of a search before you found something pertinent) or sites that you wouldn't want to visit again for whatever reason)
- For those sites that you think you might want to refer to again, there are 3 privacy levels.
  - **private** (only you and possibly a couple of close confidents/spouse, etc. should see)
  - **public** – anybody and everybody (including the queen) is welcome to see
  - **semi-public** – (something in between)

Private Bits Demo (8 minutes)

PrivateBits works on the premise that people often open up different browser windows to perform different tasks. For example, you might use one browser window to perform a literature review and in another window

have your email program open and in a third window be keeping an eye out on breaking news. PrivateBits allows you to open up browser windows in different privacy modes (public, semi-public, private) which filter what type of previous browsing is visible. The only urls, histories, etc. available in a public window would be those that have been classified as public. If the window is semi-public, both the public and semi-public sites would be visible. If the window is private, all previous visited (and saved) sites would be visible. In addition, to help classify pages at an appropriate level, private bits will automatically classify all visited pages with the privacy mode of the window. In addition to public, semi-public, and private, PrivateBits also allows you to toggle between recording and not recording browsing activity.

I'm now going to give you a guided tour of the functionality of Private Bits. We'll follow along with the user guide which you will be able to keep as a reference during the study.

Go to user guide.

Practice browsing and viewing scenario (5 minutes)

Reset Logger and Start Audio Recording

Preview of rest of session

First I'm going to give you some scenarios of varying sensitivity and ask you to use PrivateBits to browse as realistically as possible, visiting the sites that you might visit if you were doing these browsing tasks at home or at work. Please use Google for your search engine (it has been set as the home page).

After you have completed about 15 minutes of browsing, I will ask you to take a couple of minutes to make any adjustments you think are necessary to the privacy levels of the pages stored in the traces. We will then discuss four different viewing scenarios based upon your regular viewing circumstances and you will evaluate how well PrivateBits filters the visibility of information. We will then discuss the design features of PrivateBits so that you may give feedback on its functionality and usability.

Browsing Scenarios (15 minutes)

You find yourself with a few minutes free time, so you decide to do some browsing for some information. Here are two tasks that you have been meaning to do. Take a few minutes to find some relevant pages. Use PrivateBits to manage the privacy of your browsing.

Scenario interrupt 1 (at 5 minutes)

Second set of tasks (at 8 minutes)

Scenario interrupt 2 (at 13 minutes)

Privacy Maintenance (2 minutes)

Take a couple of minutes to inspect the privacy levels applied to your web browsing and make any modifications that you think are needed.

Viewing Scenarios and Questions (8 minutes)

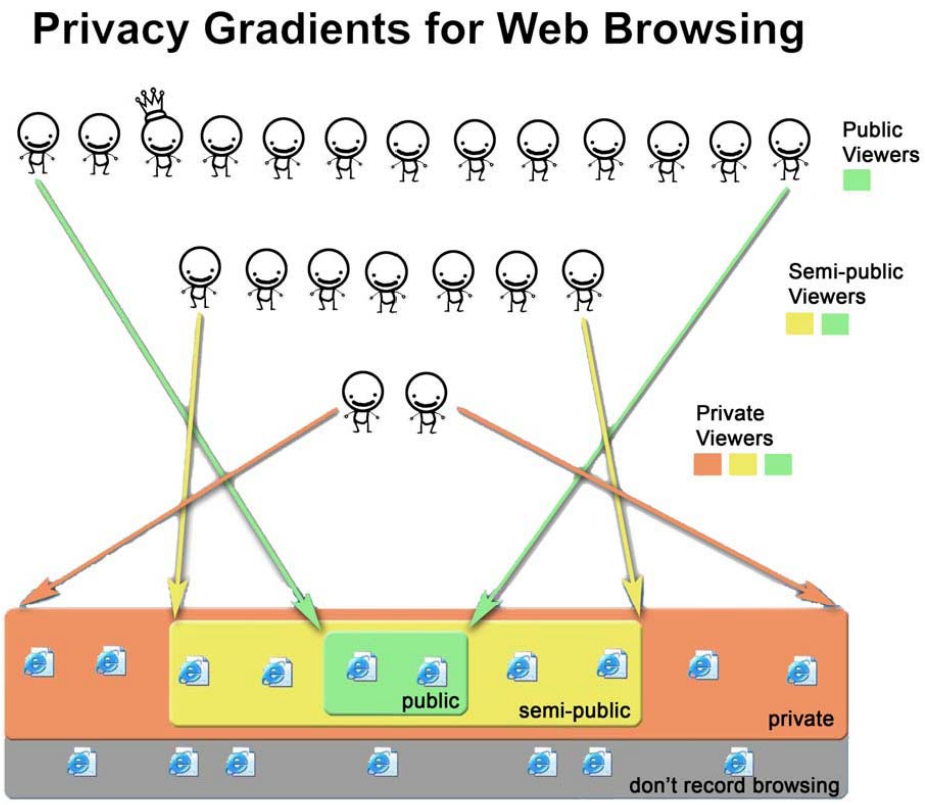
Evaluation Questions (15 minutes)

Classification tasks (5 minutes)
One thing that we've been evaluating is using content categorization of visited web pages to automatically classify browsing with an appropriate privacy level. Please complete the following two classification tasks and indicate what privacy level you think is appropriate for the various categories of web pages.

Audio recording/quotes consent (1 minute)
I will be transcribing our discussions today so that I can clarify my notes. I will only use direct quotes from the transcript in presentation of the results if you agree to it. No identifying information will be associated with a quote.

Participant payment receipts (1 minute)

Evaluation Session: Participant Tutorial and Reference Sheet



The four privacy levels that partition websites (each page belongs in only 1 of the categories) are:

- **don't record browsing** (these could be irrelevant (first 17 pages of a search before you found something pertinent) or sites that you wouldn't want to visit again for whatever reason)
- For those sites that you think you might want to refer to again (so they should be in your convenience features), there are 3 privacy levels.
  - **private** (only you and possibly a couple of close confidents should see)
  - **public** – anybody and everybody (including the queen) is welcome to see
  - **semi-public** – (something in between – whether or not it is public or private depends on the viewing context (e.g. person, location))

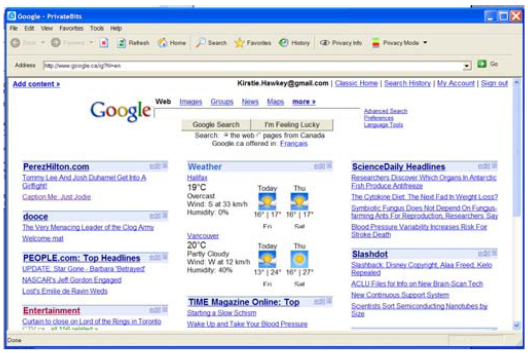
PrivateBits allows you to open a browser window in either a public, semi-public, or private mode. The only urls, histories, etc,. available in a public window are those that have been classified as public. If the window is semi-public, both the public and semi-public sites are visible. If the window is private, all previous visited (and saved) sites are visible. PrivateBits also tags all visited sites with the current privacy mode of the browser, allowing you to have windows of different privacy modes open for browsing tasks with vary content sensitivities and to easily toggle a window between different privacy modes as the nature of the visited sites changes. PrivateBits allows you to choose to save or not save traces of activity, while still filtering previously visited pages accorded to the privacy mode of the browser.

PrivateBits

PrivateBits is a custom web browser that uses Internet Explorer's functionality. You can load a new window by using the icon on the desktop:

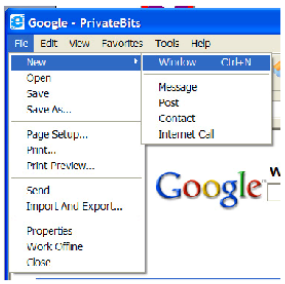
OPEN A NEW WINDOW

The default for a new window launched from the icon is to be in Public mode with no privacy feedback visible.



Once one browser window is open, a new window can be opened from within a current browser window using the file menu. New windows opened within a window have as a default the privacy settings of the current window.

NOTE: HOT KEYS TO LAUNCH A NEW BROWSER DO NOT WORK AT THIS TIME

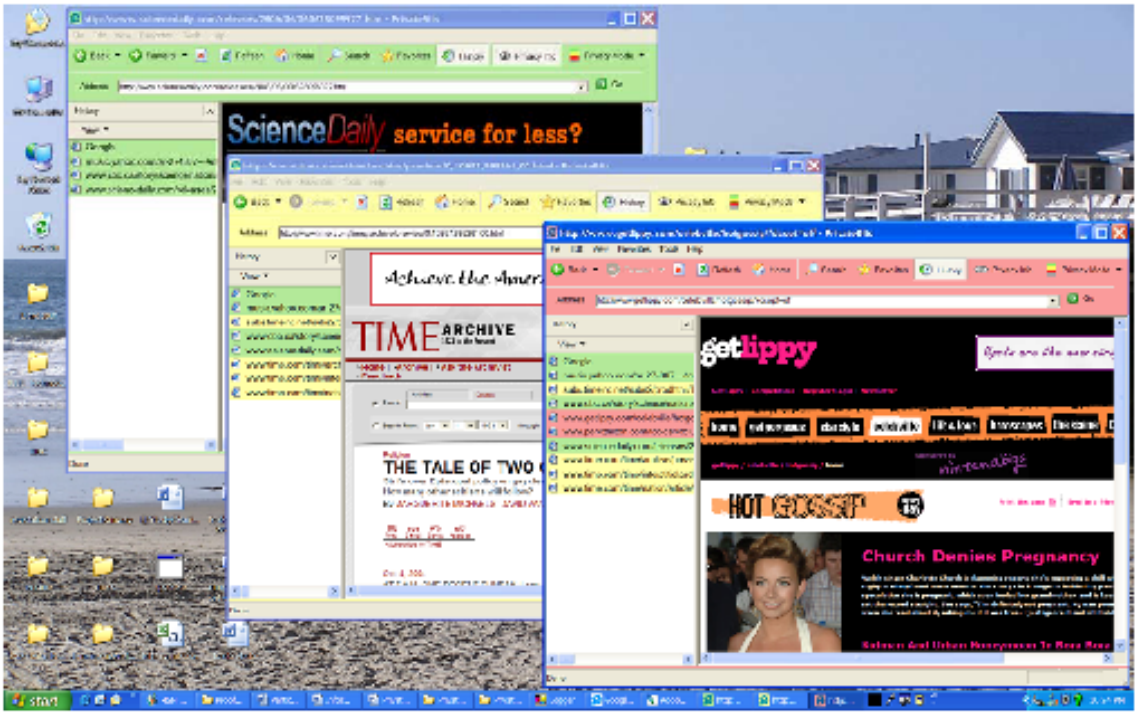


Visual privacy feedback

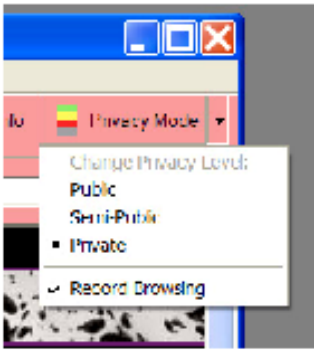
To view the privacy feedback, click on the 'Privacy Info' button'. This button toggles between showing the Privacy Info and hiding the privacy info.

The browser window will be coloured depending on the current privacy mode (red = private, yellow = semi-public, green = public). Additionally, the icon on the task bar is coloured, and traces of previous activities in the History and Favorites panel are colour coded.

Privacy Modes



To change the privacy mode of the window, select the appropriate privacy mode (Public, Semi-public, Private) from the drop down list on the Privacy Mode button:

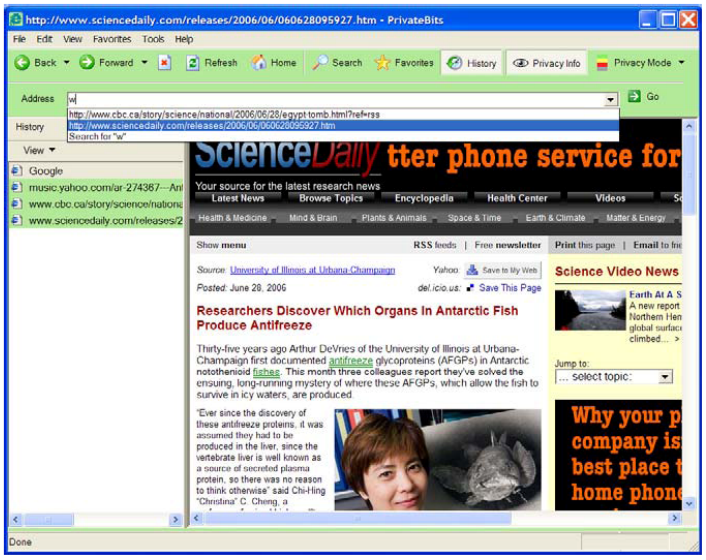


EXAMPLE: AUTO-CLASSIFICATION AND FILTERING

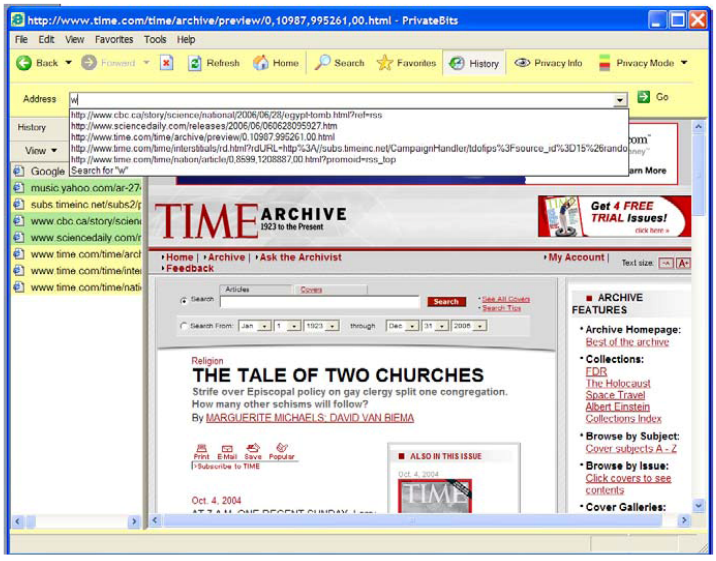
Open a total of 4 windows. Set privacy info to be visible in each window. Set one window to be in public mode, one in semi-public mode, one in private mode, and one to not record browsing (private mode for filtering).

Browse to 2 websites in each window.

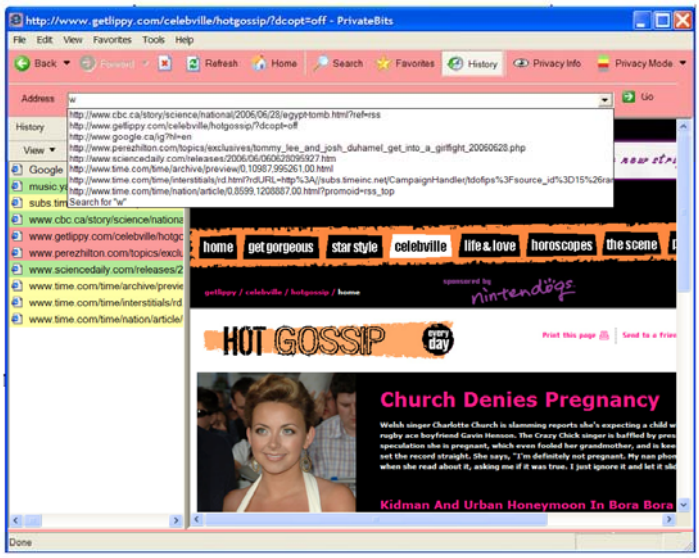
In the Public mode, all visited pages will be classified as Public and the browser window will reveal only those traces of activity that have been classified as Public. Although only the History and Favorites panel give colour coded feedback about the privacy level, the address bar, search terms, and Favorites (menu option) are also filtered.



In the Semi-Public mode, all visited pages will be classified as Semi-public and the browser window will reveal those traces of activity that have been classified as Public or Semi-public. Although only the History and Favorites panel give colour coded feedback about the privacy level, the address bar, search terms, and Favorites (menu option) are also filtered.



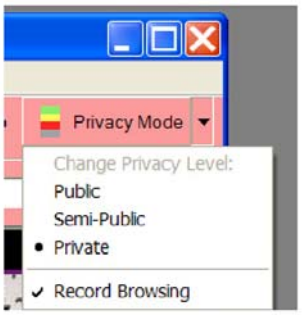
In the Private mode, all visited pages will be classified as Private and the browser window will reveal all traces of activity that have been saved. Although only the History and Favorites panel give colour coded feedback about the privacy level, the address bar, search terms, and Favorites (menu option) are also filtered.

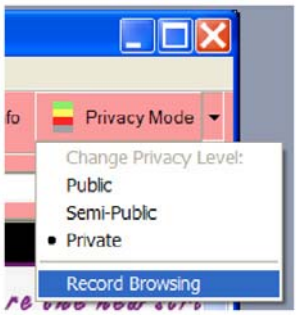


Record/ Don't Record Browsing

To toggle between saving and not saving traces of activity, check and uncheck the 'Record Browsing' option in the drop down list.

If "Record Browsing' is not checked, no activities are saved in the History or Autocompletes. The privacy mode of the web browser still filters previous activity according to the privacy mode of the browser window.

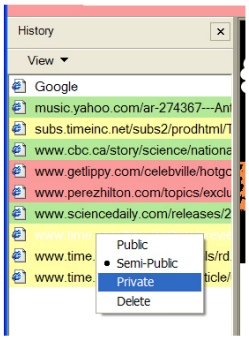




Altering the privacy classification of visited pages

When the privacy mode is changed, the privacy setting of the current page is changed to the new mode. To alter the settings of previously visited pages, view the pages in the History and right click on a page to view/change its current privacy setting.

The History panel can be sorted by date, site, or privacy level.



EXAMPLE: ALTER A PAGE'S PRIVACY LEVEL IN HISTORY PANEL.

> Open a History panel in the private window with visual feedback and also one in the public window and semi-public window. Sort the History panel by privacy level in the private window. Adjust any pages you think should be classified differently using the right click. Note how the changes are reflected in the other windows.

The Favorites panel can similarly be used to inspect and alter the privacy level of saved pages.

Auto Complete classification and filtering

URLs in the address bar auto complete are associated with the privacy level of the visited page. Changing the privacy level of the page in the history will alter what is viewed in the auto completes in the address bar.

The search term auto complete associates the current privacy mode of the browser window with the search term. To change the privacy level of a search term, set the window to be the desired privacy mode and retype the term fully in the auto complete window.

EXAMPLE: AUTO COMPLETE FILTERING

> In the semi-public window, do a search for "sort of private stuff". In one of the private windows, do a search for "super private stuff". Start doing the same search in the public window and then the semi-public window. Note that neither searches shows up in the public window, and the super private stuff search does not show up in semi-public window.

Evaluation Session: Practice Scenarios

Managing Visual Privacy in Web Browsers PID_____

Practice Scenarios

You have two minutes to fill before a meeting with your supervisor Dave to work on a project.

A friend of yours has recently lost their job and is now facing unexpected expenses. They are considering filing for bankruptcy. Take a moment and see if you can find any local support groups for people in this situation to which you can refer your friend.

Dave is about to come over to work on the project at your computer and you'll need to open a web browser. There is a chance that the URLs to the pages you've just been viewing and the search terms you've been using will be displayed as you use the browser.

Q: "Would you like to quickly check and adjust any privacy levels?"

Yes | No (what was adjusted?)

Q. "What privacy mode of browser window would you open for this collaboration?"

Public | Semi-Public | Private

Q. "Please open a window of that privacy level and open the History panel. Do you think that the data you can see in your history would be suitable for Dave to view? Yes | No

If No: "which elements would you prefer that Dave not see?"

Q: On a 7-point scale from 1-7 with 1 being extremely uncomfortable, 4 being neutral and 7 being extremely comfortable, what is your privacy comfort level of Dave being able to view this information?"

Privacy comfort scale

extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Q: Would your comfort change if you got called away and _____ was left at your computer using the keyboard and mouse? Yes | No If Yes, to what level?

Privacy comfort scale

extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Q: If you were doing this browsing again, would you change the way you managed your privacy?
Q: Would you change your browser privacy mode? Yes | No
Q: Would you opt to not record browsing? Yes | No

Evaluation Session: Browsing Scenarios

Managing Visual Privacy in Web Browsers PID: _____
Browsing Scenarios

Time 1:

Your close friend _____ has been diagnosed with {breast cancer | testicular cancer} and is understandably upset. In order to provide support to your friend, you want to learn more about the condition and treatments. Do a search for information about the disease and find three or four sites to bookmark that you think will be useful to show your friend later.

A recent topic conversation at school has been the restructuring of grocery stores to take advantage of Sunday shopping loopholes. Do a search to find out the latest information about businesses remaining open on Sundays. Try to find a recent poll about the opinion of Nova Scotians about this topic. Bookmark three or four sites so that you can share them with others later.

Time 2:

_____ has a special occasion coming up and you would like to find him a unique gift. Do a search for potential presents. Bookmark three or four sites that have items you would consider purchasing.

Time 3:

Your local radio station has a contest for free trip to see Madonna in concert. To qualify to win the trip, listeners have to call in and answer trivia questions about Madonna. Do a search and find three or four sites that contain information about Madonna's career and her current tour. Bookmark three or four of the sites so that you have quick access to the information.

One of your colleagues Mir, has been active in the volunteer community lately. Do a google search for their name and see if you can find any details about their latest activities and the charity they are involved in.

Time 4:

Your neighbour's child has a class project about reproduction and has been assigned to write a report about conception. Their computer is in the shop for repairs and your neighbour has asked you if you could print off a few sites that explain conception. Do a search for information about conception and bookmark three or four sites to print later.

Evaluation Session: Viewing Scenarios and Questions

Managing Visual Privacy in Web Browsers PID _____

Scenario ____

You are in a situation where you must use your web browser with _____ seated beside you. There is a chance that the URL's to the pages you've just been viewing and the search terms you've been using will be displayed as you use the browser for the task at hand.

Q. "What privacy mode of browser window would you open?"

Public | Semi-Public | Private

Q. "Please open a window of that privacy level and open the History panel. Do you think that the data you can see in your history would be suitable for _____ to view?

Yes | No

If No: "which elements would you prefer that _____ not see?"

Q: On a 7-point scale from 1-7 with 1 being extremely uncomfortable, 4 being neutral and 7 being extremely comfortable, what is your privacy comfort level of _____ being able to view this information (as it is now)?"

Privacy comfort scale

extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Q: Would your comfort change if you got called away and _____ was left at your computer using the keyboard and mouse? Yes | No If Yes, to what level?

Privacy comfort scale

extremely uncomfortable	very uncomfortable	somewhat uncomfortable	neutral	somewhat comfortable	very comfortable	extremely comfortable
1	2	3	4	5	6	7

Q: (if there were any elements they would prefer that _____ not see) If those elements such as _____ had been classified correctly, what would your privacy comfort level be for _____ to see the other content? Would it change if they were left alone at your computer?

Evaluation Session: Semi-Structured Interview

Managing Visual Privacy within Web Browsers

Participant #_____

Usability Evaluation (conducted in an interview format – questions read by the interviewer)

Please answer all questions honestly. Your responses will be kept confidential.

Part 1: Privacy Levels

1. "On a scale from 1-5, with one being didn't fit at all to 5 being 'fit all the time', how well did the four levels of privacy (public, semi-public, private, don't save) fit the web sites you visited during your browsing today?"

1	2	3	4	5
Didn't fit at all	Somewhat of a bad fit	Fit some of the time	Fit most of the time	Fit all of the time

2. "Were there any web sites that you weren't sure how to classify?
   YES | NO approximately _____% were hard to classify

3. If answered YES to Question 2, ask "please give examples of web sites you found hard to classify".

4. If answered YES to Question 2, ask to "what made these web sites hard to classify" – if necessary, prompt with points below:

   _____ Depends on the person looking at it
   _____ Depends on the location where it was being viewed
   _____ The website has multiple purposes
   _____ Other (please specify)_____

5. How appropriate was the terminology used for the privacy levels?

Name of the privacy level	**Is this name appropriate?**	**Suggested alternative**
Public	YES / NO	
Semi-public	YES / NO	
Private	YES / NO	
Record Browsing	YES / NO	

6. Is a 4-level hierarchical privacy scale appropriate for indicating your privacy concerns in a **work/school** setting?

   What do you think would be more appropriate?

7. Is a 4-level hierarchical privacy scale appropriate for indicating your privacy concerns in a **home** setting?

   What do you think would be more appropriate?

8. Did you ever change between privacy modes in a browser window? How easy did you find it to change between privacy levels? Was there another technique you would have preferred?

9. Did you ever change between recording or not recording your browsing activity? How easy did you find it to change between recording or not recording browsing? Was there another technique you would have preferred?

10. Did you ever change between viewing or not viewing the colour-coding feedback? How easy did you find it to change between viewing or not viewing the colour-coding? Was there another technique you would have preferred?

11. Did you find the colour-coding of private information in the web browser helpful:
    - as you were doing the intial browsing of the sites? Was there another technique you would have preferred?
    - when you were evaluating if appropriate privacy levels had been set? Was there another technique you would have preferred?
    - when you deciding what level of browser window to use for a certain viewer? Was there another technique you would have preferred?

12. Another option for feedback that would be less obvious to others is to only highlight the colour of the icon (in the browser window, in the task bar). When do you think such a less obvious technique would be beneficial? When would you want the maximum feedback?

13. As the person owning the computer, how important do you think it is to be able to conceal from those viewing your display that you are using a privacy management system to conceal some of your activities?

14. If you were the one viewing somebody else's browser, how would you feel if you became aware that they were using such a system and possibly concealing some activity traces?

15. Do you think that the buttons for Privacy Info and Privacy Mode should be concealed?

16. Do you think that it would be important to be able to password protect the Privacy Mode so that others could not change it?

17. One of the design choices made was to have the default browser window be public with no feedback as to privacy level. Do you think this is an appropriate setting? Why?

18. Another option for the default privacy mode of a new browser window is to make it semi-public, so that if you are not being careful about the privacy level, questionable items would not be mistakenly included in the public window. Do you think this would be a good idea? Why?

19. One of the design choices made was to have the privacy level of the current page change when the privacy mode of the browser changed rather than applying the privacy level just to the pages browsed next. Do you think this is appropriate?

20. Do you think that using PrivateBits would change your current privacy management techniques in your web browser?
    - How long you store your History?
    - How accurately you name your Favorites?
    - Which types of Auto Completes you choose to view?

21. If fully developed, would you use PrivateBits to manage your privacy? Why or why not?

22. Can you think of any other changes that would help improve the system?

Appendix E: Publications

Portions of the research presented in this dissertation have been previously disseminated. The following table gives details about the publications and presentations that have arisen from each of the studies conducted.

Ref.	Publication Details	Contribution Areas
IIP survey:		
E1 [63]	Hawkey, K. and Inkpen, K.M. (2006). Keeping up Appearances: Understanding the Dimensions of Incidental Information Privacy. CHI 2006, Montreal, PQ, April 2006. 821-830.	Browsing Behaviour (4.5.1) Factors of IIP (5.1, 5.3-5.7, 6.2.1, 6.2.2, 6.4.1)
PG1 field study:		
E2 [59]	Hawkey, K. and Inkpen, K.M. (2005) Privacy gradients: exploring ways to manage incidental information during co-located collaboration. Late Breaking Results: Short Papers, CHI 2005. Portland, OR, USA. 1431 - 1434.	Privacy Patterns (5.2)
E3 [60]	Hawkey, K. and Inkpen, K.M. (2005) Web browsing today: the impact of changing contexts on user activity. (Late Breaking Results: Posters) CHI 2005. Portland, OR, USA. 1443 - 1446.	Browsing Behaviour (4.1-4.4)
PG2 field study:		
E4 [61]	Hawkey, K. and Inkpen, K.M. (2006). Examining the Content and Privacy of Web Browsing Incidental Information. WWW 2006, Edinburgh, UK, May 2006, 123-132.	Browsing Behaviour (4.1, 4.2, 4.5.2) Privacy Patterns (5.2) Content Categorization Feasibility (7.3)
Exploratory Studies - Methodology:		
E5 [85]	Kellar, M., Hawkey, K., Inkpen, K.M., and Watters, C. (In press) Challenges of Capturing Natural Web-based User Behaviours. *In Use, In Situ: Extending Field Research Methods, Special issue of the International Journal of Human Computer Interaction* (accepted)	Qualitatively annotated log data (3.1, 3.2, 3.5, 3.6, 9.2)
E6 [56]	Hawkey, K. (2005). Privacy Management of Incidental Information During Collaboration: Data Analysis and Evaluation Challenges. *Workshop on Usage Analysis: Combining Logging and Qualitative Methods*, CHI 2005. Portland, OR, April 3, 2005.	
E7 [44]	Edmonds, K.A., Hawkey, K., Kellar, M., Turnbull, D. (2006) Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, WWW 2006.	Web data collection methods (3.1, 3.5, 3.6, 9.2)
E8 [57]	Hawkey, K. (2006). Mission Impossible? Capturing Rich Yet Natural User Behaviour on the Web. Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, WWW 2006. Edinburgh, Scotland, May 23, 2006.	
E9 [58]	Hawkey, K. (2006). Privacy Research: A Mixed Methodology Approach. *Workshop on Privacy and HCI: Methodologies for Studying Privacy Issues*, CHI 2006. Montreal, Canada, April 23, 2006	Mixed methodology approach to privacy research (3.1-3.7, 9.1)

PrivateBits:		
E10 [66]	Hawkey, K., Inkpen, K.M. (2007). PrivateBits: Managing Visual Privacy within Web Browsers, To appear at Graphics Interface (GI 2007). (9 pages).	Design Guidelines (7.1) Design and Implementation (8.1, 8.2) Evaluation (8.4, 8.5)
E11 [65]	Hawkey, K. and Inkpen, K.M. (2006). PrivateBits: Managing Visual Privacy in the Web Browser. Demonstration at CSCW 2006. Banff, AB, November 4-8, 2006	Design and Implementation (8.1, 8.2)
E12 [64]	Hawkey, K. and Inkpen, K.M. (2006). Managing Visual Privacy within the Web Browser. Poster presentation at the Symposium on Usable Privacy and Security (SOUPS) 2006. Pittsburgh, PA, July 12-14, 2006.	Design Requirements (7.1) Interface (8.1)
Future Directions		
E13 [62]	Hawkey, K. and Inkpen, K.M. (2006). Incidental Information Privacy and PIM. Presentation at Personal Information Management: Now that we're talking, what are we learning? (PIM 2006), SIGIR 2006 2-Day Workshop. Seattle, WA, August 10-11, 2006, 67-70.	IIP and PIM 10.3.2.1