Incognito Online: Why and How People Hide Their Information

by

Ruogu Kang

September, 2015
CMU-HCII-15-105

Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Thesis committee:

Sara Kiesler, Carnegie Mellon University (Chair)
Laura Dabbish, Carnegie Mellon University
Lorrie Cranor, Carnegie Mellon University
Alessandro Acquisti, Carnegie Mellon University

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

This work was supported by NSF grants CNS1345305 and CNS1221006.

Abstract

The communication landscape online has changed significantly from the early days of the Internet. In most developed countries, people are constantly connected through the Internet to almost everyone else in their lives everywhere they go. The Internet makes their lives more convenient, but unintentional exposure of personal information to unexpected audiences can cause emotional and tangible damage. After information leakage, some people adopt remedies such as self-censoring posts on social media, changing their passwords, not registering on websites, and using anonymous communication tools. Many people, however, do not take any action. Some feel that anything they do will be ineffective. This thesis investigates the circumstances under which people hide their information online, their motivations, and how they do so.

My findings show that many people who use the Internet, at least sometimes, want to hide their identity, content, or interactions from threats to their informational and social privacy. I used interviews, surveys, and online experiments to examine various factors that influence people's intentions and decisions to protect their privacy online. People's technical knowledge background and awareness of personal information access (as informed by system interfaces) have mixed effects on their behavioral intentions. My results show that an increased awareness of social privacy threats (measured by perceived access of other individuals to their data) leads to a higher intention to take privacy protection actions, but this intention may not always translate into actual disclosure behavior.

This work provides implications for future design and research related to Internet privacy. The findings indicate that a higher level of system transparency or more user education might not be effective in influencing people to take more secure online action. The findings suggest we need more research effort to improve policies and systems that can protect users' privacy and security online without undue reliance on their own behavior.

Table of contents

List of tables

List of figures

1 Introduction

The social nature of human beings prompts them to share information with other people. In much of the world, this communication happens on the Internet through websites and communication tools. Large amounts of personal information are available and traceable online. Widely disseminated news reports (Greenwald, 2013; Weise, 2014; Mcmillan, 2014) have raised people's concerns about government surveillance, company data leakages, and tracking techniques launched by websites, apps, and mobile service providers. Their own activities on social media for example, if revealed in unintended ways, can endanger people's social relationships (Litt et al., 2014). These risks have made many people increasingly worried about their privacy and security online. Some who have unintentionally revealed personal information have suffered emotional and tangible damage (Kang, Brown, & Kiesler, 2013; Shay, Ion, Reeder, & Consolvo, 2014; Woodruff, 2014). Ordinary people, however, have very limited control over the use of their information. They have modest technical knowledge of the Internet, and do not fully understand what information about them is revealed to others, where their information is held, and who has access to it (Bernstein, Bakshy, Burke, & Karrer, 2013; Lin et al., 2012; Rainie, Kiesler, Kang, & Madden, 2013). Helping people manage their own privacy and security on the Internet is more important today than ever before.

Internet users are concerned not only about how companies or governments collect and use their personal data, but also about what can be seen by their friends and families, or random strangers who come across their Facebook profile through a friend of a friend. Sometimes the latter concerns can feel even more threatening than their data being accessed by governments or commercial organizations.

Prior research on how people manage their privacy in relation to organizations or other individuals is not well integrated. Some studies show how people perceive data collection by organizations such as advertisers and companies (Leon et al., 2013; Smith et al., 2011); others focus on access by individuals, and how people manage their profiles and interactions on social media sites (Ellison, Heino, & Gibbs, 2006; Marwick & boyd, 2011; Stutzman & Hartzog, 2012). We may ask: When people are online, are they equally concerned about both types of privacy threats?

There are so many potential threats to different aspects of privacy from different sources, that as the various issues and concerns accumulate, it becomes evident that we need to understand how people see their online privacy overall, and what they try to do to restore privacy or prevent loss of privacy, not only from organizations, but also from their social relationships. What's known is that people take various steps towards protecting or restoring their privacy but very few have a comprehensive strategy. For example, they might manage website cookies (Turner & Dasgupta, 2003; Ur et al., 2012), delete posts from a social media site (Sleeper et al., 2013) or anonymize their communications using coded languages (Vitak et al. 2014). Privacy researchers have investigated particular solutions such as improving the privacy policies of mobile applications (Sadeh et al., 2009) and the privacy settings on social media sites (boyd & Hargittai, 2010). However, much of the literature (reviewed below) suggests that people often neglect to do anything (Dommeyer & Gross, 2003), or simply give up (Shklovski & Kotamraju, 2011). The conditions under which people make explicit and implicit choices to protect their privacy — or some aspects of their privacy — are just beginning to be understood.

In order to inform the design of future Internet technology and policy, this thesis set the goal of finding out who the people are who seek to control their privacy (from both informational and social threats) online, what actions they take, and why they do so. The following research questions are addressed in this work:

RQ1: Why do people try to hide their identity and information online?

RQ2: How do people hide their identity and information online?

RQ3: How do people understand different privacy threats?

RQ4: How do people's understanding of privacy threats affect their decisions to hide their information online?

1.1 Overview of thesis

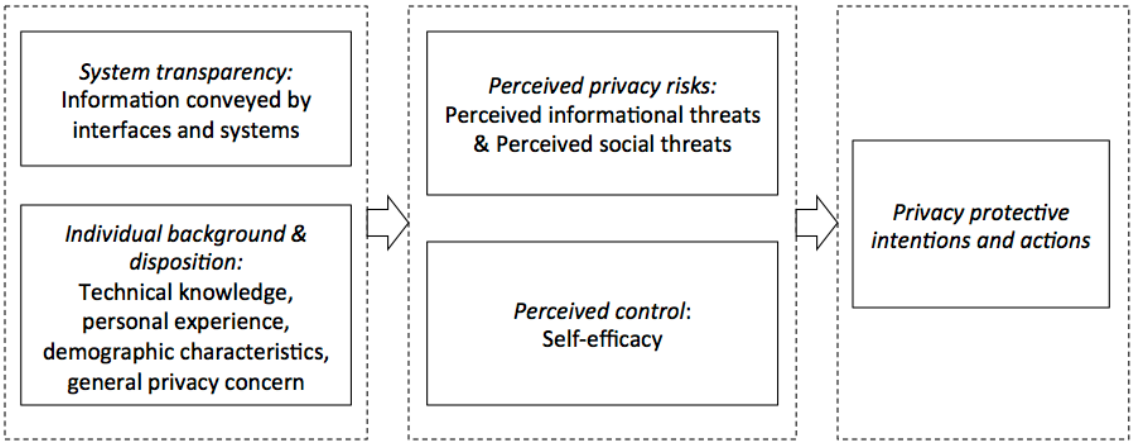


Figure 1. Overview of the research model.

The overall research model of this thesis (Figure 1) is derived from the APCO model of information privacy (Smith, Dinev, & Xu, 2011), the theory of planned behavior (Ajzen, 1991), and fear appeals theory (Witte, 1994). The APCO model links a number of antecedents (e.g., privacy experiences, demographic differences, etc.) to privacy concerns (e.g., beliefs, attitudes, perceptions), which lead to behavioral reactions (e.g., willingness to disclose information). I investigate some antecedent factors similar to the APCO model's, including privacy experiences, demographic differences, and cultural background. This thesis departs from the APCO model, however, in further exploring the influences of individuals' technical background and interface-acquired information on their privacy perceptions and behaviors because people's knowledge of how the Internet works should increase their understanding of privacy threats and how to mitigate them. Another important distinction is that, whereas the APCO model focuses only on informational privacy (privacy issues related to how institutions collect and use personal information), this thesis examines both informational privacy and social privacy (privacy issues related to how information is accessed by other people). The present work, applying the theory of planned behavior and fear appeals theory, examines the influence of both kinds of perceived privacy risk (social as well as informational) and people's perceived control on their intended and actual privacy protection behavior.

Table 1 provides an overview of the research questions answered by each chapter and the methodologies employed.

Research Questions	**Chapter**	**Methodology**
Why do people try to hide their identity and information online?	2, 3	Interview
How do people hide their identity and information	2, 4, 5	Interview, survey

online?		
How do people understand different privacy threats?	2, 5, 6	Interview, survey, online experiment
How does people's understanding of privacy threats affect their decisions to hide their information online?	6	Online experiment

Table 1. Overview of the chapters

The first part of this thesis starts with an investigation of the right-hand factors in the general model. The goal was to find some Internet users who are protecting their privacy online by trying to be anonymous, and to discover what they are trying to hide, and the reasons they give for doing so. People who have sought anonymity online (chapter 2) or have used anonymous communication applications on smartphones (chapter 3) were interviewed and queried as to their motivations and experiences. This research identified a wide range of motivations for anonymity-seeking behavior online, such as protecting family from unpleasant gossip (social threats) or self-protection from hackers or government (informational threats). As described in Chapter 2, I found that people use both technical- and behavioral-protective actions to hide their identity online, such as using advanced encryption techniques, or simply using an alias. In Chapter 3 I describe the mainly social motivations for anonymity-seeking behavior online through mobile apps: either to aid boundary management in real life or to share momentary feelings without any consequences. In Chapters 2 and 3 I discuss the tradeoffs of anonymous and identified communications and how people make the decision to be anonymous or identified online.

The second part of this thesis explores several antecedents of privacy protection behavior outlined in the first chapter. These antecedents include the desire to manage boundaries, a prior bad experience online, and technical knowledge from, for instance, former computer science education. In Chapter 4 I describe findings from two survey studies examining the prevalence of anonymity-seeking behavior among U.S. Internet users and the threats from which they hide. I also studied the kinds of privacy protection actions people perform online. The results show that prior bad experience is associated with hiding from all kinds of privacy threats and that higher technical knowledge is associated with hiding from informational privacy threats, especially. In Chapters 5 and 6 I explore the effects of technical knowledge (particularly, people's understanding of how the Internet works and sources of threat). I desribe a think-aloud qualitative study that examined how technical and non-technical users understand the Internet and how they envision various privacy threats that might be incurred in the act of transmitting information over the Internet. The results show that although people have drastically different awareness of the structural components of the Internet and of privacy threats, their reported hiding behaviors do not differ significantly. Based on the findings of chapter 5 and the above-noted behavioral theories, I carried out an online experiment, described in chapter 6, to investigate the relationship between privacy antecedents, perceived risks and efficacy, behavioral intentions, and actual actions. The participants were provided with different visualizations of the Internet structure and

privacy threats, and then answered survey questions that measured their privacy perceptions and behaviors. The findings are that seeing the list of threats increased the participants' awareness of others' access to their data. Such enhanced threat awareness, especially social threats, along with perceived self-efficacy, led to more privacy protection intentions but not actual protective behavior.

In Chapter 7, I conclude this thesis with a general discussion of the findings, and suggest some implications for future research, design, and policy. In the studies, I identified both social and informational reasons that motivate people to hide their information online. I found that people are more concerned about, and are more likely to change their behavioral intentions because of, social privacy threats than informational threats. Finally, I show that people's actual privacy protection behavior may hardly change at all, regardless of their knowledge or the information they learn from system interfaces. The present findings provide design implications for the level of transparency that future Internet systems might follow. I suggest that greater emphasis should be put on the development of privacy protection technology and policy that can remove the burden of privacy protection from Internet users themselves.

1.2 Concepts and prior work used in this thesis

1.2.1 Definition of privacy

Even after years of research, the concept of privacy is still considered difficult to define by many scholars in different domains. Law researchers define privacy as "the right to be let alone" (Warren & Brandeis, 1890). Westin (1967) defines privacy as "the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." Altman (1975) treats privacy as a dynamic and context-dependent "boundary regulation" process. Solove (2007) synthesizes a wide range of discussions centering around the conceptualization of privacy, and defines it as a "plurality of different things" rather than one single concept. The taxonomy he developed comprehends the collection, processing, dissemination, and invasion of personal information.

Much of the early research on privacy focuses on how institutions such as government or companies dealt with personal data before the proliferation of user-generated content on social networking sites or other places online. Some recent work has begun to recognize the distinction between informational privacy and social privacy (Raynes-Goldie, 2010; Young & Quan-Haase, 2013). Rader (2014) categorizes privacy concerns into information privacy and social privacy. *Information privacy* is "the control of access to personal information by organizations and institutions, and the technologies they employ to gather, analyze, and use that information for their own ends (p52)." *Social privacy* is "how we manage self-disclosure, availability, and access to information about ourselves by other people (p52)." Palen & Dourish (2003) address both concerns in their paper, using examples of surveillance, personal identity theft, and interpersonal privacy

matters. Similarly, Iachello & Hong (2007) contrast data protection with personal privacy. They summarize: "data protection refers to the management of personally identifiable information, typically by governments or commercial entities (p11);" and "personal privacy describes how people manage their privacy with respect to other individuals, as opposed to large organizations (p12)." In this thesis, I use the categorizations *informational privacy* and *social privacy* to distinguish these two concepts.

1.2.2 Informational privacy threats

Users' concerns about informational privacy threats mostly include two types of threats: companies (businesses people directly interact with and third parties) and authorities (government or authorities).

Recent advances in technology and "big data" analytics make information collection and processing by companies and other third parties more visible to users. Early research in direct mail marketing (Culnan, 1993) examined what influences people's attitudes about secondary information use. They found that those who perceive more benefits of shopping by mail, have lower privacy concerns about the loss of control of their information, and are more able to cope with unwanted mail have more positive attitudes toward secondary information use, but general privacy concerns and previous privacy invasion experience do not predict people's attitudes about secondary information use. When online behavioral advertising (OBA) penetrates widely into our everyday life, a lot of work has been done to investigate users' attitudes toward secondary information use on the Internet. Most Internet users are aware of personalized advertisements but their attitudes about advertisers using their information to send tailored ads are mixed. They perceive OBA as "annoying", "an invasion of privacy" and they are scared about being followed and monitored", but many people also think these ads are useful. People's attitudes also depend on which company collects information – they are most concerned about unfamiliar companies but least concerned about familiar brands like Google (Ur, Leon, Cranor, Shay, & Wang, 2012). Awad and Krishnan (2006) examined the relationship between people's willingness to be profiled online for personalization and how they value information transparency (it means informing users about what information a company has collected about them, and how that information is going to be used). People who place more value on information transparency are less willing to be profiled for online personalized service and advertising. The influence of privacy invasion experience, however, is different for service and advertising – experiencing previous invasions does not influence people's attitudes towards personalized service, but increases people's concerns towards personalized advertising, probably because the perceived risk associated with advertising is more salient.

In addition, people may not be aware that their activities on different sites can be linked together to identify them (Zwerdling, 2013). People's incorrect or incomplete understanding of how these technologies work contributes to their underestimation of

the potential threats. Although social networks sites and search engines do not explicitly share personally identifiable information (PII) with third parties or advertisers, research has shown that leakage of PII could occur when third party servers track user behaviors through tracking cookies. It is therefore possible for third parties to link user actions on social networks sites with specific individuals' identity or with user activities on other sites (Krishnamurthy & Wills, 2008). The Internet of things is connecting multiple devices and objects, which generates more diverse and rich data about people, even covering transportation, healthcare, and home energy use (Atzori, Iera, & Morabito, 2010). It is now much easier to track and characterize individual users through their mobile phones, Internet activities, and other ubiquitous devices or sensors. Nguyen et al (2008) examined users concerns about everyday tracking and recording technologies, including credit cards, loyalty cards, RFID, etc. Their participants overall were quite concerned about information privacy, but had much lower concern towards their data being recorded by the above technologies. Interestingly, when asked about specific sources of threat, their participants were more concerned about RFID data found out by thieves and strangers, less concerned about government and companies.

Government surveillance and intervention can affect how people manage their information online. In certain countries, government censorship can also shape how people use the Internet. In Shklovki and Kotamraju (2011)'s study of how government blocking and censorship influence people's daily use of the Internet, people execute self-censorship and may avoid contributing content online so as not to cast suspicion on themselves. Another paper (Farrall, 2012) shows that anonymity is more valued in country where individuals are aware their Internet activities are being constantly tracked by the government. Internet users who support government surveillance are found to be more willing to provide personal information online, and have lower general privacy concerns (Dinev, Hart, & Mullen, 2008). Previous surveys seem to show mixed evidence about users' opinion about their personal information being accessed by authorities: some work shows a declining trust of American public in large institutions (Twenge, Campbell, & Carter, 2014), and some shows that more than half of the public agree that government surveillance is acceptable to investigate terrorism (Pew Research Center, 2013). Solove (2007) analyzes why most people state "I've got nothing to hide" when talking about government surveillance and data mining. He suggests that people do not consider the disclosure of personal information to NSA or data mining as a strong threat to individual's privacy because those data are only accessible by government officials or computer programs. A related note is that the sense of deindividualization and the notion of "lost in the crowd" make people feel less concerned when their data being tracked and recorded (Nguyen, Kobsa, & Hayes, 2008). It is still unclear what factors influence these perceptions and opinions, such as one's cultural background, political environment or personal experiences.

1.2.3 Social privacy threats

Social privacy threats are closely related to how people regulate their social boundaries (Altman, 1975; Ashforth, Kreiner, & Fugate, 2000; Petronio, 2002). One motivation for controlling or hiding certain information from others is to manage one's self-presentation (Sleeper et al., 2013). The presentation of self, as defined by (Goffman, 1959), is how people express themselves in the presence of others. People manage their self-presentations to gain social approval (to be liked or accepted) from others (Baumeister, 1982). The Internet now becomes the grand stage for self-presentation shared by everyone. An early piece of research examined how home pages reveal about ones' identity, and found people not only used factual descriptions, but also depicted fictional personas on their home pages (Walker, 2000). Their participants, who were early adopters of the Internet in 2000 and probably only used very few Internet applications, were highly aware of what impression they gave to their readers. A later paper looked at the "true self" vs. "presented self" on the Internet (Bargh, McKenna, & Fitzsimons, 2002). Bargh and colleagues argue that Internet can be a place to express people's alternative personas such as the ideal self, future self, or potential self. Their experiments showed that people were more likely to express their true self over the Internet versus face to face, and were also more likely to project an ideal friend image to the partner they met over the Internet but not the one they met face to face. Besides people's username, the picture on their home page or their avatar in an online community, even the communication channels people choose can reflect their identity (Suler, 2002).

The rise of social networking sites further complicates the way people manage their image online. Everyone has numerous roles in life, such as parent, friend, and co-worker. People are now able to manage their images online to reflect the multiple facets of selves (Suler, 2002). Some people's roles are more integrated, whereas some others' roles are more separated (Markus & Kitayama, 1991). Sometimes these roles are incompatible with each other. For example, I interviewed a fan fiction writer who is also also a school teacher. She uses multiple Facebook accounts to maintain separate identities (Kang et al., 2013; details described in Chapter 2). She stated, "*When you work with kids, a lot of people feel like you don't have a right to a personal life. You have to be a role model at all times, even when you're not at work.*"

For people similar to the school teacher, the spread of personal information poses a serious threat to the differentiated image they want to present to different groups (Litt et al., 2014). Individuals who want to present a different image to different groups often vary in the extent to which they monitor their own behavior to make it fit the particular audience (Snyder & Gangestad, 1986). Farnham and Churchill (2011) suggest that people with a strong need for a "faceted identity" who present a different image to different groups, are particularly concerned about sharing information online. Marwick and boyd (2011) have proposed the concept of "context collapse." They argue that people always have an imagined audience in mind when communicating or sharing

information, but social media created a context collapse problem where multiple audiences are collapsed into the same context, bringing extra challenges for people to manage their self-presentation. An article used the term "peer surveillance" (Dryden, 2014) to describe the phenomena that being watched by our social connections on social media could be even more threatening than being monitored by authorities who we are usually referring to when we talk about surveillance.

Some other work adopts quantitative methods to analyze antecedents of privacy protective actions and concerns about social privacy threats. Stutzman, Capra, & Thompson (2011) studied information disclosure on Facebook. Their study shows that privacy protection behaviors such as editing privacy settings, and the extent to which participants had read the privacy policy mediated the effect of privacy concerns on people's disclosure behavior. Mohamed & Ahmad (2012) found that perceived severity of potential privacy problems, self-efficacy of protecting their information, perceived vulnerability and gender predict privacy concerns in social networking sites, but they only found a weak link between privacy concern and privacy protection behavior.

1.3 Strategies for protecting users' privacy

There have been decades of research on strategies for people to protect their information security and privacy. Some work studies strategies for protecting computer security, such as using anti-virus technology and firewall, keeping email hygiene, avoiding phishing websites and using secure passwords (Wash, 2010). Some other work reviews technologies to protect one's privacy on the Internet, including anonymizers (e.g., proxy server, and SSL), tools to block certain URLs, anonymous emailers, and Web cookie managers (Turner & Dasgupta, 2003). Chen and Rea (2004) categorized different privacy control techniques people use into three categories: falsification (falsification to access a website or to obtain software, and knowledge of cookie deletion); passive reaction (dismissal of marketing calls and unsolicited email, filtering out unwanted emails; use of new email account); and identity modification (use gender-neutral ID, dismissal of chat requests; use of multiple email accounts). Paine et al (2007) surveyed ICQ users about what actions they take to guard against privacy concerns, finding that the most commonly used actions are firewall and antivirus software. Their respondents also mentioned social actions such as limiting the amount and type of information they give away (e.g., do not share real name or contact information).

Other social actions include using privacy settings on social networking sites, using multiple profiles or multiple sites, editing or deleting posts, or limiting the amount of information shared. boyd et al (2011) studied how teenagers use privacy settings on Facebook to prevent strangers from seeing their content. Both adults and teens use what they called "social tools" to manage different social boundaries, such as using different sites (Facebook and Myspace) to communicate with different connections, and switching communication channels (Facebook vs. text message). Some of their interviewees took

extreme strategies such as constant deactivation, or constantly deleting comments they have read. They also found social strategies such as using encoded language so that only a subset of their friends is able to interpret the meaning. Similar to their findings, Vitak et al. (2014) found people selectively share information on Facebook to exhibit parts of their identity while suppressing other parts of their identity. The interviewees in their study mitigate risks of personal disclosure by moving communication to other channels or cloaking the communication using jokes or coded languages so that "only a portion of one's network understands." Stutzman and Hartzog (2012) interviewed people who maintain more than one profile on a single site or multiple profiles on multiple sites to manage boundaries in their lives. The majority of their interviewees used a strategy called "practical obscurity": using a profile that is not completely concealed but not easy to find out. Das and Kramer's study (2013) reveals that 71% of the Facebook users they sampled employ self-censoring behaviors, which means they started writing some content but did not post in the end. DiMicco and Millen (2007) studied how people manage their college connections and work connections on Facebook. People differ in the extent to which they select which photo and what information they disclose to different connections (e.g., exposure of hobbies, quotes, parties, and books for college friends vs. more conservative and professional information for professional friends). In sum, boundary regulation mechanisms people use on SNS may include filtering (using several accounts), ignoring, blocking (using pseudonym), withdrawal (self-censorship), aggression (starting arguments), compliance (accepting all requests), and compromise (Wisniewski, Lipford, & Wilson, 2012).

1.4 Prior literature: Why don't people act on privacy threats?

Although many technical and social strategies are available for people to protect their privacy, several barriers prevent them from taking effective strategies to protect their online information. Researchers consistently find people who have significant privacy concerns measured by questionnaires but who do not make privacy-preserving choices (Berendt, 2005; Jensen & Potts, 2005; Woodruff, Pihur, & Consolvo, 2014). Prior work has documented a "privacy paradox" – people often disclose more than they intended to (Brandimarte, Acquisti, & Loewenstein, 2013; Norberg, Horne & Horne, 2007), and individuals' actual behaviors do not align with their concerns (Spiekermann, Grossklags, & Berendt, 2001).

1.4.1 Lack of knowledge and awareness

Prior work tells us that people are concerned about their online information for informational and social reasons, but most lay people have poor knowledge of where their information is, how privacy-protection strategies work and what strategy to use. Meanwhile, many advanced tools are hard to use or have slow performance (e.g., Tor, encryption), and are only known by a few technically sophisticated users. For example, a qualitative study (Albrechtsen, 2007) surveyed people's security actions in

organizations, but found that although people know the importance of information security, they do not perform many actions in daily work. Their participants listed reasons include not knowing how to perform, thinking it's no their responsibility to act, or not willing to sacrifice time or functionality.

Many lay persons' privacy protective strategies stay at the browser level or simple mechanisms. Ur et al (2012)'s survey found the most commonly known strategy to stop OBA is "deleting cookies" and was only mentioned by 25% of their participants. Biddle, Patrick, & Sobey (2009) did an empirical study looking at the interface design of SSL certificates. They argue that lay users do not understand technical terms such as "server" or "encryption", and suggest that some technical details of a security protocol is only understandable by more technically advanced users therefore should not be shown in general dialog boxes. Schechter et al (2007) found users do not understand encryption or what HTTPS does to their Internet connections, and usually ignore those lock icons. They invited participants to an online banking task, and found that all their participants still provide passwords even when "HTTPS" signs are removed from the website they are accessing.

Ordinary Internet users may have little knowledge of how the Internet works and how their information can be accessed and used. In Nguyen (2009)'s study, some participant expressed uncertainty about how store loyalty cards information will be used, but avoided taking any strategy to protect it: *"You know, I have no idea, and that scares the crap out of me. But I don't really... I don't really think about these things* (p187)*."* This uncertainty also exists for social privacy threats. People can hardly estimate the actual size of the online community they participate in and the visibility of their profiles (Acquisti & Gross, 2006), and often underestimate the audience size of their posts on Facebook (Bernstein et al., 2013). In social network sites like Facebook, their privacy sometimes can be violated by what others share about them (Lampinen et al., 2011; Litt et al., 2014) .

A lack of knowledge can cause people to experience confusion, insecure, or learned helplessness. Interviewees in Shklovski and Kotamraju (2011)'s study expressed that government blocking caused some confusion when they use the Internet, such as not being able to know if some websites are accessible or not, and whether spotty connection is caused by government blocking or technical reasons. Internet users with little technical knowledge may have developed a form of learned helplessness in the face of uncontrollable data about them online. Learned helplessness is a mental state in which an organism forced to endure aversive stimuli becomes unable or unwilling to avoid subsequent encounters with those stimuli, even if they are escapable, presumably because it has learned that it cannot control the situation (Seligman, 1972). Consistent with this argument, Woodruff (2014) describes people who experienced online reputation damage and described these experiences not only as "unpleasant" but also "disempowering".

It is likely that the awareness of how information can be accessed by others and the knowledge about what strategies to use could empower people to take actions to protect their information. However, only having higher awareness and more knowledge cannot guarantee more secure actions. People's self-efficacy in information security (self-efficacy is individual's self-evaluation of their behavior) is shown to be associated with more use of security software and features (Rhee, Kim, & Ryu, 2009). Research has also shown that people's self-reported knowledge is usually inaccurate, and higher than their actual knowledge about privacy technology (Jensen & Potts, 2005), suggesting that people may be overconfident in the reported self-efficacy.

1.4.2 Bounded rationality

Part of this discrepancy between people's concerns and their actions is driven by the unstable preferences (Tversky & Kahneman, 1981). People are usually not good at estimating future risks. Most people may consider security breaches and privacy invasions as small probability events, but people cannot accurately estimate outcomes associated with small probabilities. Furthermore, people tend to focus more on immediate gratification and benefits, and ignore or underestimate risks (Acquisti, 2004; Nguyen et al., 2008). The tendency towards status quo can also influence their privacy decision – people often prefer to maintain the current status (strongly influenced by the default choices) even if the alternatives are more advantageous (Kahneman, Knetsch, & Thaler, 1991).

It is well established in prior research that people have optimism biases when estimating risks (Weinstein, 1989). People like to believe they have better chances of experiencing a desirable outcome than others, and have lower chances of experiencing a negative event compared to others. Some work argues that the 'illusion of control' contributes to the optimism bias about negative events, showing that people are more optimistically biased about negative outcomes that they perceive as controllable (Harris, 1996). This is probably due to the fact that people's actual control to most negative events is low. On the other hand, for events that people actually have a great deal of control, they tend to underestimate their controllability (Gino, Sharek, & Moore, 2011).

Due to these biases, people's privacy management behavior may be malleable (Acquisti, Brandimarte, & George, 2015). We can possibly use interface cues or the framing of a question to change people's behavior to a safer or riskier direction. For example, Knijnenburg et al. (2013)'s study shows that changing the sharing choices in location privacy setting interface might trigger people to choose even riskier decisions. Angulo et al. (2014) examined how framing influences people's attitudes toward information being collected in the emerging cloud computing environments. For non-sensitive data, people are willing to give away control of their data when provided with free cloud storage, but this effect does not hold for sensitive data. The work of Brandimarte et al. (2013) shows that feeling of control over the publication of their information increases people's feeling

of overconfidence, causes people to disclose more personal information and dismiss the access and use of their information. John, Acquisti, & Loewenstein (2011) found a number of environmental cues can change people's willingness to disclose sensitive private information (e.g., adding an ethicality rating question, changing the presentation of the survey website), by removing privacy risks from their decision making process.

1.5 Summary of previous literature

Overall, the previous literature suggests that many people who are concerned about privacy have a specific source of threat in mind. People are concerned about being tracked, monitored, and analyzed by governments or companies. People also have social and relational reasons to conceal or limit access to their information online, such as maintaining their self-presentation online or managing different boundaries of their lives. To deal with their concerns, people can choose from a wide range of strategies with different levels of technical sophistication; prior work, however, reveals many challenges to the effective prevention of these privacy threats, including the lack of knowledge of protection strategies, the lack of awareness of the threats' origins, and biases in risk estimation.

Most of the previous research has examined informational and social privacy threats separately. Yet we do not yet know whether the same group of antecedents predicts hiding from both types of threats. People might perceive informational privacy threats to be more prevalent but be less likely to act on them, perhaps because they think that it is impossible to hide from governments or companies, or because they feel themselves to be safer when "hidden in a crowd." (Nguyen et al., 2008) Social privacy threats might bring more direct consequences to people's lives, but it is harder to envision the audience, and it might be tricky to safely maneuver around different sites. This thesis compares people's perceptions of both types of threats and their effects on behavior.

Part I. Hiding their identity online

The goal of the following two chapters is to discover why and how some people hide their identity online, what they want to hide, and whom they want to hide from. Both studies find social reasons for seeking anonymity such as managing one's boundary in their lives. People use anonymity to protect their privacy, vent frustrations, or just for fun or entertainment. Their technical background, prior Internet experience, and cultural background seem to influence their perceptions of how anonymous they are and what they do to achieve anonymity. Both studies provide insights about the tradeoffs people consider when deciding to be anonymous or identified. For example, anonymous communities provide more honest feedback from a more diverse audience than what people would get from identified social networks. Results of the two studies suggest implications for future online communities to include the anonymous feature, and ways to improve anonymity tools and educate users about the different routes and threats to anonymity on the Internet.

2 Why and how do people seek anonymity on the Internet[1]

2.1 Introduction

Should people have the right to anonymity on the Internet? Or should online anonymity be banned? These questions are matters of debate among security researchers, politicians and policy analysts, community designers, architects of the future Internet and the public. Although hundreds of laboratory and field studies describe positive and negative social effects of anonymous communication (e.g., Christopherson, 2007; Suler, 2004), there is a dearth of research on Internet users' own perspectives on anonymity, and the literature that exists mainly derives from studies of one or a few online communities or activities (e.g., the study of 4chan in Bernstein et al., 2011). We lack a full understanding of the real life circumstances surrounding people's experiences of seeking anonymity, their feelings about the tradeoffs between anonymity and identifiability and the factors influencing their decision to seek anonymity.

Anonymity, one of the four privacy states according to Westin (1967), is defined as "individual in public but still seeks and finds freedom from identification and surveillance." The definition I use in this chapter is based on Gary Marx's analysis (1999): being anonymous means a person cannot be identified according to any of seven dimensions of identity knowledge, that is, the person's legal name, location, pseudonyms that can be linked to the person's legal name or location, pseudonyms that cannot be linked to specific identity information but that provide other clues to identity, revealing patterns of behavior, membership in a social group, or information, items, or skills that indicate personal characteristics. A main purpose of this chapter is to examine how people think about online anonymity and why they seek it.

---

[1] This chapter is adapted from: Kang, R., Brown, S., & Kiesler, S., (2013) Why Do People Seek Anonymity on the Internet?: Informing Policy and Design. In Proceedings of CHI 13 (pp. 2657–2666). New York, NY, USA: ACM.

What we know about these reasons is derived mainly from studies of particular activities or groups who intentionally seek anonymity, including whistle blowers (Greenberger, Miceli, & Cohen, 1987), members of stigmatized groups (McKenna & Bargh, 2000), people conducting sensitive searches (Conti & Sobiesk, 2007), hackers (Coleman & Golub, 2008), and lurkers (Preece, Nonnecke, & Andrews, 2004). Anonymity lifts inhibitions and can lead to unusual acts of kindness or generosity, or it can lead to misbehavior, such as harsh or rude language and acts that are illegal or harmful (Suler, 2004). People use the protection of anonymity to reduce the social risks of discussing unpopular opinions and taboo topics, and to create different personas online than they exhibit offline (Bargh et al 2002; Yurchisin, Watchravesringkan, & Mccabe, 2005).

Another purpose of this chapter is to investigate the strategies people use in trying to achieve anonymity online. Most tools available to achieve online anonymity are poorly understood. More than 85% of the interviewees in one study said that they did not know how to surf the Web anonymously (Conti & Sobiesk, 2007). Indeed, the average person has only a vague notion of how the Internet works (Pew Internet Project, 2010;Poole, Chetty, Grinter, & Edwards, 2008) and the potential threats for users (Jensen & Potts, 2005). This knowledge may be important because anonymity is no longer assured just by using pseudonyms or relying on the obscurity of large numbers. People shop online using credit card information often revealed to third parties. They search and browse, and their clicks are recorded. A user's comments in a blog post may be searched and connected to his professional website. Even personal health records, despite attempts to keeping them confidential, are not necessarily safe (Señor, Fernández-Alemán, & Toval, 2012). How well do people understand this context of increasing social transparency and third party use of their information? Prior work has listed different ways for people to anonymize their Internet activities, including the use of proxy servers, Secure Sockets Layer technology, anonymous emailers, and cookie managers (Turner et al 2003). These options are used by comparatively few Internet users, despite their concerns about privacy and security (Albrechtsen, 2007; Berendt, 2005; Zhang, 2005). People more often modify their own behavior to manage their identity presentations to other users, for instance, by falsifying their personal information or using multiple email accounts (Chen & Rea, 2004), or adjusting their profiles on social networks sites (Tufekci 2007). We wanted to discover how users try to achieve anonymity, and whether they are confident that they have achieved it.

2.2 Method

We recruited Internet users who said they had done something anonymously online in the past, and who volunteered for a confidential interview study. We conducted one-hour semi-structured remote interviews with them from October 2011 to March 2012 via cell phone, Skype or IM chat from an anonymous client. All chat logs and audio recordings were coded anonymously.

The interviewer asked interviewees what activities they had done online anonymously, telling them that "anonymous" meant having no connection with personal information such as their legal name or persistent email address. For each activity, the interviewer asked interviewees why they wanted to be anonymous. Interviewees were prompted to give concrete examples of anonymous activities and the history of those activities. The interviewer asked them to describe the methods they used to achieve anonymity and to evaluate their level of anonymity when taking those actions (i.e., unidentifiable to the rest of the world, to some users on the site, to some of their friends, to website moderators, or to anyone outside the community). In the second part of the interview, the interviewer asked interviewees about the activities they did using their real names or other personal information that identified them. They were asked why they used their real names for those activities. The interview ended by asking interviewees to evaluate the pros and cons of anonymous and identified communication online.

We interviewed 44 participants, 23 women and 21 men. They were recruited using Amazon Mechanical Turk, Craigslist, and university forums. We told recruits that we were interested in online anonymity and asked them to participate if they had ever used the Internet anonymously. All of our interviewees said they used the Internet frequently, and had at least one prior experience with anonymous browsing or another type of anonymous online activity. Interviewees were from the United States (15), mainland China (14), Taiwan (9), Hong Kong (1), the Philippines (1), the United Kingdom (1), Romania (1), Greece (1), and Ethiopia (1). Their ages and occupations varied widely; there were students, employees, and retirees. Interviewees reported a range of technical computing skills from practically none to advanced.

We performed qualitative data analysis using a grounded theory approach (Corbin & Strauss, 2008). The data were coded in NVivo software. In the first stage of analysis, we performed open coding, identifying anonymous activities, behaviors, and attitudes in the interview transcripts. Two coders independently coded the same subset of the interviews, discussed and resolved differences, and clarified code definitions. We then performed axial coding. We discussed the body of coded transcripts, and performed affinity diagramming to group similar concepts and generate connections. These were clustered into themes. We returned to the interviews to clarify ambiguous codes and to divide themes that were too broad into separate parts. We then examined the relationships between these thematic categories, looking for patterns in reported behaviors and motivations. We refined themes during the writing process.

2.3 Results

2.3.1 Anonymous activities

About half of the interviewees (53%) used anonymity for illegal or malicious activities such as attacking or hacking others, or they engaged in socially undesirable activities like browsing sites depicting violence or pornography. Other socially undesirable

activities included downloading files illegally, flaming others, 'peeping' others, or searching for others' personal information online. The line between illegality and undesirability was sometimes fuzzy, and many whose behavior was acceptable in some situations, for example, within a discussion forum, were fearful it would be unacceptable in others, for example, at work. It was also impossible to cleanly separate "bad guys" from "good guys" in our data because many of those who reported antisocial behaviors (e.g., behaviors that are unfriendly, antagonistic, or detrimental to social order) also reported prosocial behaviors (e.g., behaviors that are altruistic, or intended to help others).

Sixty-one percent of the interviewees mentioned instrumental activities they did anonymously, including browsing websites and downloading files. Many search engines provide personalized search results and recommendations, but some interviewees browsed anonymously to avoid tailored results and access a wider range of information or to avoid personalized advertising. Some interviewees browsed anonymously because they felt that registering or logging in was unnecessary and only benefited a company.

Ninety-three percent of the interviewees reported anonymous social interactions online. Some anonymous social activities were idiosyncratic, seemingly done for fun or amusement. An interviewee in mainland China created a fictitious profile on a social networking site to play a trick on a friend.

> *I created a profile similar to my friend's profile on Renren.com. Then I added all the contacts from his 'friends' list, and posted some funny updates daily … since he was on good terms with me, I liked to play tricks on him. He did that to me too. (#30)*

Many anonymous social activities, however, were associated with groups. We categorized seven categories of social activities that people participate in anonymously. The first category is participating in online interest groups. More than half of our interviewees were anonymously involved in various hobby groups on topics such as fiction, music, pets, games, technology, and sports. One popular reason for anonymity was that the norm of those groups was to be anonymous. In a few cases, the group had an implicit or explicit membership standard that encouraged anonymity in those who did not conform. For instance, interviewee #27 joined a Japanese video sharing community anonymously to hide his American identity, because the community excluded foreigners.

Although social networking generally requires using one's real identity, half of our interviewees reported using fictitious profiles to go on social networking or dating sites, or used false personal information when chatting online. Some interviewees used different social network profiles to separate the information they shared with different groups of people. A teacher (#17) was very active in a fandom group, and often posted fan fiction online. She wanted to keep in touch with other members of that community,

but she was afraid that she might be criticized if her family or her boss found out about her writing because it was not "real" fiction. She therefore maintained two Facebook accounts, one under her real name for family and co-workers and one under a fictitious name for fandom friends.

Nearly half of the interviewees reported posting original artwork, photographs, videos, and writing online to share with others and receive feedback. We expected interviewees to attach their real names to original works to gain status and reputation, but many interviewees chose instead to sacrifice recognition to avoid links to their offline life. Interviewee #1 participated in various online music communities every week. She always posted her songs anonymously so that no one at work would find them and judge her by them.

> *The reason I won't use my real name is to not connect my real life with the online community… I don't want my supervisors and colleagues to know about the other side of my life, since that may make my image look bad. (#1)*

Consistent with McKenna and Bargh (2000), some interviewees sought help in online support groups anonymously. Some joined online domestic abuse or parenting support groups. Others went to forums to ask questions about finances or gaming. In addition, some interviewees provided support or help to others anonymously. Interviewees chose to be anonymous to preserve their public or self-image, or to manage their online relationships. The same interviewee who liked to play tricks on his friend told us that he also visited technology forums and helped people solve technical issues. He was happy to help, but sought to avoid unwanted commitments.

> *Once I helped a guy solve a problem, then he asked my real identity and kept coming back to me. It was hard to refuse him since he knew who I was. I don't like this kind of thing being turned into an obligation. (#30)*

Thirteen interviewees mentioned buying or selling products or services with other users. Nine lived in Asian countries where BBS or forums allow people to purchase goods from other users anonymously. Four interviewees from the West also bought and sold goods online. Of these four, two mentioned using fictitious information to buy and sell items on Craigslist to avoid being identified or tracked down by online predators. The other two said they typically used their real information to pay a seller using a credit card, but sometimes they initially communicated with the seller under a pseudonym.

Nine interviewees joined political discussions on anonymous forums to contribute their views and debate with other users. Some also engaged in anonymous online voting, made online donations, or participated in social protests. Interviewees from several different countries mentioned browsing news sites and political blogs and forums anonymously to access information from blocked sites and to protect themselves from social censure or legal consequences.

Four participants anonymously posted their views about products and services. They mentioned their concerns about not knowing who would access their reviews and having their reviews stored online forever. They sought anonymity to avoid negative reactions from the subjects of the reviews or from people with opposing views. One woman explained that she always signed her postal letters with her real name because they were addressed to one person or organization, but that she preferred to write anonymously when online.

> *I posted a very bad review [of a restaurant]. And I guess I did that [anonymously]. I live in a small town, so I certainly didn't want to put my real name, although I would have no problem speaking face-to-face with the restaurant owner … If you speak to somebody face-to-face, you know who you spoke to. But when it's online, you're really potentially speaking to billions of people, and the information will last. (#21)*

In sum, we identified a variety of instrumental and social online activities that people did anonymously. Consistent with prior work, people preferred to be anonymous when seeking help or doing other activities that might make them seem socially undesirable or needy, such as when they were using online dating sites or asking for support in groups, but we also found that people pursued anonymous activities when being identified might expose to them to personal threat.

2.3.2 Reasons for seeking anonymity

Our study examined users' experiences and understanding of online anonymity. From the narratives interviewees told, we gained some insight into their decision making processes for choosing anonymity over revealing their real identity.

The prior literature suggests three factors that may lead people to seek anonymity. These include technical constraints and misunderstanding of the Internet, the online community context, and personal privacy preferences. Our interviews with people about their experiences of seeking anonymity exposed some other important factors that influenced their activities and their strategies for attaining anonymity: their prior negative experiences, their desire to manage the boundaries between their online and offline worlds and their concern about specific privacy threats.

2.3.2.1 Managing boundaries

Interviewees' decisions to seek anonymity were often influenced by their desire to control and manage the boundaries between their different social networks, groups, and environments. Interviewees often sought anonymity to prevent conflict with friends or family, to maintain a professional public image, or to avoid government attention. They wanted to preserve separate identities in real life and online, in different online groups, and in different real life groups. Twelve interviewees viewed anonymity as a way to protect their real-life relationships. Potential risks to relationships included opposing views, conflicts of interest, and loss of trust. Ninety-two percent of those who talked

about anonymity as a way to protect their real-life relationships were from Eastern countries. The relational benefits of anonymity might be more important for members of Eastern cultures, consistent with the literature on communal societies and collectivism in Eastern cultures (Hofstede, 1983).

Some interviewees wished to create boundaries between different online activities. One interviewee had frequented a website about preparing for zombie attacks. Because some of the members liked to post pictures of the weapons they owned, he was more cautious about disclosing personal information on that site than on the game sites he visited:

> *In my head, there's a big difference between video game enthusiasts and firearm enthusiasts... whenever I was interacting with the firearm enthusiasts, I wanted that extra level of protection. Not that I thought everyone was bad... I just happen to know all the guns they own. (#13)*

Interviewees who liked to express different social identities in different online settings often created and maintained multiple IDs and personas to reflect how they wanted to appear to work contacts, family and friends, or other members of their online communities. They sought to keep these personas separate by maintaining separate profiles and social circles. One woman (#16) maintained separate email, Facebook, and Twitter accounts for fandom activities and for communicating with real-life friends and colleagues. Another interviewee (#36) told us he kept two Flickr accounts, one for his friends and another he used only to share photos with his parents and older relatives.

Interviewees also used anonymity to manage restrictions in the online environment such as government policies that blocked content. When the websites that participants wanted to browse violated government policy restrictions, interviewees sometimes chose to browse anonymously. Other interviewees in this situation, however, decided not to be anonymous in order to appear "normal" (see Shklovski & Kotamraju, 2011). One man told us that he liked to visit subversive websites out of curiosity but would never register or post for fear of drawing government suspicion.

> *I just want to be perceived as a harmless voyeur of this stuff, because to me it's like spy novel stuff, and.... I don't have the money to defend myself if some overzealous cyberauthority sees me doing more than browsing. (#22)*

2.3.2.2 The role of prior experience

Prior negative experiences influenced interviewees' perceptions of how using their real identity might pose a threat and how anonymity would protect them from future threats. Fifteen interviewees used anonymity because of a prior unpleasant or frightening experience. A European woman told us she used false information in every online activity she participated in because she was once lured to another country by online criminals who pretended to offer her a job. She escaped, but the experience was terrifying.

> *My life was in danger… I was even afraid to go on the Internet at that time. But... I learned a lot of things about the Internet, and the most important, you don't have to use real information about yourself. (#19)*

Friends' or other users' prior experiences also influenced people's decisions. For example, a Chinese woman who always shopped online using fake identity said,

> *Actually I'd used my real name before, but I heard of stories like this: a retailer received a bad review, so she posted the buyer's identity information to the web and said some very bad things about the buyer. So I started to use fake names. (#8)*

Having been attacked in the past was not correlated with using a more effective or technical method for attaining anonymity. Many interviewees did not have the technical skills to avoid detection. The woman who had been lured overseas by online criminals began to change her Internet service provider every six months, believing that this action anonymized her on the Internet.

2.3.2.3 Personal threat models

Interviewees' reasons for seeking anonymity reflected a personal "threat model" of individuals or organizations. Frequently, the source of threat lay outside the particular activity, site, or group in which the person sought anonymity. Personal threat fell into five categories: online predators, organizations, known others, other users on the site or in the community, and unknown others.

Online predators included criminals, hackers, scammers, stalkers, and malicious online vendors. Fear of identity theft and spam was the main concern of those who made online sales or purchases with credit cards or account information. Fear of stalking or harassment was a major motivation for hiding one's identity when chatting, posting on forums, and building social networks. *Organizations* that posed a threat included government and business organizations. Government was a threat because it has the power to identify and punish illegal, subversive, or undesirable online activity. Companies were a threat because they could reuse or sell information to marketers and spammers. *People that the interviewees knew in real life* were sometimes named as a threat, mostly as a precaution but sometimes because of a past negative experience. Among those named were specific family members, friends, employers, teachers, co-workers, supervisors, classmates, current significant others, and previous romantic partners. Anonymity was particularly a concern for people who wished to avoid harassment from estranged or controlling parents, former friends, or previous romantic partners. *Other users on a site or in the community* could also be considered a threat.

Finally, interviewees also mentioned nonspecific malicious entities that they felt were lurking online. Thirty-nine percent of interviewees expressed the attitude that revealing personal information online is "dangerous" without any specific threat in mind. A college student who participated in technology and gaming forums lurked almost all the

time, manually changed his IP sometimes, and used multiple email accounts, but rarely had any specific threat to hide from.

> *If I do something stupid online I want to be prepared... It's just like when you prepare for a disaster, you don't know what disaster is going to strike. (#10)*

In sum, interviewees' personal threat models generally involved protection and privacy from other people and groups; they were either attacker-centric or relationship-protective. Participants sought to protect themselves from real-world threats such as getting arrested, physical attacks on themselves or their families, stalking, harassment, and loss of property or jobs. They also feared online attacks, including online harassment, trolling, and flaming. They used anonymity to prevent potential privacy leaks, expressing concerns that once their information was online, it would be stored permanently and anyone could access it. One 4chan user almost always posted anonymously, because he felt that any information he shared online would be out of his hands.

> *To a large degree, you cannot control who views, accesses, or uses any data you put on the Internet … the Internet never forgets. (#12)*

Other interviewees made similar statements.

> *The Internet is sticky - pages stay up, info stays up, etc. (#16)*

> *I have no clue where [personal information] goes or how people could access it. (#25)*

2.3.2.4 Motivations other than threat

The literature in social psychology and online communities has described motivations for anonymity that are less about threat *per se* than about the emotional effects of anonymity and ways that anonymity can help people manage their social relationships online (McKenna & Bargh, 2000). In accord with this literature, a few of our interviewees said that using a pseudonym or fictional identity made them feel "cool" or "sophisticated." Some mentioned feeling more relaxed talking to anonymous strangers than to friends. One student told us that he sometimes added random people to his online chat list to talk about things that bothered him.

> *When I'm talking to someone else and neither of us knows who the other person is, there's no apprehension. Whatever you want to say, you can just say it; you can go ahead and vent some of your frustrations. (#31)*

2.3.3 Strategies people use to attain anonymity

Participants reported using both technical and behavioral strategies to achieve anonymity. The most commonly used technical method was to change one's IP address. Interviewees used proxy servers, VPNs, and anonymizing techniques like Tor to hide

their home IP address, or they changed their IP address manually. Two interviewees used proxy servers every time they went online, and 15 interviewees applied proxies when participating in potentially compromising activities such as torrenting, accessing blocked sites, revealing sensitive information, or browsing special forums (e.g., about hacking, politics, or health). Those with more advanced technical skills used encryption to protect their information. For users with lower technical abilities, one commonly used method was to change browser settings or website-specific privacy settings to control which other users had access to their profiles. Most, however, said they did not bother because, as one interviewee explained, the tools *"are quite a bit of trouble to use."* (#13)

All interviewees, regardless of their technical expertise, used behavioral methods to hide their identity. Half of the interviewees obtained anonymity within online communities by not participating. They also limited the information they shared online. Sixteen interviewees reported sharing false information to maintain their anonymity—providing a fictitious name, using a false profile photo, and inventing biographical information when other users asked for personal information.

Interviewees who liked to express different social identities in different online settings often created and maintained multiple IDs and personas to reflect how they wanted to appear to work contacts, family and friends, or other members of their online communities. They sought to keep these personas separate by maintaining separate profiles and social circles. One woman (#16) maintained separate email, Facebook, and Twitter accounts for fandom activities and for communicating with real-life friends and colleagues. Another interviewee (#36) told us he kept two Flickr accounts, one for his friends and another he used only to share photos with his parents and older relatives.

2.3.4 People are uncertain about how anonymous they are

We asked interviewees how effectively they had achieved anonymity. We did not quiz them on their understanding of the Internet, but many interviewees revealed an incorrect or incomplete understanding of the Internet and anonymity. For example, when discussing the private browsing function of a web browser, interviewee #8 said she was not sure whether it erased her traces from the computer she was using or from the website she visited. Interviewees also confused social anonymity (e.g., hiding name, location, occupation, and so forth) with technical anonymity (e.g., hiding IP address or computer information). Many did not understand that one can be anonymous within a particular group or application but not anonymous to the ISP. Only a few possessed greater understanding of the Internet and distinguished between what members of a community knew and what might be discovered about their Internet behavior more generally. For instance, interviewee #21 said she was unidentifiable in a particular online community because of the steps she took to protect her identity (using a specific pseudonym for that community, and not revealing personal information to others), but

she also said that there is no true anonymity on the Internet because anyone with technical expertise could find out who she was.

Under Marx's definition of anonymity (Marx, 1999), we found that few achieved full anonymity even when they claimed to do so. Most participants did not reveal their real name or location, and many participants mentioned using pseudonyms to hide their identity, which use would afford incomplete protection. A few participants said that they used variations of their names or something important to them in their pseudonyms, and they were aware that some other users or website administrators could identify their real identity from their pseudonyms. Some people reported creating separate identities in different online communities to prevent their friends in one group from learning of their membership in another group. Some others, however, used the same identification information across communities or platforms, which would provide clues to their real identity. Only a few participants were aware that subtle patterns of behavior across time and applications could identify them.

2.3.5 Comparing anonymity and identifiability

Nearly all of our interviewees (86%) held both positive and negative attitudes about anonymity. Two advocated anonymity as a right and felt that it was essential to privacy and security in the digital age. Twelve said that anonymity could be misused and could allow irresponsible behavior without consequences for the perpetrators, but would not give up their rights to be anonymous because of their own situations.

Ten interviewees thought seeking anonymity as a general online strategy was a futile pursuit because advances in computing and use of digital data have made anonymity virtually impossible across applications. These participants were concerned about hackers, the government, and unknown others capturing their IP address and tracking them down. They expressed concerns about personal information being used by third parties such as proxy or torrent server owners. One government employee felt very strongly that although anonymity is essential for privacy and security, it is exceedingly difficult to achieve:

> *We, to a large degree, live in a post-privacy world, where if you know how, you can find out anything about anyone. (#12)*

Table 2 summarizes the balance of factors that interviewees recalled retrospectively about their choice to be anonymous or identified. Tradeoffs included expanding the diversity of their Internet associations versus protecting their image and relationships, freely expressing their opinions versus maintaining their credibility, and getting useful, personalized recommendations versus receiving spam.

Category	Advantages of being anonymous	Advantages of being identified

Social connections	Avoid disliked others Avoid commitment to the community Lower barrier to new relationships Protect others one cares about	Connect to real life friends Have stronger social connections Encourages more participation
Reputation and trust	Give honest rating/ recommendation	Good for reputation building Gain trust from other users
Image building	Have control over personal image Avoid embarrassment /judgment /criticism	Avoid harsh criticism Consistent with self-image
Emotional benefit	Feel relax and comfortable Feel cool and sophisticated	Feel real, integrated Feel closer to people
Express opinion	Feel free to express views	Avoid irresponsible behavior
Privacy	Have more control over personal information disclosure	Look innocent
Security	Protect personal safety Avoid legal repercussion/spam/stalk/lost of property	Hide in the crowd
Ease of use	Saves effort to log in	Easy to remember account

Table 2. Perceived tradeoffs of being anonymous vs. being identified

2.4 Discussion

2.4.1 Policy and design implications

Our results show that people from all walks of life had reason, at one time or another, to seek anonymity. A main policy tradeoff is that discouraging anonymity will discourage malicious behavior (about half of the incidents in our data) but will also discourage people from engaging in creative, helpful, and harmless online activities that they might otherwise pursue. Many people would be prevented from managing personal threat and their social boundaries because identifiability increases the bleeding of social information across time, place, and group.

Current Internet design allows for anonymity at the application level (e.g., within a website), but anonymity across applications (especially in some countries) is very difficult to achieve for most users. Further, the demographic information or content that users reveal can be linked across applications and cause them to be identified even if their legal name, email address, and IP address are hidden. An important policy question is whether Internet users should have stronger controls on their levels of anonymity, and whether the risks of anonymity outweigh its benefits. In this paper, we examined only the risks and benefits for individuals rather than for communities or the society as a whole. Recent world events, such as the rapid spread of a viral incendiary video, suggest that the freedom of individuals to act anonymously will need to be balanced against societal effects.

Forty-five percent of our interviewees expressed uncertainty about what might happen to them or their data online. They also did not have an accurate understanding of how their personal information could be accessed by others and which information would be disclosed. Interviewee #16 mentioned concerns about her practice of entering her telephone number in multiple accounts, and whether that behavior connected her multiple identities. She avoided using sites that did this.

> *I think the threat for me is mostly that Google would accidentally associate my two accounts. (#16)*

Our findings suggest we should institute higher standards for telling people what use others are making of their data and what information is actually disclosed to others when they try to hide their identity via pseudonyms or other means (see Mazurek et al., 2010; Odom & Sellen, 2012). Interviewees noted the absence of user-friendly tools for achieving anonymity. Some complained that existing proxy servers were too slow or difficult to use. Others did not know how to use anonymity tools at all. If we want to support anonymity as an option online, then we must improve the usability of tools for achieving anonymity.

Online communities will sometimes want to offer anonymity for some or all members. Such communities will probably need to develop strong norms and moderation or sanctioning processes to support prosocial behavior and prevent destructive behavior (Kraut & Resnick, 2011). Online pseudonyms allow users to build reputations inside single communities or websites such as eBay while keeping their real identities hidden. However, our interviewees sometimes wanted to build reputations across different online platforms. We suggest that new mechanisms might provide better solutions for users attempting to balance their safety concerns with their desire for widespread recognition.

2.4.2 Limitations

Although the diverse demographic and technical skills of our sample provided us with a snapshot of anonymous Internet use in different cultures, government policy areas, and knowledge contexts, our sample was not a representative sample of the population of Internet users. Limited by our interview approach, we were also unable to examine how users' strategies align with their actual anonymity levels. Further research will require a more representative sample and a more fine-grained approach to find out how Internet users in general define and seek anonymity.

Our sample and the study design did not allow us to distinguish political from cultural factors in motivations for anonymity. People in countries whose governments censor the Internet say they execute self-censorship and may avoid seeking anonymity explicitly so as not to cast suspicion on themselves (Shklovski & Kotamraju, 2011), but cultural factors, such as a cultural belief in respect for authority, could be at work as well. In our

study, Chinese interviewees weighed relational factors especially heavily when choosing to hide their identity. They also were more suspicious than other interviewees about information being used against them by officials, vendors, and strangers, and many did not register on websites when they avoid doing so. Our finding echoes other work suggesting that Chinese users are particularly likely to falsify their identity on online social network sites (Wang, Norice, & Cranor, 2011). This behavior could be due to political or cultural beliefs, or to biases in our sampling. In Chapter 4, we quantitatively investigate the effect of cultural orientations.

3 Strangers on your phone: Why people use anonymous communication applications[2]

3.1 Introduction

The previous chapter shows that many Internet users seek anonymity for a broad range of reasons. In this chapter, we specifically look into how new communication applications on smartphones such as Whisper (Figure 2), Secret, and YikYak provide ways for people to connect through smartphones and interact anonymously on any topic to an anonymous audience. Earlier work has documented how people communicate on their computers in web-based MUDs, Usenet, and Second Life (Donath, 1999; Turkle, 1995; Wellman & Gulia, 1998), but the rise of mobile devices and apps has made casual, even constant, location-based anonymous communication possible. This technological change leads us to ask how people interact with others anonymously through smartphones and why they do so. What benefits do people get out of mobile communication without identification?

In this study we extend our understanding of motivations for anonymity by analyzing users' activities and motivations on anonymous communication applications on their smartphones. We find that these applications provide a place for people to share personal emotions and experiences, positive or negative, without their needing to track their disclosure boundaries and to fear negative reputational consequences. Unlike designated confessional online communities in which people mainly share confessions and secrets (Turkle, 2012), people using mobile apps such as Whisper also share funny, entertaining posts, echoing the finding described in the previous chapter whereby some people sought anonymity for private fun or amusement. Also, despite the lack of a

[2] This chapter is adapted from: Kang, R., Dabbish, L., Kiesler, S., & Sutton, K. (2016, to appear). Strangers on your phone: Why people use anonymous communication applications. In Proceedings of CSCW '16. ACM.

cohesive and persistent set of identifiers, as users have in many online communities, people in our sample reported that these apps provide social support and validation for their feelings, experiences, and self-identity. They said people are honest and open in this setting. They said they were able to open themselves because of the lack of accountability for what they said and lack of connection to their real world identity. These benefits have been observed in studies of some online communities but in the case of anonymous mobile apps, the benefits occurred even without such pro-social features as moderation and group identity markers of those communities (Chapter 3, p.79-p.83, Kraut & Resnick, 2011).

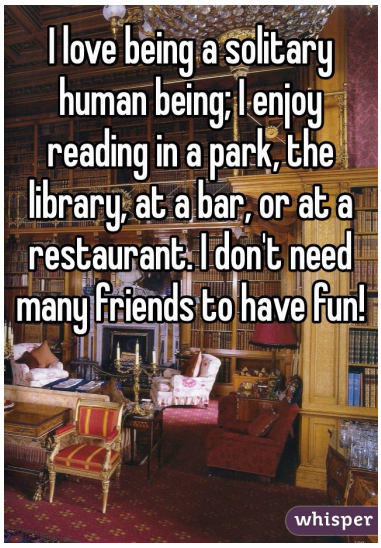


Figure 2. An example post on Whisper.

3.1.1 Anonymous communication applications

Anonymous communication applications (e.g., Whisper, YikYak, and Secret) are software programs designed mainly for mobile devices. They allow people to share messages with other users of the same application without any connection to their identity or among their messages. Unlike the quasi-anonymous email and Web communities that support group identity and specific topics such as financial trading, transgender transition, new motherhood, or cancer treatment (which often require moderation and/or registration), anonymous mobile apps are designed for crowds without an agenda or specific topic. Users can find very little identity information on these apps except others' general location at the city or state level. Previous research on these apps suggests that people using Whisper interact more often with co-located than with distant users (Wang et al., 2014).

Research so far indicates that people disclose personal information and express their personal needs and wishes on these apps. Linguistic analysis of Whispers shows a higher frequency of first person pronouns in Whisper posts than in other social media

such as Facebook and Twitter (Correa et al., 2015; Wang et al., 2014). The social psychological literature suggests that self-disclosure may be motivated by social approval, intimacy, relief of distress, social control, or identity clarification (Derlega, Grzelak, et al., 1979). People are more willing to disclose to strangers because doing so can mitigate some disclosure risks (Rubin, 1975). Disclosure risks include rejection by listeners, reduction of one's autonomy and personal integrity, loss of control or self-efficacy, and the possibility of hurting or embarrassing others (Omarzu, 2000). Disclosing online can feel safer. People disclose more about the self in the online context compared to what they do in the offline context (Joinson, 2001), and they disclosre more online when the medium contains less information about real identity (Qian & Scott, 2007). Correa et al (2015) asked MTurk workers to categorize four hundred and seventy-seven Whisper messages; they found the most popular category was "confessions," followed by "relationships," "meetup," and "QnA/Advice." Along with self-disclosure, questions and advice seeking posts are common in Whisper (Wang et al., 2014), as they are in 4chan/b/ (Bernstein et al., 2011) and Facebook Confession Board (Birnholtz, Aaron, Merola, & Paul, 2015). People may feel anonymous advice is more honest than advice from identified sources (Morris, Inkpen, & Venolia, 2014).

Ephemerality is another characteristic of anonymous communication. According to Bernstein et al. (2011), the median life of a post on 4chan /b/ is 3.9 minutes, and 43% of the posts do not receive any reply. The authors propose that ephemerality removes a "rich get richer" effect and raises community participation because users have to quickly reply to a thread to keep it alive. Similarly, more than 50% of Whispers receive no replies, and most replies arrive within 1 day after a post has been made (Wang et al., 2014). Although Whisper posts do not automatically expire, a significantly higher proportion of this content gets deleted either by moderators or by users themselves than other social media (18% vs. 4% on Twitter).

There is limited qualitative work delving into the subjective experiences of users of anonymous apps; most existing research uses text analyses to detect behavioral patterns. There remain open questions about the nature and form of participation in these apps, given their limited affordances for developing interpersonal bonds or common group identity, two theoretical factors that sustain social groups online (Ren, Kraut, & Kiesler, 2007).

3.1.2 Identity in online communities

In online communities that lack persistent user identifiers, group identity can play an important role. There is rich literature on self-expression and image management associated with one's self-identity (Goffman, 1959). Online identity signals the credibility of an information source, helps to build trust between group members, and motivates people to contribute to online communities in order to build reputation. In Donath's study of Usenet groups (Donath, 1999), anonymous accounts were commonly used

when people wanted to reveal personal information or discuss legal matters, or when they just wanted to keep their participation secret from others they knew. Studies of anonymity in the early days of the Internet suggested it could be a dangerous medium leading to angry, antisocial behavior towards others, such as "flaming"—angry or hateful speech (Kiesler, Siegel, & McGuire, 1984). Nontheless, researchers also discussed the benefits of anonymity for equalizing participation in groups (Kiesler et al., 1984), for encouring communications among weak ties (Wellman & Gulia, 1998), and for the freedom to share ideas and be open and honest without social constraint (e.g. GDSS research by Nunamaker, Applegate, & Konsynski, 1988).

The anonymity of the Internet provides people opportunities for identity experimentation. Gross (2004) proposed the Internet as a so-called identity playground for teens. Half of the students in their study had pretended to be someone other than themselves on the Internet – mainly someone older—a desired or future identity. Similarly, in Stern's work (2008), adolescents constructed personal websites as "touched-up versions of themselves."(p.106) Using the shield of online anonymity to explore alternative identities is not only seen in teens, but also in adults. In Turkle's early work of MUDs (Turkle, 1995), she describes the virtual communities as "laboratories" for people to explore and experiment with their different selves (p.12). Recent work shows that people create online dating profiles to reflect an "ideal self" instead of their actual self (Ellison et al., 2006). On the other hand, Bargh et al. (2002) argued that people are more able to disclose their so-called true self on the Internet because the anonymous environment reduces expectations and the risks of social sanction that exists in face-to-face interactions. The disinhibiting effect of anonymity online can lead to a higher level of self-disclosure (Bargh et al., 2002), discussion of taboo topics and unpopular opinions (Birnholtz et al., 2015), and unusual acts of generosity (Suler, 2004).

To explore how anonymity influences communication on mobile anonymous communication apps, we asked:

RQ 1. How and what do people post on anonymous communication apps?

We also wanted to examine why people participated.

RQ 2. What experiences or contexts motivate people to use anonymous communication apps?

3.1.3 Anonymous vs. identified online communication

Social network sites like Facebook allow people to build their images and history online and maintain relationships with real life connections. Over the years since these sites were introduced, however, people have become increasingly cautious about sharing their personal information or opinions on social network sites because of the higher risks associated with their identity tied to the content they post (Stutzman, Gross, & Acquisti, 2013). Disclosure on Facebook carries risks such as social rejection and damage to self-

presentation (Litt et al., 2014; Vitak & Kim, 2014). Others may block or unfriend posters because they post too often about politics (Sibona & Walczak, 2011). Participants in some studies describe how they attempt to mitigate posting risks by moving communication to other channels (Vitak & Kim, 2014). Marwick & boyd (2011) describes how, in order to manage different social boundaries, adults and teens both use different sites (e.g., Facebook vs. MySpace) to communicate with different connections, or to switch among communication channels (e.g., post in Facebook vs. text message). In addition to maneuvering in identified social networks to protect their identity and privacy, some people choose to share anonymously without their real names attached to the content they post. Leavitt (2015) describes how Reddit users use anonymous throwaway accounts to disclose personal information (such as asking for advice or feedback about controversial problems), and to manage boundaries between their different accounts on Reddit.

To explore these phenomena as they might or might not apply to anonymous communication apps, we posed the following research question:

RQ 3. How do people choose to use anonymous communication apps versus identified social networks?

3.2 Method

We conducted eighteen semi-structured interviews with anonymous communication application users to better understand how and why people participate in these apps. The first thirteen interviews (P1-P13) were conducted in Summer 2014, and the last 5 interviews were conducted in Spring 2015 (P14-P18). The findings were not different across these two time periods, so the data have been combined. Participants were recruited through flyers posted in a major city in the east coast of U.S., a participant pool of a research university in the east coast, and through Craigslist postings. We recruited participants who had used at least one of the anonymous communication apps Whisper, Secret, and YikYak. The majority of our participants used Whisper.

Each interview session lasted approximately one hour and focused on how participants used the application, the posts they made and viewed, and their perceptions of the application dynamics and other users of the application. We conducted interviews in person (4 out of the 18 interviews) or over the phone, Gtalk, or Skype. Participants also sent specific posts mentioned during the interview via email for later reference during analysis and coding.

3.2.1 Features of the apps

All three applications were available on iOS and Android smartphones. Users could only post via their smartphones. Whisper had a website displaying popular whispers and various categories of whispers that people can access on their computers

(https://whisper.sh/). Users on both Whisper and YikYak were not connected to other users on any existing social networks. Posts on both Whisper and YikYak contained location information: each Whisper post displayed city level location information, and each YikYak post showed a blurred area map in the background. Whisper posts always had a background image and texts on top of the image. Each reply also included a background image. YikYak posts only contained text. Secret connected users with their friends and friends of their friends based on existing social networks such as email contacts. Secret posts that were not from a friendship circle were displayed with city-level location information. At the time when we conducted the first 12 interviews, Secret allowed users to add a background image but later they removed that feature.

Whisper randomly assigned a username to each user at initial signup, but users could change their usernames any time. The default setting for posting on YikYak did not require users to attach a username to their posts, and replies on YikYak were identified by random avatars. YikYak users had the option of editing a "handle" if they want to added an identifier to each post. Secret posts did not have usernames attached, and replies were identified by avatars randomly assigned to each user. Users' avatars changed every time they replied to a different post.

Users interacted with others in three ways: posting, acknowledging another's post ("hearting" a post on Whisper or Secret; "upvoting" or downvoting in YiYak), replying to a post (on Whisper, YikYak, and Secret), or sending direct messages to another person (only available on Whisper and Secret). Users could flag posts on all three apps.

3.2.2 Participants

Participants in the study reported ages between 19 and 29 (mean age 23.5 years); 11 were female participants and 7 were male participants. Our sample is typical of the demographic breakdown of Whisper, according to a prior count by Correa et al., (2015), who reported that Whisper users are between age 17 and 28 and 70% of them are women. Thirteen participants were students, and the other participants were two administrative assistants, a research specialist, and a hospitality worker. Seventeen participants used Whisper and 5 participants had also tried Secret (but they mainly talked about their usage of Whisper during the interviews). One user used the application YikYak. Nine participants also mentioned using other anonymous communication services during the interview, including TextSecure (a private messaging app), PostSecret, FML (http://www.fmylife.com/), Facebook confession board, and Reddit. Our sample used a variety of other social apps as well (16/18 used Facebook, 11/18 used Twitter, 11/18 used Instagram, and 8/18 used Snapchat).

In recruiting, we stratified our sample across usage duration with one third of our sample being new users (1-3 months in the app), one third being moderate duration users (4-8 months in the app), and one third long term users (12 months or more). We

also split our sample across users who had posted in the application (10/18) versus those who only browsed and read posts by others (8/18).

3.2.3 Analysis

We transcribed our interview audio records and analyzed the transcripts following a grounded theory approach (Corbin & Strauss, 2008). The transcripts were coded using Dedoose software (http://dedoose.com/). Within the interviews we identified descriptions of application usage, and posting and reading behavior. Two researchers independently coded subsets of the interviews, discussed and resolved differences, and clarified code definitions. We first open coded participant responses about their motivations for using the application, posts they made and viewed on the site, and social perceptions. We then performed axial coding, conducted affinity diagramming to group categories of concepts, and developed a series of themes on each topic: application use, types of posts people made and why, and types of posts they liked and disliked, motivations for using the application, perceptions of the other users of the applications and contrasts with other social media. We refined the themes during the writing process.

We supplemented our interview responses with qualitative analysis of posts on the applications. The posts we coded were from three sources: we randomly collected 125 posts from the applications Whisper and Secret in June 2014; we also coded the posts that our participants had posted on those applications; and we extracted the posts that our participants reported seeing on those applications. Three researchers worked together to code the posts according to their topic, to the emotion expressed, and to the motivations that seemed to be behind the posts. We used the categories from this analysis to confirm and extend the themes from our interview analysis, and to validate the types of posts that participants described in our interviews. Table 3 below summarizes the result of the integrated coding of posts we collected from the applications and posts mentioned during interviews.

3.3 Results

We found that participants interacted on anonymous communication apps to disclose predominantly personal information or emotions (positive and negative), and that they felt short-lived connections with other users in response to content or aspects of content they saw. In the following sections, we first present how people used Whisper, YikYak, and Secret, the types of posts they made in the applications, their motivations for making posts and what they got out of reading others' posts, and their perceptions of community on the app. Lastly, we discuss how participants viewed anonymous communication apps versus identified social networking sites.

3.3.1 How people used the apps

Anonymous apps provided a periodic distraction and diversion from everyday life for the participants, but also an emotional outlet without social consequences. A majority of our participants (browsers and posters) visited the app for five to fifteen minutes a day. They reported browsing and liking posts before they went to bed, or to pass time when they had some downtime throughout the day, e.g., when they were on the bus or waiting in line. They tended to look at popular posts, nearby posts, and occasionally to search for topics, such as music, technology, and fashion, in which they were interested.

Some participants in our sample reported their usage of the apps declined over time. Two longer-term users of Whisper mentioned that they visited the app very frequently when they first started using it (several times a day) but used it much less now (two to three times a week) because were irritated by seeing lots of rude argumentative or overly sexual posts (P17 and P18). Participants who posted frequently early on also said they posted less frequently over time, especially after getting unwanted responses to their posts. For example, P1 who posted when he first began using Whisper switched to only browsing because he received too many direct messages and requests for further contact in response to the posts he made.

3.3.2 What people posted on anonymous applications

We next examined the types of posts people made on the applications. Table 3 summarizes main categories of posts revealed in our analysis. Our findings echo prior research showing that the majority of posts were personal disclosures (Correa et al., 2015; Wang et al., 2014). Many of these posts expressed strong personal opinions, or shared personal experiences, confessions, and negative feelings such as anger or sadness (coded in Table 3 as distress release and social venting). Some posts served the purpose of identity clarification and did not always contain strong emotions (Table 3; self-expression). People also shared posts that were more lighthearted, fun, and entertaining (Table 3; entertaining confessions, positive stories). Many posts were made seeking responses (Table 3; seeking interaction) either in person, in replies and one-on-one messages, or chat in other channels. People also made impersonal posts with quotes, facts, or information about topical interests (Table 3, general entertainment, information sharing).

Types of posts	**Subcategories**	**Example posts**
Personal		
Distress release	Sad, negative emotions, confessions of bad behavior, bad experiences	"I was like I can't believe I wanted to get through school to work full-time." (posted by P17)
Social venting	Conflict, relationship problems	"You're extremely difficult to work with. I wish I could clone you make them your

		manager so you know how much you suck." (seen by P3)
Self-expression	Self-reflection about identity, personal opinions, aspirational	"I love being a solitary human being. I enjoy reading in a park, the library, at a bar, at a restaurant. I don't need many friends to have fun." (posted by P3)
Entertaining confession	Funny self-observations, funny habits, pranks	"I swear about my bosses all of the time in a language I know they don't speak. Their dog too." (posted by P10)
Positive stories	Overcoming challenges, achievements, positive experiences, news	"I've got five kids and I'm a single mom. I woke up to my oldest son helping his younger siblings clean their rooms and get read. Parenting done right." (seen by P5)
Seeking interaction	Meet-ups, asking for advice, question discussion starters	"I never know what to do with my hair – What's a hairstyle that works for curly hair?" (posted by P12)
Impersonal		
General entertainment	Non-personal jokes, quotes, observations	"Summertime and the living is easy" (posted by P16)
Information sharing	Interests, facts, local information	"I like Game of Thrones. I can't wait until next season." (seen by P2)

Table 3. Post categories identified in our study. (These categories are not mutually exclusive.)

3.3.3 Why people post on anonymous applications

Despite the lack of affordances for social identity or relationship development, participants had mainly social reasons for posting on the applications. Sometimes people explicitly sought social interaction, asking for replies to their posts or interactions outside of the app such as chatting or a "meetup." Sometimes they just wanted to share their personal stories or momentary feelings without any expectation of responses.

3.3.3.1 Using the crowd for social validation

According to particpants, a primary motivation for creating posts in the apps was to obtain social validation from the crowd of application users. They used the diffuse members of the application as a social litmus test of their behaviors, opinions, and admissions of frailty or unusual characteristics. They wanted to know whether their opinions or thoughts were normal, whether people would disapprove of something they had done, how bad it would seem to others, or how a wide set of people thought about an issue of personal relevance.

For example, a 22-year-old hospitality worker posted about being a solitary person (Figure 2; Table 3, self-expression category example post) and said she posted it on Whisper to see whether others were like that too. She said she always felt pressured to go out with friends and colleagues in real life but what she really enjoyed was reading

by herself. The hearts to her post made her feel like there were other people who felt the same way. She said:

> *It's nice to have some validation that you aren't the only one that feels that way.(P3)*

Despite the anonymity in the apps, participants sometimes sought opinions about their own behavior indirectly. Seemingly impersonal questions to the crowd or statements were another way to acquire opinions about the user's own behavior. For example, after P4 got a tattoo, she posted a question to get people's opinions about tattoos. The post read, "tattoos on a girl yes or no?" not revealing anything about herself or the fact that she had one. As she said: *"it's interesting to figure out what people think about that."*

Participants used the apps to get opinions from the (presumably) more diverse and more objective audience than they would find on their other social media. For instance, P6 said:

> *I guess anyone who has Whisper can check it so it's beyond your Facebook friends, 1,000 people you already know. So they might have other interests, might be cooler to like Lord of the Rings on Whisper than to like it on Facebook. (P6)*

Nevertheless, participants were aware that anonymity took away accountability and responsibility. Two participants said that they did not want advice from random strangers on the Internet and they would prefer advice from friends because they knew them better.

The social validation motivation for posting we observed seems similar to the approval and identity clarification goals described in the previous literature mainly among teenagers. Teenagers share personal information on their personal webpages for the purpose of seeking connection and validation (Stern 2008); we found young adults doing so on these apps. According to prior work, a goal of self-disclosure is to be liked or accepted by others, and previous work has found these disclosures help communicators clarify their identity by allowing them to convey accurate information about themselves (Baumeister, 1982; Derlega et al., 1979). In the anonymous applications, people shared their personal opinions, habits or stories even without sustained interaction. When the anonymous audience responded positively with hearts or upvotes, even with no further conversation, they felt they were not alone and that their behavior was acceptable.

3.3.3.2 Making short-term connections

Although it was unusual in our sample of interviewees, many posts on the anonymous applications were explicitly looking for interactions or "hookups" with other users. These could be online one-on-one chat interactions or meetings in real life. For example, P1 had used the app to talk to people online and to meet them in real life. He had also posted specifically to get responses from people who wanted to chat one-on-one, with a

post saying, "if anybody wants to chat, message me." It seems like most of these posts were seeking short-term communication or sex, but we do not have evidence to discover whether these interactions ever developed into longer-term, intimate relationships. Four participants, all female, expressed an aversion to this type of post, saying they were "spammy," "disgusting," or defeated "the purpose of the app."

Some of our participants responded to requests for contact. For example, P2 sent a direct message to another user in his local area asking if anyone had Game of Thrones DVDs and offering to buy them. Posts aiming at making connections may have a higher level of risk of exposing the poster, and these posts were more likely to include identifying information such as selfies (photos of the poster), or the poster's location or personal interests. For example, P7 saw a friend posting a selfie on a post that said "Anyone in [his city] want to talk or something?" using the "Nearby" feature of Whisper (Figure 3).

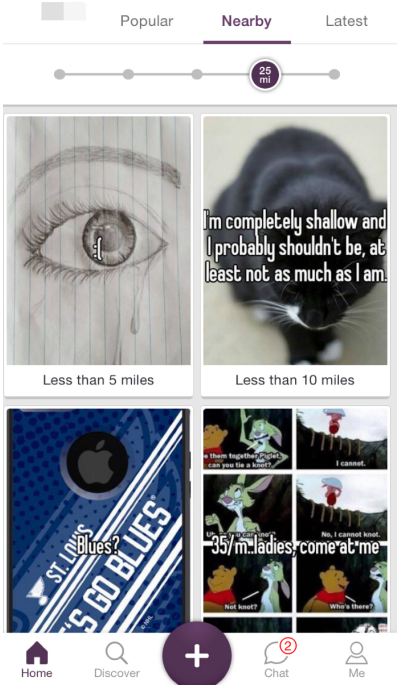


Figure 3. View of the Nearby tab in Whisper app

3.3.3.3 Avoiding social risk and context collapse

Participants also used the apps to release emotions or secrets without risking social consequences such as offending others, secrets being found out by others, or harming one's image online. In the social psychology literature, it has been reported that writing about a secret can improve people's health conditions (Pennebaker, Kiecolt-Glaser, & Glaser, 1988). Turkle (1995) has argued that the virtual world is a "safer place to ventilate" (p. 199). This motivation seemed to be prominent in the posts coded as distress release and social venting (Table 3). For example, P18 complained about work-

related issues on Whisper, and said she did not care about getting feedback or responses because she just needed to get her opinion out there.

The lack of social boundaries on Whisper meant participants did not have to deal with conflicting expectations from different social groups, or the context collapse problem in social media (Marwick & boyd, 2011). The above participant P18 indicated that she posted her complaints about her work on Whisper rather than other social media because her coworkers were friends on the other social media she used (Facebook and Instagram). This motivation is similar to what Kang et al. (2013) (chapter 2) reported about the motivations to seek anonymity. They reported that people used anonymity as a way to manage overlapping and difficult social boundaries in their lives, eschewing all of them by being completely anonymous. An administrative assistant talked about freedom from differential audience expectations on Whisper versus other social media:

> *I didn't really have an audience in mind. It's one of the things I like about Whisper is that I don't have to think about that. When I post to Facebook, et cetera, because I have so many family members and then people that I'm sort of in touch… I have to always make really careful decisions about what's public versus what's for just these people. (P10)*

In addition to removing the concern of managing one's social boundaries, sharing on anonymous communication apps, as implied in the above quote, did not require as much effort to construct or polish one's post as did posting on identified social networking sites.

Participants said they did not need to consider how to manage their online image on these apps. P1, who also used Tinder to find connections, said he cared less about his grammar on Whisper as compared with Tinder; when he talked to others on Tinder, he would phrase his words better. A woman who had shared her story about an unrequited love to PostSecret thought posts on Whisper were easy to make, in contrast to the effort required when she had to buy a post card, write her secret down, and mail the post card to PostSecret (P9).

3.3.3.4 Sharing momentary information

Some participants posted to share feelings or thoughts they had that were stimulated by the situation they were in at that moment. For example, P15 posted a joke in frustration while she was driving and stuck in traffic because of construction she saw everywhere on the road. Her post read: "In [her town] it's either winter or construction." She said she wouldn't post something like that on Facebook because it wasn't important enough, expressing that some things are not "monumental enough to post on Facebook."

These in-situ posts often involved observations or thoughts about people around the poster that couldn't be shared or said out loud. Frustration was a common driver for these in-situ posts. They also included posts about conflicts or arguments with other people (Table 2, social venting). The administrative assistant also posted after a drunken

fight with her husband, saying that, "I'm pretty sure I made a mistake marrying my husband but I have two more weddings to get through before I can get a divorce." She appreciated the anonymity of this post the next morning because there were no real consequences. She said:

> *In the morning I look at this and even though I know it doesn't matter because nobody's seen it I'm just kind of mortified that I even thought that… But that's kind of why I like Whisper is you can post something like that and there's not really consequences so to speak since it is anonymous.(P10)*

3.3.4 How people view anonymous communications

The majority of our participants used the apps for browsing posts made by others. We were interested in why they enjoyed reading posts from others and what benefit they got out of reading these posts, given the barriers of anonymity and ephemerality to forming relationships with the posters. We found that participants' motivations for reading posts fell into three main categories: connection, entertainment, and social comparison. We also noticed aversion towards offensive content people saw on the site.

3.3.4.1 Feeling connected to people like me

Our participants liked the posts that sparked feelings of similarity or empathy, and estimated that the posts they could resonate with were more likely to get popular. These posts were about funny habits, similar life experiences, or similar problems they had gone through or were experiencing (such as the loss of a close relative, being saddled with student loans). Three participants mentioned they liked to see people overcoming difficulties in their lives that seemed relevant to their own lives. For example, P5 liked a post by a single mom about raising five kids (Table 2, positive stories example post) because she had a big family and could empathize with how hard that situation must be:

> *It's nice to see that people are fighting through things, and like in this situation that would really stink. I'm a child of four, and so I know how crazy it gets just from both of our parents. So then it's like someone with five kids I'd be like wow. I'd give them props.(P5)*

People also resonated with objective topics that signaled a common social identity outside of the app, such as posts about the school attended, local news, or local restaurants. Participants did not restrict their activity or viewing to a single topic or interest, but especially liked posts that connected with their own interests. Some participants mentioned they were more attracted to posts about music, fashion, technology, or TV shows that matched their own interests. On the other hand, the anonymous apps did not seem to serve interest-based information needs very well. Two participants mentioned searching for things they were interested in but failed to find satisfying results.

3.3.4.2 Being entertained

Participants enjoyed reading entertaining or sensational posts made by others for their amusement or shock value. Several participants enjoyed reading personal negative confession posts for their shock value (Table 2, distress relief and social venting). For example, P18 said: "it's like a reality TV on your phone." P13, an undergraduate student majoring in engineering, said: "It's kind of a madhouse. Like, people will post some of their deepest, darkest secrets." He liked reading about people's "life-changing problems" because he believed people were sharing the actual problems they had.

Half of the participants said they enjoyed reading more lighthearted confessions and entertaining posts for amusement. For example, P11 said:

> *A lot of people just post these really intense confessions, or things like that. And it's kind of good to see that people can also just post these really funny things on it.(P11)*

3.3.4.3 Aiding downward social comparison

Distressing posts also stimulated downward social comparison or schaudenfreude (Wills, 1981). Two participants (P5 and P11) said learning about other people's problems helped them put their own lives into perspective, and made them feel more motivated to handle their own problems.

> *It kind of put things in perspective for me... There were times where I would be like oh my gosh in my life. And then I would see other people and their issues, and then in that respect I would be thankful for what I have, and like myself. (P5)*

3.3.4.4 Disliking misinformation and offensive content

An obvious drawback of anonymity is that it lacks social (and legal) accountability for rumors, misinformation, and offensive content. Several participants complained about seeing posts that were offensive to women, posts that included nudity or overly sexual content, and prejudicial posts about people's religion or race. As in Turkle's work (2011), a few of our participants doubted the truthfulness of posts and mentioned that stories they saw on Whisper were farfetched and did not seem realistic. The engineering student (P13) thought popular posts he saw were "presented in a way that can grab your attention." P4 said, "some people might just exaggerate details to get people to feel sorry for them."

3.3.5 Perceptions of identity and interaction

Users in the applications could not associate individuals with posts most of the time and posts did not persist for long periods. It was often difficult to find a particular post seen in the past. Yet our analysis of posts and responses to posts revealed, perhaps counter-intuitively, that participants felt a connection to the messages and emotions of other users. sometimes motivating direct interaction.

3.3.5.1 User identities

Our participants perceived other users of the applications to be mostly teenagers or young adults. Some described user identities using specific demographics such as "LGBT group, female." Most of them made this estimation based on the type of content they saw in the app. One said, "It's a lot of people talking about things like hooking up and like meeting people out at bars and stuff. It makes me think of a younger population." (P18) The YikYak user guessed there were other college students on the application because of her location (a college neighborhood) and the fact that the application used her location to show nearby posts.

Location provided an identity signal unique to these apps in comparison to other web-based online communities. Posters could use the "nearby" feature of Whisper and YikYak to focus their attention on the people geographically close to them. This shared identity increased the level of informational or emotional support they gained from interaction. There were other ways of signaling identity through posts, such as by indicating special interests or affiliations. For example, one participant posted about a local baseball team, using a phrase that only fans would understand, signaling he was a fan and intended to elicit responses from other fans (P16).

We asked participants to estimate how anonymous they were on the site, and most of them were quite confident that they were anonymous. One exception was the woman who also used PostSecret; she thought Whisper was not as anonymous as PostSecret because Whisper posts could potentially be tracked electronically. The engineering student (P13) was suspicious about all anonymous sites and took additional steps to anonymize his identity such as using a Burner app to create temporary phone number for Tinder or Craigslist and using an app TextSecure to send encrypted messages.

Self-deanonymizing and deanonymizing others were frowned upon by our respondents, and not something they had done through their posts in the app. One participant (P7) did deanonymize a friend on Whisper when he recognized him, later deleting his identifying reply because the original poster became upset after being publicly deanonymized and he did not want to hurt his friend's feelings. Our participants observed other users occasionally deanonymizing themselves by posting selfies as the background of their posts, typically in posts requesting in-person meetups.

3.3.5.2 User interactions

Participants mentioned seeing a variety of replies ranging from supportive and encouraging messages, suggestions, to extreme opinions, criticism, and offensive comments. P7 said Whisper was similar to a "giant psychologist" whereby people unload their problems and others console them. People seem to bond with others based on the similar problems they have experienced or on their shared interests, but this connection may be weak and short-lived. As described earlier, hearts (or upvotes on YikYak) could signal validation or support, and people tended to heart (or upvote) posts that they felt were relevant to them or funny posts they liked. However, the meaning of

"heart" could be ambiguous (similar to the ambiguity of likes on FCB [Birnholtz et al., 2015]). P11 said she usually hearted funny posts on Whisper, but would feel awkward if she hearted sad posts.

It can be challenging to keep one's interactions appropriate while using anonymous communication apps because of the disinhibiting effect of anonymity. The engineering student noted, regarding some spiteful replies he saw to a woman's post: "These people were just digging into this woman and showing absolutely no remorse" (P13). Replies on Whisper usually consisted of personal opinions, which sometimes could be considered as crossing the line of appropriateness, especially because they were coming from complete strangers. For instance, P4 posted about her relationship problems on Whisper, but erased her original posts after getting inappropriate responses from other men. She said:

> *They were just people putting their two cents in about things that I didn't post about. People were saying, "Oh, he must not be making you happy in other ways," and things that I didn't say anything about, so just people looking way too far into the situation. And this feels really personal really fast for things that I wasn't asking for. (P4)*

Whisper allowed people to send direct messages to others, usually initiated by posts. Some people used them to seek further interaction such as meeting in real life. These interaction requests sometimes drove people away when they were unwanted or inappropriate. The college student who was looking for responses on her tattoo received replies asking for her ASL (age, sex and location), and pictures of her tattoo. Eventually she stopped replying to the messages when they got "too personal" (P4). P1 enjoyed messaging others on Whisper, but did not pursue further interactions when a 30-year old guy asked him to hang out.

3.3.6 Comparison with other communities

Each anonymous communication app had some unique characteristics that were different from other anonymous applications or communities and had different appeal. Our participants reported being anonymous on other systems like reddit, Tumblr, Facebook Confession Board, FML app, and PostSecret but sometimes for different reasons. Three Whisper users also used Tumblr, but they thought Whisper was designed for people to vent about their problems, whereas Tumblr was for entertainment purposes and sharing impersonal fun (such as fandom and artistic content). P3 noted that she has Tumblr friends in real life and even meets them or connects with them on Facebook, but she would never do that with other users on Whisper. P11 thought Whisper was similar to FML in serving the purpose of making people feel better about their lives by looking at others' problems, but the stories on FML were more extreme, and probably only happened to one person, whereas the posts on Whisper were more common.

All but two participants we interviewed had Facebook accounts. Four participants mentioned that they used Facebook much less frequently recently and rarely shared their personal status. They considered Facebook as a place for the "older generation," and they only shared big events or links and photos on it. On average, each participant used three other social sites or apps. When we asked them to compare the anonymous communication apps with identified communities, participants said anonymous posts were more personal, more open, and more honest – "you can be yourself because there's no retribution (P9)." P1 thought people were more honest and truthful on Whisper because they did not need to worry about managing others' impressions of them. Participants thought anonymous communication app communities tolerated different religious and political beliefs than the more identified communities they were in and had a more diverse audience. As in previous research (Morris et al., 2014) and the tradeoffs mentioned by participants described in Chapter 2 (Table 2, p27), our participants reported that the feedback they received on Facebook was more personal and in-depth than anonymous feedback (P2). Some people said they saw similar types of posts and interactions on Whisper and Facebook or Twitter, such as quotes, and encouraging responses like "good job."

3.4 Discussion

This study shows that many people share their personal opinions, experiences and confessions with others on anonymous applications, for the reasons of seeking or providing social validation, building connections, avoiding problems of context collapse and impression management on identified social media, and sharing momentary feelings. A lack of accountability can mar interactions in these apps, but the unique interaction patterns and the benefits people gain from using these apps suggest design implications for online communities and social networking sites.

3.4.1 Exchange social support without identification

The main anonymous feature on these apps is that there is usually no consistent username or handle. Unlike previous work (Turkle, 1995) in which users adopt different identities with different personalities in the virtual world, people cannot build consistent identity or reputation on these apps. There is no need to play an "ideal self" on these apps since there is no reputation or personal history. People likely use these apps to disclose fragments of their multifaceted identity or the unconventional parts of themselves. It is known that self-disclosure increases intimacy among group members [8, 37]. We also found people easily bond with a complete stranger on these apps when they share similar experiences or feelings.

A primary reason that people post and view anonymous posts is to gain social validation and social support from the community. Our finding suggests that social support and social approval can be provided by strangers in anonymous communities.

This finding strikingly resembles early research on Usenet nearly 20 years ago (Wellman & Gulia, 1998) – online groups often are supportive in nature. Hearts and upvotes in these apps are lightweight methods to signal validation and connection. However, we do not know whether or not the connections users find on these apps would evolve into longer-term, intimate relationships. Even the veteran users we interviewed have only used the apps for about a year. Future research should examine whether or not this temporary support and connection lasts, and whether or not it contributes to users' well-being in the long run.

3.4.2 The ephemerality of anonymous communication

The new form of location-based interaction on smartphones enables momentary and in-situ information sharing that was previously impossible in web-based communication. Our finding suggests that some user-generated content might be ephemeral by nature and could be shaped by the design of the communication media. For instance, our participants reported that they seldom went back to find previous posts they have made on Whisper or YikYak. An important goal of self-disclosure on identified social networks such as Facebook is to keep a record of personal history for one's own use (Vitak & Kim, 2014; Zhao et al., 2013). These anonymous communication apps do not seem to support this purpose. P9 compared Whisper to reddit, and pointed out the lack of history in Whisper:*" It's kind of meteor flashes in the pan."* The fact that people's identities are not tied to the content they created also makes it impossible to establish reputation or history as in other online communities. The recent demise of Secret (Constine, 2015) raises the question of the sustainability of these kinds of communication apps. Post ephemerality might contribute to decreased user engagement: without a reputation attached to a community, it is very easy for user to quit and join other communities.

The lack of identity association, however, also reduces the burden of generating and sharing content, which could lead to more content generation. People sometimes post very momentary feelings, and these types of content may not need to stay long on the Internet. Snapchat's feature of allowing user to set how long their content lives before they share it with their friends seems to support this need for ephemerality. An important design question is: can people accurately estimate how long their data need to persist before they share the content?

3.4.3 Mitigate negative interactions

The challenge of introducing anonymity into other online communities is how to mitigate negative interactions such as offensive or sexually explicit content. Birnholtz et al (Birnholtz et al., 2015) shows that using anonymity in a localized identified community (FCB) only produces a small amount of negative comments, but the responders on FCB are identified by their own Facebook profile. We could expect the amount of negativity to increase if the identities of those who responded are removed. A

possible solution is to use identity signals to mitigate the negativity of anonymous communication, such as location, affiliation, or domain of interest. Our finding also suggests that shared identities such as location or school might motivate more positive interactions in anonymous communications.

Future work could conduct experiments to examine the relationship between the level of identification and the amount of negativity people receive in anonymous or quasi-anonymous online communication environments. Some existing communities already give users the choice of temporary anonymity such as Quora (an online Q&A site), and a significant minority chooses to be anonymous when replying or following certain questions (Peddinti, Korolova, Bursztein, & Sampemane, 2014).

3.4.4 Limitations

A limitation of this work is that we did not get rich enough data about negative interactions on these apps. A few of our participants talked about their discomfort when received unwanted connecting requests or harsh comments, but we did not talk to anyone who has sent these messages. Social desirability bias might make this problem challenging for qualitative research: people tend to under-report behaviors that may be viewed unfavorable by others. To examine these particular types of behavior, future research could use other methods such as textual analyses to examine the characteristics of negative or undesirable interactions on these apps or on other communities such as Quora or Reddit. The result of these studies could help build automatic content filters of these apps.

Another limitation is that we have a comparatively small sample size and most of our participants lived in the same city. Although our sample matches the demographic characteristics of Whisper users described in other larger scale study (Correa et al., 2015), our small sample size may prevent us from detecting all possible reasons that motivate people to use these apps. Individual demographic differences such as gender identity, profession, and location may also influence people's behavior and motivations. These questions can be more appropriately answered in a larger-scale quantitative study using methods such as online survey.

Part II. Managing privacy threats to personal information online

In 2013, a University of Pittsburgh researcher murdered his wife using cyanide. Some key evidence presented at the trial was that his Google search history contained multiple searches for "cyanide" (Daley, 2014). The defendant also searched for information on how to remove his computer search history, but obviously he did not succeed in hiding his search traces. In this case, a person's failure to use the correct strategy to hide worked to the advantage of law enforcement and news audiences; it also demonstrated the technical barrier lay people confront when choosing an appropriate strategy to hide their information.

In the second part of this thesis, I present three studies examining the relationships between people's background (including their technical knowledge and Internet experience) and their perceptions of privacy threats, and how these background and perceptions influence their decisions about managing their personal information online. One persistent finding is the lack of strong evidence that technical knowledge influences how people manage privacy threats to their personal information online. Technical knowledge gives people the capacity to understand privacy threats and to use privacy tools, but they often do not use this capacity to a greater extent than do ordinary lay users. Thus, it is not uncommon to find that people with technical knowledge have damaged themselves by what they revealed about themselves online, just as the Pittsburgh researcher did ("Software engineer hooked on child porn jailed for three years," 2008).

4 How people perceive and manage online privacy threats[3]

4.1 Introduction

The findings described in previous chapters are that many Internet users try to control access to their online personal information in numerous ways. By personal information, I mean not just personal demographic data such as age or home address, but also people's posts, interactions, and communications with others online. I define privacy threat as any risk to the loss or misuse of personal information by other individuals or organizations, or entities. In this chapter, I describe how people conceptualize and use different strategies to manage threats to their personal information online, specifically in regard to whom they want to hide their information from.

Understanding how different Internet users manage their personal information online will help us improve the design of privacy-enhancing technologies and policies. We need to help people manage risks to their interpersonal relationships and risks to the misuse of their personal information. The adversarial threat models used in security-related research emphasize software adversaries that threaten personal information (e.g., the STRIDE threat model by Microsoft) but we also need to understand social privacy concerns such as risks to reputation and damage to people's relationships (e.g., Woodruff 2014).

In the general model for this thesis, three factors associated with people's individual background may be important in perceived privacy threat. These factors are people's enduring social orientation (derived from personality and culture), their own past Internet experience, and their technical knowledge. Here I use empirical data from two

---

[3] This chapter is based on the datasets described in:

Rainie, L., Kiesler, S., Kang, R., & Madden, M. (2013). Anonymity, Privacy, and Security Online. Pew Research Center, 35. Retrieved from http://www.pewinternet.org/Reports/2013/Anonymity-online.aspx

Kang, R., Brown, S., Dabbish, L., & Kiesler, S. (2014). Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. In Proceedings of SOUPS 2014 (pp. 37–49). USENIX Association.

surveys I have conducted with colleagues to show how these factors predict people's threat perceptions and what strategies they use to mitigate different threats. I also compare behavioral versus technical coping strategies people use to manage different threats and how people with different levels of technical knowledge adopt those strategies.

4.1.1 Demographic characteristics

The demographic characteristics of a group of people predict their attitudes. For instance, younger people tend to be more politically liberal than older people, and tend to have more concern about privacy and supports of freedom from surveillance. Younger people take more privacy-protection strategies, and men take more actions than women (Dommeyer and Gross, 2003). Because social media tends to elicit personal information from people and increases people's awareness of their information being exposed, using social media should predict more concerns about privacy as well.

4.1.2 Social orientation

Individuals' orientation to their social world varies within and across their social and cultural environments, and shapes the way they think about and act to protect privacy, mainly by influencing their desire to manage boundaries in lives. Prior literature suggests that the collectivism vs. individualism distinction is particularly important in distinguishing individual's social orientation (Brewer & Chen, 2007; Hofstede, 1984). Triandis (1989) described collectivistic cultures (e.g., Asian countries) as those that socialize people to develop a more public and integrated identity whereas individualistic cultures (e.g., North American countries) socialize people to develop a more private and independent identity. Collectivists have a sense of responsibility to share information for the good of their group or company, even if doing so is potentially disadvantageous and harms individual privacy. By contrast, individualists share information in their personal interest, and what they share depends on their assertiveness and personal choice (Chow et al. 1999). Another social orientation that should be important in how people perceive and treat privacy threats is whether they have a more or less segmented identity. Those who desire to segment their social lives, presenting a different "self" to different groups, would be particularly threatened by publication of personal information or leaks of their online interactions across groups. They would be expected to mitigate the threat by hiding content from certain groups.

The collectivist vs. individualist distinction and ideas surrounding faceted identity leads us to the following predictions.

H1a. Internet users whose social orientation is more individualistic or less collectivistic will be more likely than others to identify social privacy threats.

H1b. Those whose social orientation is more segmented will also be more likely than others to identify social privacy threats.

4.1.3 Prior negative Internet experience

Prior negative experiences on the Internet are likely to influence people's perception of privacy threat. Shay et al. (2014) report that 30% of an MTurk sample and 15% of Google Consumer Survey respondents had experienced unauthorized access to their email or social networking accounts. Attackers include both unknown groups and known social ties. Research shows that having experienced privacy invasions on social media motivates people to take more actions to protect their privacy on those sites. Litt and Hargittai (2014) examined how a variety of negative experiences which they called "online turbulence" affects people's behavior managing their personal information online, but they mainly focused on turbulence to people's social relationships such as trouble with friends or parents. The interviewees described in Chapter 2 mentioned a variety of negative experiences that motivated them to seek anonymity, such as having been criticized or stalked online, or experiencing computer attacks and security breaches. Therefore we hypothesize prior negative experiences affect people's responses to both social and informational privacy threats:

H2. Internet users who have experienced negative events online will be more likely than others to identify privacy threats.

4.1.4 Technical knowledge

People with more computer-related technical knowledge might have a greater awareness of the different ways that personal information can be accessed by others. Conversely, less technical knowledge might lead to less awareness of how the Internet works. One of the interviewees in Chapter 2 said, "*I have no clue where [personal information] goes or how people could access it.*" (p.23) Much previous work mentioned in Chapter 1 show the effect of technical knowledge on people's privacy perceptions and behaviors. We therefore hypothesize:

H3. Internet users who have more technical knowledge will be more likely than others to identify privacy threats.

People's general privacy concern has been studied a lot, and is shown to influence people's perception of privacy and their behaviors (Joinson, Reips, Buchanan, Schofield, & Carina, 2010; Smith et al., 2011). Because it is not the focus of this chapter, we use general privacy concern as a control variable in the following analyses.

4.2 Method

With co-authors, I conducted two survey studies of privacy and anonymity, one a representative telephone sample of U.S. Internet users and the other, a few months later, an online survey of MTurk workers. Most items for both surveys were the same. Because the surveys given to the representative U.S. sample were conducted by phone with voice responses, and the MTurk surveys were conducted online, with typed responses, the response options were never identical. However, as much as feasible, the questions themselves were identical. The survey questions we analyzed in this chapter are attached in Appendix I.

The first survey was administered by the Pew Research Center's Internet Project (referred to here as "Pew") in July 11-14, 2013. We collaborated with Pew researchers on constructing questions for this survey. The survey items were developed based on the interview questions about anonymity in Chapter 2 and questions on privacy that the Pew Research Center fielded in its previous surveys (Madden & Smith, 2010; Madden, Fox, , Smith, 2007; Pew Research Center, 2013). Pew surveyed a representative sample of U.S. adults consisting of 1,002 U.S. adults ages 18 and over, with 500 surveys using landline telephones and 500 surveys using cell phones. Respondents were not paid, except any cell phone charges were reimbursed. When conducting the survey, interviewers asked respondents if they would be willing to participate in a confidential and anonymous survey. Participants were then asked a series of questions, first to determine if they were Internet users, and then about their activities online. Of the total participants, 775 said they used the Internet and our analysis is based on responses from these Internet users.

The second survey was conducted on Amazon MTurk (www.mturk.com), a crowdsourcing platform, from February 16 to 20, 2014. We recruited 418 people using the same sampling criteria as in previous studies to increase quality (Kelley, 2010; Paolacci, Chandler, & Ipeirotis, 2010), by restricting the participants to those with an approval rate of at least 95% and at least 100 approved HITs. Each participant was paid $1 for completing the survey. They were told that the survey was about how people use the Internet. Separate HITs were released to recruit participants from the U.S., India and other countries. After accepting the HIT, MTurk workers were directed to a SurveyMonkey survey. The survey was completely voluntary and confidential. Participants could opt out of the survey at any time. Twenty-two responses (5%) were excluded because they failed the attention check questions or entered invalid responses. To rule out potential confounding variables such as different government policies in different countries, the analyses shown in this chapter only include the 775 U.S. public sample and 182 U.S. MTurk users. A comparison of the two samples is shown in Table 4.

Demographic characteristics	**U.S. Public**	**U.S. Turk**

N	**775**	**182**
Age		
18-24	12%	24%
25-34	14%	41%
35-44	13%	23%
45-54	17%	9%
55-64	24%	3%
65+	19%	1%
Mean age	49.8	32.7
	F [1,955] = 133.94, $p < .001$	
Gender		
Female	50%	42%
Male	50%	57%
	X^2 [N = 956] = 2.89, $p = .09$	
Education		
High school or less	26%	12%
Some college	31%	45%
College and more	42%	43%
	X^2 [N = 955] = 20.02, $p < .001$	
Percent who use social media	68%	90%
	X^2 [N = 957] = 35.57, $p < .001$	

Table 4. Demographic characteristics of two survey samples: U.S. telephone representative sample (referred to as U.S. public), U.S. Turk sample. Total N = 957.

The survey presented a series of questions related to people's social orientation, their experiences on the Internet, and their computer and Internet knowledge (independent variables in the models), demographic information (control variables), worry about information (control variable), and privacy protection behavior (dependent variables). We do not report on a few questions that we asked in the survey that are not relevant to the topic of this chapter.

4.2.1 Protecting their personal information

We asked people about behaviors they used to protect their personal information online. We also asked about their general privacy concern about information using this question: "Do you ever worry about how much information is available about you on the Internet?"

Both surveys also asked respondents whether they had tried to hide their identity online: "Have you ever tried to use the Internet in a way that hides or masks your identity from certain people or organizations?" Those who answered "yes" to this question were coded as having tried to hide their identity.

Internet users may be differently concerned about protecting their personal information when they communicate with different groups. To study whether respondents were selective in hiding content (such as posts and other personal information) that they had communicated online, the national sample Pew survey asked participants "Have you ever tried to use the Internet in ways that keep ___ from being able to see what you have read, watched or posted online?" They were asked if they had done this to "family members or a romantic partner;" "certain friends;" "people from your past;" "an employer, supervisor, or coworkers;" "the companies or people who run the website you visited;" "hackers or criminals;" "law enforcement;" "people who might criticize, harass, or target you;" "companies or people that might want payment for the files you download such as songs, movies, or games;" "advertisers;" "the government?" In the MTurk survey, we slightly modified the format and asked people the same question for each of the five groups: family, friends, co-workers; employers and supervisors; unwanted ties; authorities; and other third-parties.

In the Pew survey, we asked whether people have used 11 strategies to hide their digital traces: "While using the Internet, have you ever done any of the following things: used a temporary username or email address; used a fake name or untraceable username; given inaccurate or misleading information about yourself; set your browser to disable or turn off cookies; cleared cookies and browser history; used a proxy server, Tor software, or a virtual personal network; encrypted your communications; decided not to use a website because they asked for your real name; deleted or edited something you posted in the past; asked someone to remove something that was posted about you online; used a public computer to browse anonymously?" This list of strategies was generated based on interview results from the previous study described in chapter 2.

In the MTurk survey, we asked this question for each group of people or organizations that the respondent said he or she has tried to hide from: "Which of the following methods did you use to prevent ___ from seeing what you have read, watched, or posted online?" with the same list of strategies to select from. We repeated the strategy question at the end of the five threats questions to make sure everyone had seen this question even if they answered "no" to all five questions about hiding from the five groups.

4.2.2 Social orientation

To test Hypothesis 1a and 1b, we adapted existing scales to measure three types of social orientation in the MTurk survey: collective identity (Brewer & Chen, 2007), individual identity (Brewer & Chen, 2007), and segmented identity (a combined scale from self-

monitoring in Snyder & Gangestad, 1986 and faceted life in Farnham & Churchill, 2011), as shown in Table 5. These questions used 5-point Likert scales (1 = "strongly disagree," 5 = "strongly agree"). These three factors accounted for 56% of the overall variance (using varimax rotation, eigenvalue for three factors is 1.12).

Social orientation items	α	Factor loading	Mean (s.d.)
Collective identity	0.69		3.15
In general, belonging to social groups is an important part of my self-image.		0.79	(.78)
The social groups I belong to are an important reflection of who I am.		0.69	
To me, pleasure is spending time with others.		0.51	
My happiness depends very much on the happiness of those around me.		0.41	
Individual identity	0.48		3.85
I often do "my own thing".		0.72	(.68)
I enjoy being unique and different from others in many ways.		0.36	
Segmented identity	0.78		2.85
In different situations, I often act like very different persons.		0.85	(.77)
I'm not always the person I appear to be.		0.83	
I guess I put on a show to impress or entertain others.		0.44	
I have parts of my life that are really very different from each other.		0.64	
I would probably make a good actor.		0.31	
I prefer to keep different parts of my life separate.		0.53	

Table 5. Measures of social orientation (using varimax rotation, eigenvalue for three factors is 1.12, accounting for 56% of the overall variance).

4.2.3 Negative Internet experiences

To test Hypothesis 2, we asked respondents in both surveys if they had experienced any of ten different negative experiences online in both surveys: been stalked or harassed online; something happened online that led you into physical danger; experienced trouble in a relationship between you and a family member or a friend because of something you posted online; lost a job opportunity or educational opportunity because of something you posted online or someone posted about you online; had your reputation damaged because of something that happened online; had important personal information stolen such as your Social Security Number, your credit card, or bank account information; been the victim of an online scam and lost money; had an email or social networking account of yours compromised or taken over without your permission by someone else; had your personal information leaked by a company; got into trouble with local authorities or government because of your online activities. The list of negative experiences is generated from instances mentioned by participants in Chapter 2 study. To simplify the analysis, we used presence of any negative experience as an independent variable in the following analyses.

4.2.4 Technical knowledge

To test Hypotheses 3, we measured respondents' computer and Internet knowledge (but only in the MTurk survey) using their self-rated familiarity with nine technical terms on a 5-point scale (IP address, cookie, encryption, proxy servers, SSL, Tor, VPNs, privacy settings, and privacy browsing modes in browsers), and eight true/false questions about security and anonymity knowledge (e.g., "No one, except for the sender and intended receiver, can reveal the content of an encrypted email."). We developed the knowledge questions by consulting domain experts in computer security and tested their reliability with two independent samples. We combined respondents' self-reported familiarity with technical terms and their accuracy on the true/false questions (r = .55) to provide our measure of technical knowledge.

Additional items asked for demographic information such as nationality, gender, age, employment, and level of education.

4.3 Results

First we looked at the demographic characteristics of the Pew sample and the MTurk sample. Consistent with previous studies (Berinsky, 2012), our MTurk sample is much younger (MTurk mean age: 32.7; Pew mean age: 49.8) and has more male than female respondents (MTurk: 57% vs. 43%; Pew: 50% vs. 50%). The MTurk sample is also much more likely to use social media (MTurk: 90%; Pew: 68%). MTurk sample is more likely to have experienced negative experience than the Pew sample (MTurk: 49%; Pew: 36%).

4.3.1 Hiding identity and hiding information from specific groups

One purpose of these surveys is to find out how prevalent is anonymity seeking online among U.S. Internet users. Table 6 shows the percent of people from both surveys who have tried to hide their identity online. About 17% of the Pew sample reported purposely trying to hide their identity online. We found an even higher percentage of anonymity seekers among U.S. MTurk workers (31% vs. 17%, t [939] = 4.30, $p < .001$).

We also asked whether respondents try to hide their online contributions or content selectively, from different groups such as friends or employers. In the Pew survey, more than half of the entire sample had hidden content from at least one individual or group, but their implicit threat models differed by virtue of the different categories of people or groups avoided. Again, significantly more participants in the U.S. MTurk sample reported having tried to hide content from at least one group than in the Pew sample (73% vs. 53%, t [955] = 4.94, $p < .001$). In Table 6 we listed five groups of threats and the percent of Pew and MTurk respondents who have reported hiding from each threat.

	Pew sample	**MTurk sample**

Percent who have tried to hide identity	17%	31%
Percent who have tried to hide content or interactions from at least one group	53%	73%
Hide from social threats	32%	65%
Hide from family; friends; coworkers	20%	54%
Hide from employer	10%	27%
Hide from unwanted ties (people who might criticize or harass, and people from the past)	22%	27%
Hide from informational threats	47%	37%
Hide from authorities	10%	18%
Hide from other third parties	44%	28%

Table 6. Percent who have tried to hide their identity and percent who have tried to hide from different groups

As shown in the above table, more MTurk respondents reported hiding from social threats than from informational threats (65% vs. 37%), whereas more Pew respondents reported hiding from informational threats than from social threats (47% vs. 32%). I want to note that the different ways we asked those questions might bias people's responses. In the Pew survey, we asked people about whether or not they have tried to hide from the 11 people or organizations one by one without inserting any other question in between. In the MTurk survey, people were asked about what strategies they used to hide from each audience after they answered "Yes" to whether or not they have tried to hide from each specific group.

4.3.2 Strategies people use to hide information

Eighty-one percent of Pew respondents and ninety-six percent of the MTurk respondents had taken at least one action to hide their information online. When analyzing the data, we categorized nine of strategies we asked people in the survey whether they have used to hide their information into the following four categories: *mange cookies* ("set your browser to disable or turn off cookies"; "cleared cookies and browser history"), *use alias* ("used a temporary username or email address"; "used a fake name or untraceable username"; "given inaccurate or misleading information about yourself"), *edit content* ("deleted or edited something you posted in the past"; "asked someone to remove something that was posted about you online"), and *use technical methods* ("used a proxy server, Tor software, or a virtual personal network"; "encrypted your communications"). Because there were no differences across groups in whether people used a public computer to hide their identity and or said they had decided not to use a website because it asked for their real name, those two items were dropped from further analysis.

Table 7 shows the percent of respondents from each survey who had reported using each type of strategies. The most commonly used strategy by both samples is managing cookies. Respondents might have believed that managing their privacy in a local application protected their privacy at all levels of the network. The popularity of these approaches might have been due to their comparatively high usability rather than because respondents thought they were highly effective.

	Pew sample	MTurk sample
Percent who have used at least one strategy	81%	96%
Percent who had ever managed cookies	72%	88%
Percent who had ever used alias	35%	77%
Percent who had ever edited online content	42%	57%
Percent who had ever used technical methods	24%	42%

Table 7. Percent who have used each category of methods to hide their interactions online

In the MTurk survey, we were able to ask which methods people use to hide from each specific threat. Figure 4 shows what kind of strategies respondents used to hide from each threat. From the figure, we see that participants were most likely to use the strategies of managing cookies (including clearing a browser history) and using alias to protect themselves from almost all privacy threats. Those who had tried to hide from informational threats (including authorities and other third parties) were more likely to use technical methods, but less likely to edit content than those who hid from the three types of social privacy threats.

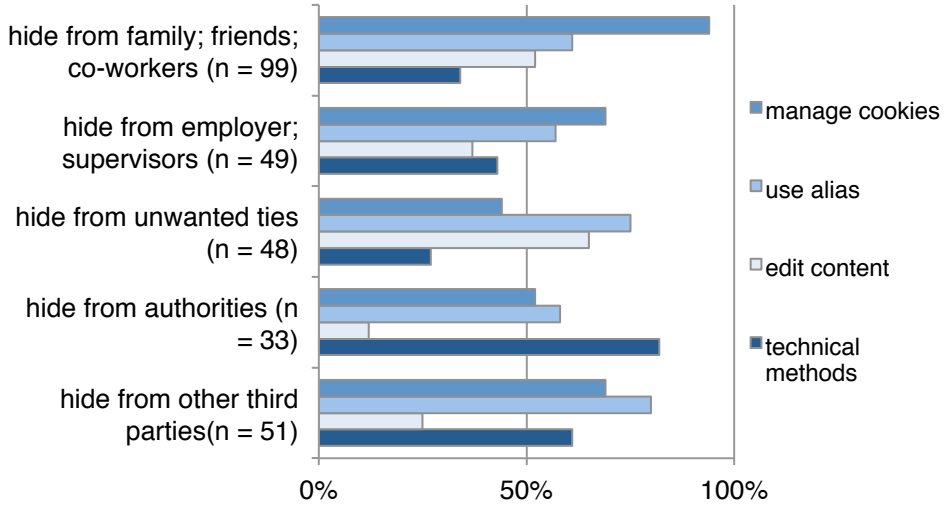


Figure 4. Percent of respondents who used each category of strategies to hide their interactions, divided by source of privacy threat. Data shown in this figure is from the U.S. MTurk sample (N = 182)

4.3.3 Factors of individual background affecting how people protect their information

Because the Pew sample has a wide age range and contains social media users and non-social media users, we are able to look at how demographic characteristics like age, gender and social media use affect people's hiding behaviors. Because we did not measure technical knowledge specifically in the Pew survey, people's education level is used as a proxy for their technical knowledge in the following analysis. In contrast, the MTurk sample has a relatively narrow age range (majority of respondents are below 45 years old), and includes mostly social media users (90%). We added several questions in the MTurk survey to capture people's technical knowledge, social orientation and previous negative Internet experience. In both surveys, people's worry about online information is used as a proxy to capture their general privacy concern. Because of these sample differences, I analyze the two surveys separately in the following sections.

4.3.3.1 The effect of demographic characteristics, social media use, and prior negative experience

As shown in Table 8, people who are younger, use social media, are more worried about information, and those with higher education level are more likely to report having sought anonymity online (hide identity).

Age significantly predicts hiding from all social privacy threats and authorities – younger people are more likely to hide from family & friends, employers, unwanted ties, and authorities. Gender predicts hiding from informational privacy threats – men are more likely to report hiding from authorities and other third parties than women. Using social media predicts hiding from all social privacy threats and other third parties, which is probably caused by social media users' higher awareness of tailored advertising shown on social media sites. Those who said they are worried about information online are also more likely to hide from almost all threats except for family and friends. Having had negative experience online consistently predicts hiding from all kinds of privacy threats. H2 (prior negative experience predicts identifying privacy threats) is supported for the Pew sample.

Lastly, education does not show a strong effect in hiding from different threats, except that people with lower education are less likely to hide from other third parties.

		Hide from social privacy threats			**Hide from informational privacy threats**	
	Hide identity	Family; friends; co-workers	Employers; supervisors	Unwanted ties	Authorities	Other third-parties
Age	-.10*	-.20***	-.19***	-.21***	-.15***	-.04
Gender [Male = 1]	.04	-.06†	.01	-.02	.07†	.08*
Use social media	.12**	.11**	.08*	.09**	.04	.17***

Worry about information	.11**	.03	.08*	.07*	.07†	.08*
Bad experience	.03	.26***	.10**	.28***	.07†	.15***
Education [HS or less]	-.13**	-.00	-.07†	-.02	.03	-.09*
Education [some college]	.10*	.01	.00	.06	-.03	.00
R^2	.07	.19	.17	.19	.09	.07

† p < .10, * p < .05, ** p < .01, *** p <.001. Values in the table are standardized beta estimates. All models are logistic regression models because the dependent variable is binary (hide or not hide).

Table 8. Factors predicing hiding identity and information from people or organizations. Data shown in this table is from the Pew survey (N = 775).

Then I used the same group of independent variables to predict the strategies they use to hide information (Table 8). People who are younger, use social media, and have bad experience are significantly more likely to use all types of strategies. Gender only influences how likely people manage their browser cookies: men are more likely to manage cookies than women. People who are more worried about their information online are more likely to use technical methods, but not the other three strategies. If we consider education as a proxy for people's knowledge, those who have lower knowledge (high school or less) are less likely to manage cookies and use technical methods to hide their information. Education has no effect on the use of behavioral methods (editing content or using alias).

	Manage cookies	**Use alias**	**Edit content**	**Use technical method**
Age	-.10**	-.16***	-.31***	-.11**
Gender [Male = 1]	.08*	.03	.01	.07
Use social media	.21***	.12**	.19***	.12**
Worry about information	.04	.05	.05	.09*
Bad experience	.13***	.18***	.21***	.13***
Education [HS or less]	-.25***	-.02	-.06	-.15***
Education [some college]	.07	-.02	-.01	.02
R^2	.16	.12	.27	.11

Table 9. Factors predicting strategies they use to hide. Data is from the Pew survey (N = 775)

Then I looked at three other factors by adding some measures in the MTurk survey. Because almost every respondent in MTurk survey uses social media, I did not include social media use in the following models.

4.3.3.2 The effect of social orientation, prior negative experience, and technical knowledge

First, the effect of demographic information like age and gender is weaker in this sample. Younger age only shows a marginal effect in hiding from authorities, not in

other hiding behaviors. Men are more likely to hide from authorities than women in this sample, but not in hiding from other privacy threats. Because we measured people's technical knowledge separately in this survey and the correlation coefficient between their technical knowledge scores and self-reported education level is low (r = .07), I put both variables in the model.

The three social orientation scales seem to mainly predict hiding from family and friends, and hiding identity. The model shows that respondents whose social orientation is low in collective identity and high in segmented identity were more likely to hide their identity and hide their online interactions from family, friends or co-workers. High segmented identity also predicts hiding from employers and supervisors, and hiding from more groups. We find no effect of individual identity, but this could be partly due to the low reliability of this scale (α = 0.48) compared to the other two orientation scales. H1a (low collective identity predicts identifying social privacy threats) and H1b (high segmented identity predicts identifying social privacy threats) are partly supported for the MTurk sample.

Prior negative experience predicts hiding from social privacy threats in this sample, but the effect is not significant for predicting hiding from informational threats. Therefore H2 is only supported for hiding from social privacy threats in the MTurk sample. Technical knowledge strongly predicts hiding from informational threats, hiding from employers and marginally predicts hiding one's identity. H3 (higher technical knowledge predicts identifying privacy threats) is supported for hiding from informational threats, but not social threats. In addition, prior negative experience and technical knowledge strongly predict the number of groups they hide from – suggesting that these people have identified more levels of privacy threats. Prior negative experience and technical knowledge have almost no correlation (r = -.03).

			Hide from known groups			**Hide from organizations**	
	Hide identity	**# of groups they hide from**	Family; friends; co-workers	Employers; supervisors	Unwanted ties	Authorities	Other third-parties
Age	.03	-.06	-.03	.02	.02	-.15†	-.07
Gender [Male = 1]	.04	.05	.10	-.01	.06	.13†	.01
Education [HS or less]	-.12	-.15	-.07	-.06	-.02	-.08	-.22*
Education [some college]	.06	.06	-.05	-.06	.00	.10	.22*
Worry about information	.01	.15*	.16*	.11	.07	.01	.06
Social orientation							
Collective identity	-.15†	-.06	-.18*	.00	-.03	.04	.02
Individual identity	.07	.00	.03	-.13	.03	.06	.01
Segmented identity	.16†	.14†	.19*	.18*	.13	.03	-.13
Bad experience	.08	.28***	.17*	.19*	.33***	.10	.05

Technical knowledge	.16†	.25***	.12	.20*	-.06	.25**	.25**
R^2	.12	.25	.19	.14	.16	.18	.11

† p < .10, * p < .05, ** p < .01, *** p <.001. Values in the table are standardized beta estimates.

Table 10. Factors predicting hiding identity and interactions from people or organizations. Data is from the U.S. MTurk sample (N = 182)

In addition, we examined the effect of social orientations, prior negative experience, and technical knowledge on the strategies people use to mitigate different threats (Table 11). High segmented identity and low collective identity orientation predicts more use of editing content (a behavioral strategy). High segmented identity is also associated with using technical methods.

Having bad experience is associated with more use of editing content and technical methods. Technical knowledge marginally predicts the use of managing cookies, and strongly predicts the use of technical methods, but does not predict the use of the other two methods. This finding echoes previous research (Joinson et al, 2010) that both technically sophisticated and naive users use behavioral methods to protect their privacy online (using alias and editing content in this study).

	Manage cookies	**Use alias**	**Edit content**	**Use technical method**
Age	.06	.03	-.00	.04
Gender [Male = 1]	-.06	-.04	-.03	.02
Education [HS or less]	-.13	-.09	-.14	-.08
Education [some college]	.09	-.03	.08	.06
Worry about information	.18*	.19*	.14†	.06
Social orientation				
Collective identity	-.03	-.11	-.17*	-.00
Individual identity	.06	.12	-.05	-.10
Segmented identity	.02	.11	.22**	.21**
Bad experience	.05	.03	.25**	.12†
Technical knowledge	.15†	.10	-.07	.48***
R^2	.08	.12	.16	.30

Table 11. Factors predicting strategies they use. Data is from the U.S. MTurk sample (N =182)

4.3.3.3 The effect of technical knowledge and negative experience on perception of anonymity

We found that those with more technical knowledge reported hiding from more threats, and using more technical methods to protect their information. Did they feel more secure than those without technical knowledge, or did they feel less secure than those without technical knowledge, who might benefit emotionally from their ignorance? Having more technical knowledge could make people feel more empowered because

they know how to use tools to protect themselves, or they could feel even more helpless because they are more aware of the possible threats and difficulty of mitigating them than those with less knowledge.

We asked two preference questions related to people's opinion about anonymity in the MTurk survey and examined the effect of bad experience and technical knowledge on answers to these questions. Age, gender, and education were added to the model as control variables. Those who had had a bad experience were significantly less likely to think that it is possible to be completely anonymous (see Table 12), but they were also less likely to agree that people should have the ability to be completely anonymous. Those with more technical knowledge were more likely to agree that people should have the ability to be anonymous. We noticed a significant interaction effect between technical knowledge and have had a bad Internet experience on whether or not respondents thought it is possible to be anonymous online. We divided participants into high technical users and nontechnical users by doing a median split on the technical knowledge measure (a continuous variable). As shown in Figure 5, Among those with higher technical knowledge, those having had no prior bad experience seemed to be more confident and to think they could be anonymous than those who had had a bad experience (45% said yes vs. 16% said yes, t [93] = 3.18, p < .01). Among nontechnical respondents, a bad experience had no significant impact on their perceptions, and only a minority thought it was possible to be anonymous online (37% vs. 24%, t [62] = 1.14, p = .26).

	Think that it is possible to be completely anonymous (31% said yes)	**Think that people should have the ability to be anonymous (86% said yes)**
Age	-.09	-.05
Gender [Male = 1]	.07	.03
Education [HS or less]	-.00	-.08
Education [some college]	-.16	.11
Technical knowledge	-.01	**.19***
Bad experience	**-.23****	**-.17***
Technical knowledge × bad experience	**.15***	-.04
R^2	.12	.09

† p < .10, * p < .05, ** p < .01, *** p <.001. Values in the table are standardized beta estimates.

Table 12. Logistic regression examining factors that predict policy preferences. Data shown in this table are from the U.S. MTurk survey (N = 182). Those who answered "not sure" were treated as missing values.

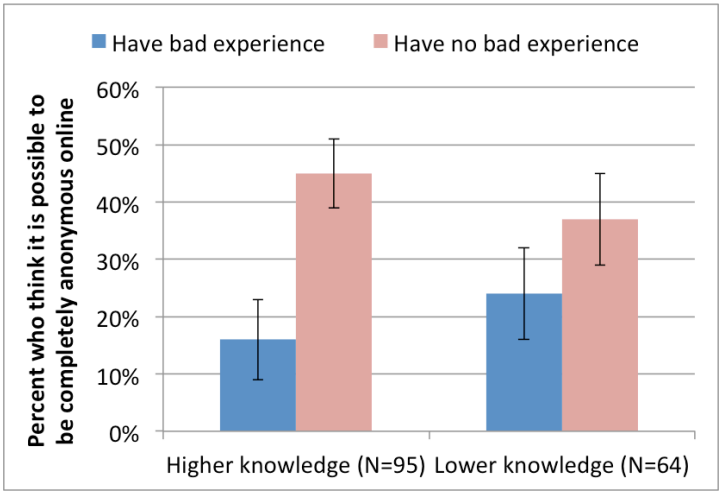


Figure 5. The Interaction effect of knowledge and bad experience on perception of whether or not anonymity is possible.

4.4 Discussion

Overall, these findings provide direct empirical evidence on the impact of individual background and experience on how people identify and act upon different sources of privacy threats online. The majority of the Internet users we surveyed had tried to hide some of their information online. A small proportion of them had explicitly tried to hide their identity. The surveys are correlational but showed how the three factors (identified in previous chapters) of social orientation, prior negative experience, and technical knowledge, predict how people identify and try to mitigate different threats to their privacy. In addition to the privacy threats from government, hackers, and companies that are widely acknowledged, a significant number of respondents in both surveys took action to mitigate privacy threats from people within their own social networks.

4.4.1 Desire to manage social boundaries

The measures of social orientation (especially the scale of segmented identity) were intended to capture people's desire to manage their social boundaries. Managing social boundaries was identified as an important motivator for anonymity seeking online in both studies described in Part I of this thesis. The MTurk survey described in this chapter replicated this finding. In general, the measure of segmented identity (adapted from the self-monitoring and faceted life scales published in previous work) significantly predicted respondents' hiding from social privacy threats and editing their online content. The more segmented people's identities were (i.e., they wanted to present different images to different audiences), the more likely they were to hide from social privacy threats, and to edit the content they had posted. Our findings suggest that segmented identity might be an important personality measure to consider in future privacy research.

4.4.2 Should we supplement people's technical knowledge and add to their threat perception to motivate privacy protective behavior?

Both surveys showed that more technical knowledge (measured by education in the Pew survey and technical knowledge tests in MTurk survey) was associated with respondents' hiding their identity, hiding from informational threats, and using technical methods to hide their information. Technical background might also moderate the effect of other individual background factors. We found that prior negative experience online elevated people's concerns and made them feel less anonymous online, and this effect was more pronounced for technical respondents (Figure 5). Those with higher technical knowledge who did not have any prior bad experiences on the Internet were more confident, and thought it is more possible to achieve anonymity than technical people who had had a prior bad experience. They might have assumed their knowledge protected them and that the steps they took were effective in achieving privacy.

To motivate more privacy protective behavior on the part of users, we might consider educating Internet users, giving them more knowledge about how the Internet works and different ways their privacy can be threatened. Prior negative experience consistently predicted more privacy protective behavior in both survey samples. Therefore we also might exploit the power of negative experiences. Previous work shows that simulating a bad consequence might be more effective than privacy warnings alone. Kumaragaru et al. (2007) made users fall for phishing email and then educated them about phishing attacks. Educating users about other people's experiences may also be helpful. Knowing who among a user's friends fell prey to a recent password breach might motivate people to adopt a stronger password (Das, Kim, Dabbish, & Hong, 2014). Greater transparency of the links among a user's own Internet practices and privacy threat might also change behavior. One study showed that revealing password weaknesses in a highly visual way motivated people to create stronger passwords (Kim et al., 2014).

4.4.3 Limitations

Our findings might be more illuminating if we added other individual difference measures (Smith et al., 2011), such as extraversion and self-esteem (Pedersen, 1982), or a measure of actual-ideal self (Triandis, 1989). For instance, some of the interviewees described in Chapter 2 said that being anonymous can help people express an ideal self-image that they do not express in real life. The desire to present an "ideal" or different selves differs among people and, when strong, could motivate them to manage their online information differently. For example, Ellison et al. (2006) reported people constructing their online dating profiles to reflect an ideal self they desire, which include concealing some parts of their actual identity (i.e., representing less weight).

We asked about five kinds of concrete threats in the surveys but did not ask respondents about threats from unknown others. As described in Chapter 2, some people feel their privacy is threatened, but they do not know where this threat originates--who the attackers are or who might harm them in the future. Asking people about unknown threats might suggest a somewhat different predictive model. I did not use established scales from previous research to measure people's general privacy concern (such as IUIPC scale by Malhotra et al, 2004) besides controlling for people's worry about information. In a later study (described in Chapter 6), I used the widely adopted IUIPC scale to measure people's general privacy concern.

5 "My data just goes everywhere:" User mental models of the internet[4]

5.1 Introduction

Today, the Internet is a ubiquitous vehicle for information, communication, and data transportation, but it is not an automated device that works in a simple and secure way. Prior literature and the findings described in previous chapters reveal that many people who use the Internet everyday know little about how it really works. People have to make many decisions that affect their privacy and security, ranging from whether to access public Wi-Fi at an airport to how to share a file with a colleague, to how to make up and remember a new password for a shopping site. It is therefore important to understand what people have to know to protect themselves, and whether or not their technical knowledge of the Internet influences their daily privacy and security practices.

From the surveys in chapter 4, we learned that people's technical knowledge might be associated with the use of technical strategies to protect their privacy from informational threats (authorities or third-parties). However, we did not capture their overall understanding of the Internet and still do not know which specific part of the respondents' technical knowledge influenced their privacy and security practices.

Understanding how users think about the Internet could help us design privacy and security interfaces that match user perceptions. Understanding how users think the Internet works also will help us develop educational programs so that users, as citizens, can be better informed about privacy policies and other aspects of Internet governance. A clearer picture of how users think about the Internet also could help system designers develop technologies and policies that meet users' expectations and help policy makers communicate in ways that are easily understood by lay people (Sen, Joe-Wong, Ha, &

[4] This chapter is adapted from: Kang, R., Dabbish, L., Fruchter, N., & Kiesler, S. (2015). " My Data Just Goes Everywhere :" User Mental Models of the Internet and Implications for Privacy and Security. In Proceedings of SOUPS 2015 (pp. 39–52). USENIX Association.

Chiang, 2013). Towards these goals, I and my co-authors examined users' mental models of how the Internet works.

Part of the challenge in understanding the Internet is its rapid evolution. The Internet is now massive and embedded into many contexts. It connects billions of individuals around the world through many different types of devices (Zheng, Simplot-Ryl, Bisdikian, & Mouftah, 2011). Many entities are involved in transmitting data and tracking user behavior including third party caching services, first and second level ISPs, cellular network providers, web services, search engines, and ad networks. More personal data than ever is transmitted via the Internet as mobile access proliferates (Brown, Mortier, & Rodden, 2013) and service providers expand their tracking, creating privacy and security challenges far beyond the ability of end users to manage (Tbahriti, Ghedira, Medjahed, & Mrissa, 2014). Network security tools are not widely used and do not help users understand why or how well they work.

We conducted a qualitative study in which we asked users to describe and explain how the Internet works, both in general and while they did different common, Internet-based tasks. We sampled users with and without computer science or related technical or computational backgrounds. We identified patterns in their conceptual models of the network and awareness of network-related security and privacy issues. A mental models approach, in contrast to surveys or other methods, revealed subtle differences in people's knowledge of the Internet. Our results suggest that user perceptions do vary as a function of their personal experiences and technical education level. Users' technical knowledge partly influences their perception of how their data flows on the Internet. However their technical knowledge does not seem to directly correlate with behaving more securely online.

5.1.1 Users' mental models

A commonly used method in psychology to elicit users' understanding about a problem is mental models, which are "psychological representations of real, hypothetical, or imaginary situations (Jonassen & Cho, 2008)." Mental models describe how a user thinks about a problem or system; it is the model in the person's mind of how things work. These models are used to make decisions by supporting mental simulation of the likely effect of an action. Mental models of a system can be useful in informing interface design or educational materials because they suggest natural ways to visualize complex system components or user interactions with them. A number of researchers have adopted the mental models approach to understand users' perceptions of the Internet (Klasnja et al. 2009; Poole, et al. 2008) and Internet-related systems or technologies, such as home computer security (Wash, 2010), firewalls (Raja, Hawkey, & Beznosov, 2009), and web security (Friedman, et al., 2002).

Diagramming exercises are considered a good way of capturing mental models in addition to traditional verbal reports (Jonassen & Cho, 2008), and this method is

frequently used in user-centered Internet research. Poole et al. (2008) used a sketching task in order to understand lay persons' knowledge of home networks. Their results suggest that most users, even those who are technically sophisticated, have a poor understanding of home networking structures. Klasnja et al. (2009) also used a diagraming task when studying how users understand and use Wi-Fi. Their study revealed that users had an incomplete understanding of important privacy risks when they were connected to Wi-Fi, such as malicious access points, and did not protect themselves against threats, such as seeking SSL encryption. Four out of the eleven participants they observed were aware that other people could possibly access their information being transmitted over Wi-Fi, but this understanding did not raise concerns.

Having a deficient mental model may indicate a lack of awareness of the security risks surrounding Internet activities. Some prior work specifically examined users' perceptions of security systems. Wash (2010) interviewed people about how they understood security threats to their home computer and summarized different folk models about home computer security including models centered on viruses and models centered on hackers. Friedman et al. (2002) also addressed security risks, interviewing 72 participants and asking them to do a drawing task to illustrate their understanding of web security. They found that the majority of participants relied on simple visual cues like the presence of HTTPS and a lock icon to identify secure connections. Raja et al. (2009) studied users' mental models of personal firewalls on Windows Vista using a structured diagramming task. They gave participants images of a computer, firewall, and the Internet depicted as a cloud, and asked participants to connect those pictures with arrows. They then improved understanding of firewalls by showing participants an interface prototype with contextual information.

Many studies show that more technically advanced users have a different understanding of the Internet and computer systems compared to more novice users. Bravo-Lillo and colleagues (2011) compared advanced and novice users' differences in their mental models about computer security warnings, finding that advanced users had much more complex models than novice users. Vaniea et al. (2014) interviewed people about their experiences with a specific application, Windows Update. They found that a lack of understanding might prevent people from installing important security updates for their computers, thus increasing security risks. Their study suggests that a reasonable level of technical knowledge is essential to guide correct user decisions.

Besides privacy-specific research, we can also draw from literatures about people's general understandings of complex systems. Researchers in cognitive psychology argue that complex systems often include multiple levels of organization and complex relationships. Hmelo-Silver and Pfeffer (2004) compared experts' and novices' conceptualization of a complex system and found that novices' understanding focuses more on "perceptually available" (concrete) components, whereas experts mention more "functional and behavioral" (conceptually abstract) components. A few other studies

(Jacobson, 2001; Resnick & Wilensky, 1998) found that people often assume centralized control and single causality, especially domain novices, whereas experts think about decentralized control and multiple causes when asked to describe a complex system.

The previous work on Internet mental models provides some insight into the nature of users' understanding of the Internet and its anchoring in personal experience. Much of this work, however, is task-specific or focuses on a specific security tool or application. A number of other researchers have conducted interviews or surveys to study users' general or privacy-related Internet knowledge.

5.1.2 Users' knowledge of the Internet

Various attempts have been made to measure users' knowledge of the Internet. Page & Uncles (2004) categorized Internet knowledge into two categories: the knowledge of facts, terms or attributes about the Internet (declarative knowledge), and the knowledge of how to take actions or complete tasks on the Internet (procedural knowledge). Following this argument, Potosky (2007) developed an Internet knowledge measure (iKnow) that asks people to rate their agreement as to whether or not they understand terms related to the Internet (e.g., "I know what a browser is"), and whether or not they are able to perform Internet-related tasks (such as "I know how to create a website"). An important question researchers have asked is what impact these two kinds of knowledge have on user security and privacy protection behavior.

Park (2011) measured user knowledge in three dimensions: technical familiarity, awareness of institutional practices, and policy understandings. He found higher user knowledge correlated with online privacy control behavior. Other studies emphasize the role of user skills. Das et al. (2014) proposed three factors influence the adoption of security and privacy tools: awareness of security threats and tools, motivation to use security tools, and the knowledge of how to use security tools. Litt (2013) found that higher Internet skills were positively associated with more content generation online and managing one's online presence. boyd and Hargarttai (2010) found that users with more Internet skills were more likely to modify their privacy settings on Facebook. Hargittai and Litt (2013) developed a scale to specifically measure privacy-related skills. They asked people to evaluate their level of understanding of privacy-related Internet terms such as "privacy settings," "tagging," and email "bcc." Their survey showed that higher privacy-related knowledge was positively associated with better privacy management of social media profiles.

Having more declarative knowledge or skill has not always been shown to predict more secure online behaviors. Dommeyer and Gross (2003) found that consumers are aware of privacy protection strategies, but do not use them. In a study by Nguyen and colleagues (2008), some participants expressed uncertainty about how store loyalty cards would be used, but they did not take any protective actions to protect their personal information. Furnell et al. (2007) studied how people manage security threats to home PC systems

and found advanced technical users did not use more effective security practices than novice users.

The Internet today is much different than what it was 10 years ago, so people may perceive or use it very differently today, especially in managing their privacy. In 2003, the majority of American Internet users expressed strong concern about information used by governments and corporations, but they had little knowledge of how their data flows among companies (Turow, 2003). A more recent 2011 review of the literature suggests that people's awareness of organizations collecting their personal information increases their privacy concerns (Smith et al., 2011), but there remains little understanding of how people think the Internet works. In late 2014, Pew Research Center conducted a national U.S. sample survey to test Internet users' knowledge of the Web by asking 17 questions about Internet terms (e.g., "URL"), famous technology celebrities (e.g., identifying Bill Gates' photo), and the underlying structure of the Internet (e.g., explanation of Moore's law) (Pew Research Center, 2014). Their survey indicated that the majority of Internet users recognize everyday Internet usage terms, but very few are familiar with the technical details of the Internet and most do not understand Internet-related policies.

In sum, there is mixed and indirect evidence of whether or not an accurate mental model and more advanced Internet knowledge are associated with more secure online behavior. In light of the new data privacy and security challenges associated with the Internet's evolution, we wanted to assess how people currently understand the Internet, their perceptions of how their data flows on the Internet, and what they are currently doing to protect their privacy or data security. Our work aims to examine the relationship between people's knowledge and their privacy and security behavior in today's Internet environment, and to move towards a better understanding of the kinds of Internet knowledge users need to have.

5.2 Method

We conducted semi-structured interviews with twenty-eight participants about their mental models of the Internet. A list of all the participants is shown in Table 1. In addition, after completing the interviews with technical and nontechnical participants, we invited 5 domain experts (faculty members in computer networking or computer security domain at a research university) to review and evaluate several mental model drawings generated by technical and nontechnical participants. Here, we first introduce the method and results of the interviews with participants. Then, we discuss the implications of our results and incorporate experts' comments into the discussion and implication section.

5.2.1 Participants

We did three rounds of data collection and recruited a total of 28 participants. Each participant was paid $10 for a 30-45 minute interview session.

The first two rounds of participants were recruited through flyers, personal contacts, and an online participant pool at a US east coast research university. At the outset of this study, we used educational level and college major as a proxy for technical knowledge (used for N01-N09, T01-T03). For other technical participants recruited in the second round (T04-T10), we developed a screening survey for technical knowledge, only accepting participants who scored 5 or higher in an 8-item survey as technical participants (see Appendix II). Those who scored lower than 5 counted as non-technical participants (N10, N11). These nontechnical and technical participants included people from the local area, university staff members, and students pursuing all levels of degree study. Non-technical participants had a mix of backgrounds. Technical participants all had computer-related college majors.

Identifier	Gender	Age	Education background
Lay participants (N = 17)			
N01	M	19	Finance
N02	M	22	Finance
N03	M	22	Biomedical Engineering
N04	F	18	Geology
N05	F	22	English
N06	M	22	Law
N07	F	21	Cognitive science
N08	F	19	Statistics; psychology
N09	F	22	Legal studies
N10	M	30	Music; foreign languages
N11	F	18	Neuroscience
C01	M	64	Engineering; public health
C02	M	32	Culinary arts
C03	M	62	Communication arts; religion
C04	M	49	Psychology
C05	F	58	MBA
C06	F	30	Foreign policy
Technical participants (N = 11)			
T01	F	19	Computer science
T02	F	21	Computer science
T03	F	27	Computer science & HCI
T04	M	25	Information technology
T05	F	24	Electrical/CS engineering
T06	M	26	Computer science
T07	M	25	Information technology
T08	M	23	Computer science

T09	M	27	Software engineering
T10	M	24	Software engineering
T11*	M	32	Computer science

Table 13. Study 5 participants (Total = 28; N = non-technical participants; C = community participants; T = technical participants; *T11 was recruited with the community sample).

Because our initial two samples were similar in age and university education, we also recruited a third group of participants from the local community by posting an advertisement on craigslist with the inclusion criteria of age 30 or older (C01-C06). One of these participants (T11) had a computer science background, so was treated as part of the technical sample. Both the nontechnical and community participants had non-computer science related education backgrounds, so we refer to them together as "lay participants" in the following sections. Participants who had had formal computer science or computing education are referred to as "technical participants."

5.2.2 Procedure

In the interview study, participants were brought into a room equipped with pen, paper, and a desktop computer. After an overview of the study, participants completed a short survey regarding Internet experience, smartphone literacy and computer knowledge. They were also asked about the number and types of devices they owned.

After completing the survey, participants were guided through the main drawing tasks. Every participant was first prompted to explain how the Internet works, and asked to draw a general diagram of it in whatever form they chose on a large sheet of paper in front of them. Participants were instructed to verbalize their thought process as they drew, consistent with traditional think aloud protocols (Ericsson & Simon, 1980). A video camera captured participants' drawings and voices. All recordings were labeled using anonymous identifiers. No personally identifiable information was collected or recorded.

Each participant was then asked to draw several diagrams about specific tasks they did on the Internet following the same procedure. The tasks used were a subset of the following: *watching a YouTube video, sending an email, making a payment online, receiving an online advertisement* and *browsing a webpage.* After each model drawing was completed, participants were asked several follow-up questions, clarifying drawings and explanations as needed. Additionally, participants were asked to draw a separate diagram for each task if they thought it worked differently on mobile devices. The interview script is attached in Appendix III.

After the drawing tasks, participants filled out a post-task survey with demographic questions, as well as a series of Internet knowledge questions. The Internet knowledge questions were the same as the knowledge questions used in the MTurk survey described in Chapter 4 (attached in Appendix I). All 28 participants filled out the same

post-task survey. Besides differences in academic background, technical participants performed significantly better than lay participants in both the self-rated familiarity questions (mean: technical = 3.59, lay = 2.47, t [26] = 4.32, $p < .001$) and correctness on the true/false questions (mean number correct: technical = 4.27, lay = 1.53, t [26] = 5.83, $p < .001$).

5.2.3 Data Analysis

We qualitatively analyzed participants' think aloud responses to identify key differences across mental models. We conducted our analysis iteratively, carrying out three rounds of data collection and subsequent analysis, allowing the first analysis process to guide our second round of data collection, and then the third. Our initial analysis occurred after the first 12 sessions with participants (predominantly non-technical participants). We focused on the diagrams they generated during our sessions as well as the video and audio recorded during our sessions. By comparing and contrasting across user models, we generated a set of codes that indicated dimensions on which the models varied. To verify and extend codes and themes identified in our first round of data analysis, we conducted a second round of analysis, extending codes identified in our first round based on new features of the second set of models. In the last round of data collection, we added a few questions to the interview based on results from the previous two rounds. The third round of data collection expanded the age range of our sample and let us examine the influence of users' past experience and concerns on their perception and behavior. Six interview recordings were lost due to equipment problems but field notes on paper were available. The remaining 22 of the 28 interviews were recorded and transcribed (9 technical, 7 nontechnical, and 6 community participants). Aside from analyzing the drawings, we performed qualitative data analysis of the verbal transcripts and field notes using a grounded theory approach (Corbin & Strauss, 2008). The data were coded in Dedoose (http://www.dedoose.com/). A second researcher independently coded 15% of all the interviews. Our analysis showed a good inter-coder agreement between the two researchers (kappa = 0.79).

5.3 Results

Our analysis showed that participants with different technical education and personal experiences had very different mental models of how the Internet works. These models were related to participants' perceptions of privacy threat and what happens to their data on the Internet. However, technical education and mental models did not seem to be very predictive of how participants acted to protect their privacy or security. Those actions appeared to be more informed by participants' personal experience. In the following sections, we first discuss users' knowledge of how the Internet works as a system and their awareness of security and privacy features in the system. Next, we present people's different perceptions of their personal data on the Internet. Lastly, we

show the methods participants take to prevent their data from being seen and discuss the connections between their knowledge, perception and the protective actions.

5.3.1 Users' knowledge of the Internet

Participant models varied in their representation of the Internet as a simple system or service (the "Internet" in Figure 6) or as an articulated, technically complex system (Figure 7 and Figure 8).

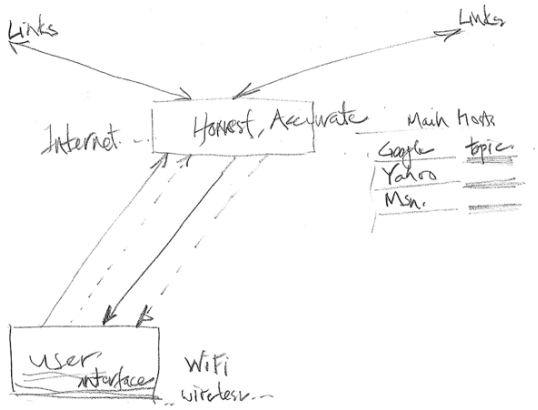


Figure 6. Internet as service (C01)

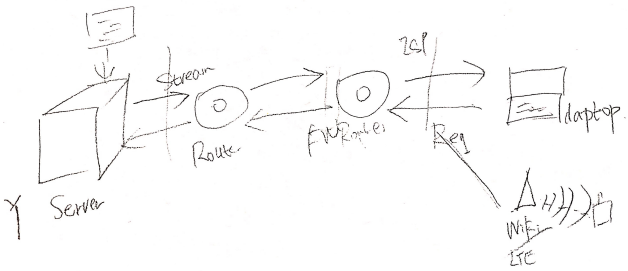


Figure 7. Articulated model with hardware components (T10)

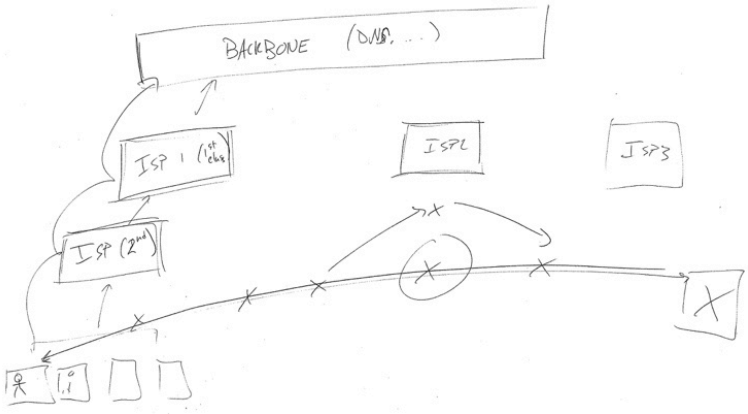


Figure 8. Articulated model with multiple layers of the network (T06)

5.3.1.1 Simple vs. articulated system mental models

A majority of the lay participants represented the Internet as a comparatively simple system or service consisting of the user connected to a "server," data bank, or storage facility. These participants used metaphors such as earth, cloud, main hub, or library that receives and sends out data. Thirteen lay participants and one technical participant belonged in this category. Their models showed that the Internet receives and sends out data, indexes webpages, and responds to their different requests. A few users considered Google or Yahoo the main provider that connected them to other webpages.

> *So everything that I do on the Internet or that other people do on the Internet is basically asking the Internet for information, and the Internet is sending us to various places where the information is and then bringing it back. (C01, Figure 6)*

Most lay participants only expressed surface-level awareness of organizations and services that they interacted with directly such as Google and Facebook, but did not mention any of the underlying infrastructure. When talking about making online payments, for example, they mentioned a number of different organizations involved in the process such as "the bank," "Amazon," and "PayPal." Some were aware of physical objects that helped them connect to the Internet (see N05's drawing of a router in Figure 9). Three lay participants also drew mobile towers when describing a cellular network. Three thought satellites played a role in connecting them to the Internet, but none of the technical participants mentioned this.

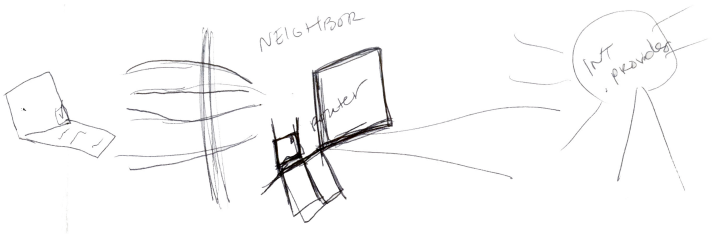


Figure 9. Drawing of how she uses neighbor's Wi-Fi (N05)

In most technical participants' drawings, we seldom saw a simple system or service representation of the Internet. Instead, users had more articulated models of the Internet as a complex system with varied hardware components and a more involved set of connections among components (Figure 7 and Figure 8). Ten technical and four lay participants belonged in this category. The number and presence of entities and organizations within participants' sketches mirrored to some extent their Internet literacy levels. The presence of other computers, servers, ISPs, DNS, routers, servers/clients, and infrastructure hardware spoke to a participant's knowledge and understanding of the Internet as a complex system.

Some technical participants articulated their view of multiple layers of the network (Figure 8), whereas most lay participants described one layer of the network. A few technical participants mentioned physical layers ("fiber cable", T05), or concepts potentially associated with a physical layer such as physical location (such as a "U.S. server," or a university as a physical entity). Most technical participants (9 out of 11) expressed broader awareness of entities and organizations involved in the Internet. For example, 6 technical participants noted there were many different ISPs. Furthermore, technically advanced users had specialized knowledge. Five technical participants mentioned network protocols such as "TCP/IP", "SMTP", or "IMAP", but none of the lay participants mentioned these concepts. Some technical participants also mentioned

logical elements such as "routing" or "peering." The differences between these two types of mental models are explained in Table 14.

	Description of the models
Simple and service-oriented models: 13 lay participants; 1 technical participant	Represent the Internet as a vague concept or a service; Only show awareness of organizations or services they directly interact with; Lack awareness of underlying layers, structures and connections; Use inconsistent or made-up terminologies.
Articulated technical models: 4 lay participants; 10 technical participants	Represent the Internet as a complex, multi-level system; Show broader awareness of components and organizations in the network; Express awareness of layers, structures and connections; Use accurate, detailed, consistent terms.

Table 14. Differences between simple and articulated models

There were aspects of the mental models both groups had in common. Regardless of their technical background, participants said that the Internet connects computers and supports communications. For instance, a 49 year-old local flower shop owner was quite excited about all the changes the Internet has brought to his life, and mentioned that the Internet enables him to *"talk to friends that I've lost contact over the years."* (C04) A technical participant focused more on the infrastructure: "*There's a level at which there're ISPs that communicate with each other*." (T06)

5.3.1.2 Awareness of security and privacy[5]

We analyzed the comments related to security and privacy that naturally emerged during the interview as a measure of people's general awareness and attention to security and privacy. We did not explicitly prompt people to talk about security mechanisms of the Internet. The concepts that emerged concerned private vs. public spaces, protection mechanisms, trust, and perception of security on mobile phones vs. computers.

5.3.1.2.1 Public vs. private communication

Six lay participants and two technical participants talked about distinctions between public vs. private information or connections. For instance, one nontechnical participant thought that home Wi-Fi is more secure than public Wi-Fi because it has firewall and security settings (N09). Several participants thought sending an email or doing an online payment is private while watching YouTube videos is public. A few participants

[5] This section and following sections are based on the 22 interview transcripts, including 9 technical and 13 lay participants.

mentioned privacy settings on YouTube or Facebook that they could use to control whether their information was public or private.

> *I think there's a user profile [on YouTube]. I mean that to me is a much more public space. (C06)*

5.3.1.2.2 Protection mechanisms

We coded users' expressed awareness of protection mechanisms such as encryption, passwords, certification of websites, and verification steps implemented by websites. One lay participant and seven technical participants said that their email, online payments, or connections could be encrypted. T04 said, *"If I'm going to use Gmail then I assume that, by default, the connection is going to be encrypted between my PC and the Gmail server."* Another technical participant drew a little lock sign in his model to indicate that the connections are encrypted (Figure 10).

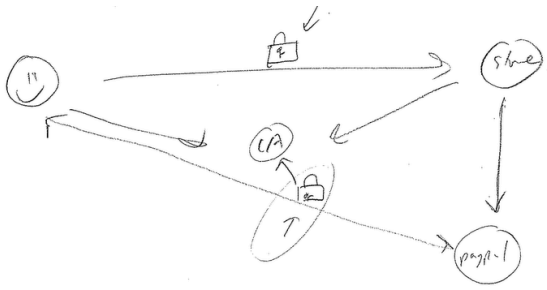


Figure 10. Model of making an online payment to a shoe store (T09)

One lay participant (N06) said, *"I don't put [my credit card info] in when there's not like that little lock up on top of the screen. I think it's pretty secure."* Also, when talking about sending an email or making an online payment, some participants mentioned the bank or email server would verify the requester's identity (T04, T08, and N11). In Figure 5, the technical participant included a certificate authority ("CA") in his model of online payments.

5.3.1.2.3 Trust

Eight lay participants and three technical participants expressed shared beliefs about the security provided by big companies or institutions, and considerable trust in those they knew. The cues participants used to decide whether or not they would trust a website included their knowledge that other people had used the same service, that it was a reputable brand, terms of service, certificates, warnings, and whether or not they had had a bad experience on the site.

> *I think if this was Amazon, their site is probably protected. (C05)*

One participant transferred his trust of the physical bank to the online world.

> *I talk to the employees there in person a lot, and they just seem to have a level head on their shoulders. I don't think they would give out their information to anybody over the phone without verifying who they were with some kind of credential verification. (T11)*

5.3.1.2.4 Mobile phones vs. computers

Participants offered mixed opinions about whether it is more secure to connect through the phone or through their computer. N10 said it is less secure to do banking or payment related activities on a mobile phone, because he felt it was like "*sharing wireless connections with other people in a public network*." He thought the difference between connecting from his computer vs. connecting from his smartphone was that the connection on mobile phone was wireless.

By contrast, T10 always used his smartphone to make payments because he was worried that his computer might have a virus or tracking software and thought his phone would be more secure. C01 thought a mobile hotspot was more secure than connecting to a public Wi-Fi at a coffee shop because he was the only one on it.

5.3.2 Users' perceptions of their data

A great deal of privacy-related policies and research efforts concerns organizational practices in the collection, retention, disclosure, and use of personal information. In our study, we asked users about their perceptions of how personal data is dealt with on the Internet.

5.3.2.1 Where does my data go?

Most participants were aware that their data is sent to the servers of the company who provides them services such as Google. Two lay participants had a very vague idea of where their data went (C03 and C04). When asked about where his data goes on the Internet, the flower shop owner said:

> *I think it goes everywhere. Information just goes, we'll say like the earth. I think everybody has access. (C04, Figure 6)*

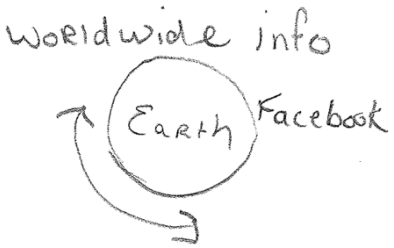


Figure 11. A depiction of where his information goes online (C04)

Regarding where their data is stored, participants mentioned "Google's large storage banks," cloud storage, ISPs, and advertising companies. One participant said, *"Once*

something is online, it's there forever." (T11) A few others were not sure if information would be stored permanently, using the evidence of having seen webpages removed.

Many participants were familiar with the partnerships among different organizations, an idea they mostly learned from news articles or personalized advertisements and services. N11 mentioned the *"paid relationship between Google and Amazon."* C02 said, "*Government can piggyback off the different servers and get all the information of what they are looking for.*" Eight lay participants and eight technical participants talked about personalized advertisement and personalized service such as tailored search results and video suggestions. Recommendations or ads tailored to their interests made people aware of a data partnership among different companies, but most of them could not spell out to whom their data was sold.

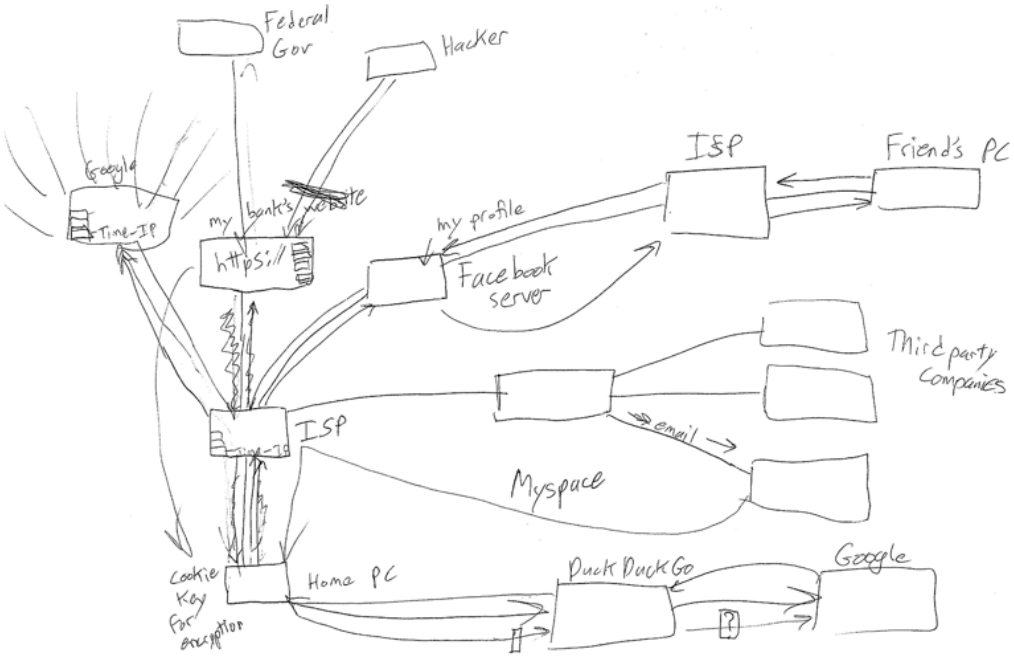


Figure 12. Model of the Internet including who can access his information (T11)

5.3.2.2 Who can see my data?

After each participant completed their drawing of the Internet, the interviewer asked, "Are there any other people, organizations, or companies that can see your connections and activities?" Privacy threats participants identified in frequency order include: companies that host the website (e.g., YouTube, Amazon) (mentioned by 18 out of 22 participants), third parties (e.g., advertisers or trackers) (mentioned by 14 participants), the government (mentioned by 12 participants), hackers or 'man in the middle' (mentioned by 12 participants), other people (e.g., other users online, other people using the same Wi-Fi) (mentioned by 11 participants), internet service providers (mentioned by 8 participants), employer (mentioned by 2 participants), and browser owners (mentioned by 1 participant). Figure 12 shows a fairly complete representation of all the people and organizations that the participant thought had access to his information,

including the government, hackers, company, ISP, and third parties. This participant (T11) studied computer science in school, but stated that his current job was not related to technology.

We compared how much lay and technical participants' mentioned the six most frequently mentioned threats. These two groups did not differ significantly in their general awareness of who has access their data. Lay participants mentioned on average 3.23 threats (out of 6), whereas technical participants mentioned on average 3.67 threats, a small non-significant difference overall. As shown in Figure 13, technical participants were significantly more likely, however, to mention hackers having access to their data than lay participants did. Across the categories of threat, they were more specific in identifying threat such as ISPs, whereas lay participants mentioned more vague threat such as third parties: "*whoever tries to make money off of you.*" (C02) This generality was probably due to the more simplistic mental models lay participants had about the Internet.

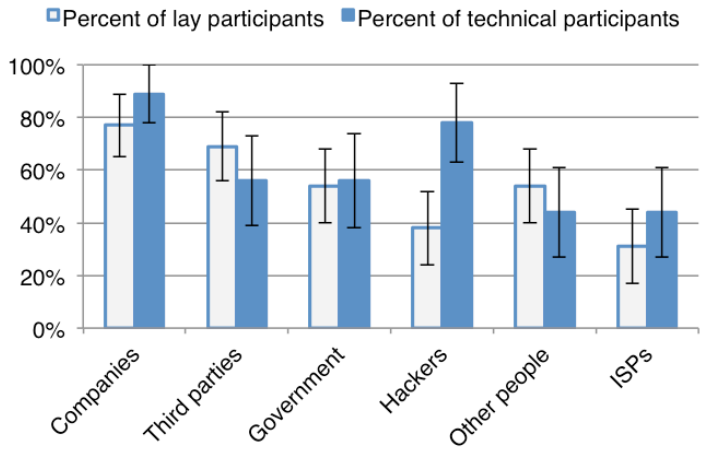


Figure 13. Percent of lay or technical participants who mentioned each group that might have access to their data.

Although technical education did not seem to influence participants' overall perception of privacy threat, the mental models (simple vs. articulated) were somewhat predictive of the number of threats people perceived. We found that, on average, participants with articulated models mentioned more sources that might have access to their data than those with simple models (mean number of threats mentioned by people with articulated models = 4 and the number mentioned by those with simple models = 2.56, t [20] = 2.80, p = .01). Those with articulated mental models expressed higher awareness of privacy threats from government, hackers, and ISPs. This higher level of awareness may be caused by these people's better understanding of where risks could occur in the network. For example, with a mental model like Figure 6, there is no way the user would know what privacy risk his ISP could bring to his data on the Internet.

Besides these specific threats, some participants thought that "everyone" could access their information, either in the general sense, or in certain situations. T06 stated that, *"the Internet is not designed to be private"* and explained the technical details of why this is the case – *"at the end of the day you're relying on correct implementations of logically sound security protocols, and historically most implementations aren't correct and most protocols aren't logically sound. So, it's just a question of an arms race of who's paying more attention."* Two lay participants also held similar opinions about their information online – *"anybody that has the capability of getting through passwords or encryptions can get it [personal information]"* (C02 and C03). N07 thought that YouTube is open to "*a lot of other people*," so the data is available to everyone. Similarly, two community participants (C04 and C05) thought the Internet in general grants everybody access.

As described earlier in the paper, participants tended to deem sending an email and making online payment as more private than activities like posting on social media. Therefore, when asked whether others could see their transaction of an online payment, T10 said *"I don't think so."* Two other technical participants (T07 and T08) thought no one could intercept their email, because they had a password or encryption to protect their email content. N06 also thought that no one could see his email, but was not able to provide any further explanation except that *"email is more private."* A technical participant (T11) mentioned that he expected Netflix not to sell his data because it's a paid service, but he was uncertain of how exactly it works: *"I try to browse through the terms and conditions but there's so much there I really don't retain it."*

5.3.2.3 Different types of information

Previous research has shown that users consider some personal information more sensitive than others (Ackerman, Cranor, & Reagle, 1999). From our interviews, we saw different user privacy expectations for different types of information, including not only personal information, but also technical identifiers. For instance, three lay participants thought companies could access their purchase history but not credit card information (N06, C02, C06). N11 suspected companies would be more interested in what she watched on YouTube than her emails, so she expected more protection on emails. Some participants were aware of the differences between identifiable information (such as names) and non-identifiable information like an ID number or IP address (C05), but she also said *"they could find it [my name] from this ID."* T09 pointed out that even for encrypted messages, his ISP could see all the packets and they could still tell *"where the origin, which is me, and what it's going."*

5.3.3 How do people protect their information?

5.3.3.1 Protective actions

We asked people: "Did you do anything to prevent any others from seeing your connections or activities?" Participants mentioned a wide range of protective actions they had tried, such as not logging in to websites, watching for HTTPS, and using cookie blockers or tracker blockers. We categorized the actions participants used into four

categories as shown in Table 15. *Proactive risk management* includes general precautionary steps people take in daily use of the Internet. *Event-based risk management* includes people's actions towards specific requests or intrusions. *Controlling digital traces* includes actions that mask or remove people's digital or physical footprints. *Securing connections* indicates methods people take to make sure their connection to a certain site or their general Internet connection is secure or anonymized.

Types of protective action	**N**	**# of lay participants who have used this type of action (out of 13)**	**# of technical participants who have used this type of action (out of 9)**	**Actions**
Proactive risk management	15	9 (69%)	6 (67%)	Use anti-virus program Back up personal data Be cautious when using public Wi-Fi Change password regularly Do not use or use less social media Take care of physical safety of credit card Use tape to cover computer camera Switch devices
Event-based risk management	8	5 (38%)	3 (33%)	Change email password when asked Do not accept many friend requests Do not give email address when asked Do not open pop ups Exit malicious website Not sign up or not log in
Controlling digital traces	15	10 (77%)	5 (56%)	Use anonymous search engine Use cookie blocker or other tracker blocker Cut off address from package Limit or change information shared online Delete cookies, caches, history Use private browsing mode Use fake accounts or multiple accounts
Securing connections	12	5 (38%)	7 (78%)	Encrypt data Watch for https in websites Use Tor Use password to secure Wi-Fi

Table 15. Protective actions used by lay participants and technical participants.

Although our technical participants were more knowledgeable of how the Internet works in the backend, they did not in general take more steps to protect their information online, in comparison with lay participants (Mean types of actions used by

technical participants = 2.33, lay participants = 2.23, *n.s.*). As shown in Table 3, the only difference was that technical participants were somewhat more likely to mention securing their connections than lay participants and the comparison shows a trend approaching significance (t [20] = 1.99, p = .07). This finding contrasts with some of the prior work that has shown a correlation between technical knowledge and privacy practices (Park, 2011), but this ostensible contradiction may stem from how we and other authors explored the influence of technical knowledge. Our study was focused on how people understand the Internet and its infrastructure whereas other studies (Hargittai & Litt, 2013; Park, 2011) have mainly focused on users' Internet literacy and their familiarity with privacy practices.

We counted the diversity of privacy threats that participants mentioned among six frequently-mentioned sources of threat in Figure 13: companies, third parties, government, hackers, other people, and ISPs. We then compared how perceptions of threat were related to protective action, by conducting a nonparametric correlation analysis on the number of threats they mentioned and the number of protective action types they took. The analysis yielded a moderate correlation (r_s = .40, p = .06). This result indicates that the awareness of privacy threats is probably a stronger indicator of people's protective actions than their general technical background. This comparison points to a difference in the impact of general technical knowledge, which does not seem to predict actions, and the awareness of Internet privacy risks.

Many participants had some knowledge of protective actions but had not used them. This may be one consequence of the privacy paradox (Acquisti et al., 2015) whereby people have general desire for privacy but do not act on this desire. We wanted to know what our participants would say about why they did not take steps to protect their information.

5.3.3.2 What prevents people from taking action?

Four categories emerged when participants talked about why they did not take actions to protect their information from being seen. The most common explanation was similar to the statement, "I've nothing to hide"(Solove, 2007).

Eleven participants (8 lay and 3 technical participants) said they were not worried about their information being accessed or monitored or did not have the need to use tools. Many participants were not concerned because they did not do anything very subversive, illegal, or had little to protect. T10 said, *"I don't care who sees and reads my email"* although he was aware that *"hackers can act as mail servers."* Two participants were not worried also because *"I don't put that much information out there."* (C03 and C04) Three participants said they had too little money to protect, *"I don't have much money to worry about."* (C03) C01 said he was not worried because his data is among *"an awful lot of data."* T11 said he knew a lot of methods that other people had used to mask their IP address, such as proxy servers, but he never pirated so much music that he felt the need

to do so. T04 mentioned Tor as a protective method during the interview, but also said, "*Till now I haven't had the need to use Tor.*"

The second reason given for not taking protective measures was that doing so would sacrifice effectiveness or convenience. T11 started to use DuckDuckGo (https://duckduckgo.com/), an anonymous search engine, to conduct anonymous searching but switched back to Google after several months, because Google gave better search results, tailored to his interests. T06 quit Facebook but did not quit Google, because "*their services are a lot more useful*." C06 said she is willing to take risks because doing things online is much more convenient than the "*old-fashioned way*."

Another reason given for not taking protective measures was the poor usability of privacy protection tools or software. T07 said that it is hard to do incognito browsing on smartphones. N10 knew that he could get a blocker but suspected some of the blockers might include viruses and would add clutter to his browsing experience.

For a minority, a feeling of helplessness and lack of procedural knowledge prevented them from taking any action (Shklovski, Mainwaring, Skúladóttir, & Borgthorsson, 2014). C05 said that hackers would probably hack into the website servers instead of individual users, and there was nothing he could do about it. Four lay participants said they lacked enough information to discuss actions they could do to prevent others' access to their information. C03 said he deleted cookies and then said, *"I don't know how to do anything else."*

The relationship of risk perception and action is also shown in participants' remarks. A technical background could influence awareness of threats and risks to some extent, but risk perception could also be shaped by personal experience. T11 started using DuckDuckGo after hearing about news related to Target's data breach and NSA monitoring. He became worried about how many people could see his information online. T11 had also been harassed by a Craigslist job poster because he gave out his phone number and email address. The Target data breach was also mentioned by C02, C07 and T11. After C07 was notified of the breach, she was not sure whether she was a victim or not, so she kept checking her statements carefully for a few months. Consistent with the findings in Chapter 2 and Chapter 4, these instances suggest that past negative experience triggers more secure online behavior and a heightened level of privacy concern. In contrast, people who had not experienced a negative event seemed to be habituated to the convenience brought by the Internet and were less motivated to take protective actions online. A community participant (C04) had a friend who experienced identity theft, but hearing about this story did not make him worry about his information, and he stated, *"unless it happens to you it's hard to walk in somebody else's shoes."*

5.4 Discussion

The findings of this study help us understand more deeply the differences between technical and nontechnical users' knowledge of the Internet. Technical education determined whether people viewed the Internet as a simple, service-like system or as an articulated technical system. The more technical participants had a more articulated model of the Internet and expressed greater awareness of the different people and organizations that could access their data. However, in this study, technical participants did not take more steps to protect their online information than those with less technical knowledge.

After the second round of data collection, we invited five networking and computer security experts to review several lay and technical participants' models and discuss implications for security and privacy.

5.4.1 The role of knowledge in privacy decisions

Previous research is unclear as to whether or not Internet knowledge is associated with better management of one's privacy and security. Our studies also show mixed results in the effect of technical knowledge in people's self-reported privacy protective actions: the surveys described in Chapter 4 show that more technical knowledge is associated with more hiding from informational threats and using technical methods to hide personal information. In this study, technical participants did not differ from nontechnical participants in the number of methods they reported using to hide from threats, although they were slightly more likely to use advanced technical methods to secure their connections than nontechnical participants were (consistent with the findings described in Chapter 4). Many technical participants said that they did not need to take actions to mitigate risk, and that the tools were inconvenient. These observations echo the finding described in Furnell et al. (2007) that technical users complained about practical factors that prevented them from taking secure actions (e.g., "security is too expensive"). Also, expert reviewers pointed out that technical participants might be overconfident about their knowledge, causing them to have a *"skewed view of security"*.

In comparison to general Internet knowledge, people's knowledge of privacy threats and risks might be more predictive of their privacy behaviors. Expert reviewers identified overlooking privacy and security risks as an important limitation of simpler mental models. They indicated that users who lacked awareness of Internet entities or organizations would have difficulty identifying the source of a problem or error when attacks, leaks, or other security issues occurred. One expert reviewer said that the lack of entity awareness in the simple mental model might engender too much trust in data privacy and security:

> *When it's just a magic black box, you tend to say well, I trust the magic black box, and so I would worry a little bit more that someone with this level of abstraction would not think*

> *as much about who could be sniffing on their communications or changing it or how they interpret security warnings and things like that.*

Our data supported this argument, by showing that participants with an articulated model, on average, expressed greater awareness of who could access their data than did participants with a simple model. The number of threats people identified seemed to be correlated with protective actions they took.

Another dimension of knowledge is that of privacy protection tools or systems. Expert reviewers were concerned that insufficient knowledge of encryption mechanisms could lead to data security risks. They speculated that users who were more aware of encryption would be better at controlling their data privacy and security. However, we did not find this association in our data. Participants who were more aware of protective mechanisms such as encryption or website certifications did not report taking more protective actions. There might be some skewness in our data because the majority of our participants were aware of protective mechanisms (17 out of the 22 we coded), so the relationship between knowledge of privacy tools and people's actual action requires further investigation.

5.4.2 Uncertainty in knowledge

Across all three rounds of data collection, participants expressed a great deal of uncertainty or lack of knowledge about how the Internet works, how their data is collected, shared or stored, what protective actions they could use, and whether the protection is effective or not. This finding echoes Acquisti et al.'s work (2015) demonstrating broad privacy uncertainty. For example, N11 used a Google app to block trackers but she was not sure how effective it was and was still concerned: *"I don't think it blocks everything."* Several nontechnical and community participants were confused about how attacks or problems happened. Finally, three technical participants expressed doubts about who had access to their data. These different uncertainties may prevent people from accurately estimating their privacy and security risks.

Another dimension of uncertainty in people's knowledge is whether or not their mental models can adapt to changes in technology. A few nontechnical participants' perception of the Internet seemed to be dominated by names of well-known content providers (e.g., "Yahoo", "Google", and "Facebook"). They also used name recognition as a safety heuristic—deciding that a website is secure because it is a well-known brand. However, advances in technology, security breaches reported in the press, and the rise of new companies could change these attitudes. As noted by one expert reviewer, participants did not seem to update their models as fast as the Internet changed. Only a few participants expressed awareness that their models might be outdated.

5.4.3 Limitations

Because we used a think-aloud style qualitative study, our observations were influenced by the questions we posed and the knowledge people recalled. Participants may have had more knowledge of the Internet or security mechanisms than they expressed. Another limitation of conducting a qualitative study is that we have a comparatively small sample size. The small sample size may prevent us from detecting small but real effects of declarative and procedural knowledge on motivations and behavior.

In the next chapter, I will describe an online experiment with a larger sample size. Participants in the next study were given different visualizations that mirror the two main components of Internet knowledge identified in this chapter (structure and entity awareness). The online experiment will allow me to draw causal inferences between manipulations of Internet knowledge and people's awareness of privacy threats, and their privacy protection actions.

6 The effect of privacy threat visualization on people's behavioral intentions and actual behavior

6.1 Introduction

As the Internet becomes more and more complex, system designers and researchers face the question of what people need to know about the Internet to protect their privacy and security. As a means of avoiding unintended information disclosure, some argue for greater transparency about what others can do with their data, (Solove, 2007), whereas others warn of transparency's potential tradeoffs, including an increase in complexity and demands on people's attention and cognitive capacity (Stuart, Dabbish, Kiesler, Kinnaird, & Kang, 2012). User-centered design principles dictate that system status be made visible, that feedback be provided, and that systems should be mapped to users' mental models of the system (Norman, 1988). When a system is too complex, however, greater visibility could cause information overload, increase feelings of uncertainty, and, sometimes, backfire.

Privacy and security researchers have explored many ways to provide greater system transparency by increasing user awareness of information leakage over wireless networks (Kowitz & Cranor, 2005), smartphone applications (Almuhimedi et al., 2015; Balebako, Jung, Lu, Cranor, & Nguyen, 2013), and social network sites (Wang et al., 2014). Much of this work found no direct evidence that increased transparency leads to more secure behavior online; an exception is Almuhimedi et al. (2015), who found that showing people the frequency of data access by smartphone apps prompted them to review the apps' permission settings. It is not clear how long this influence lasted. My previous studies show that knowing more about the Internet and data practices does not necessarily lead to more secure user behavior online.

The results from the two surveys reported in Chapter 4 show that higher levels of educational attainment and greater technical knowledge was associated with hiding from informational threats and using technical methods to do so; however, we saw no such connection between knowledge and hiding from social threats or using behavioral

methods to hide. Further, in the study of people's mental models of the Internet (described in Chapter 5), we categorized users' knowledge of the Internet along two dimensions: their awareness of the structural components of the Internet, and their awareness of privacy threats (i.e., of those who had access to their data). Lay participants in the study tended to comprehend the Internet in a simple, abstract way, and could not spell out the entities involved in the process of delivering or receiving content over the Internet. Technical participants had a far more articulated mental model of the Internet. Despite these differences in knowledge, the two groups did not differ in the actions they took to protect their personal information, except that the more technical participants were slightly more likely to secure their connections using advanced technical methods. These findings were based on self-report, correlational data. We do not know whether or not the behaviors identified by participants reflect actual privacy protection behavior online, nor do we know whether the two types of knowledge cause a different privacy-decision-making process.

The goal of this chapter is to answer this causal question: Will increased system transparency – visualizations of underlying structure and data access – change people's privacy perceptions and behavior?

6.1.1 Two kinds of privacy risks

Prior theoretical frameworks suggest that privacy awareness ("the extent to which an individual is informed about organizational privacy practices," Smith et al. 2011, p.998) might lead to privacy concerns and, in turn, lead to behavioral reactions (Malhotra et al. 2004; Smith et al. 2011). In the interview-based study discussed in Chapter 2, most interviewees reported hiding their online information from a specific source of threat, reflecting a personal "threat model" of individuals or organizations. In this study, I extended the scope of privacy awareness to include perceived personal information access not only by organizations but also other individuals such as one's family and friends, employers, or other users online. Informational privacy threats in the general model are represented here by a variable, *perceived organizations' access to data*, and social privacy threats are represented by a variable, *perceived individuals' access to data.*

As noted in Chapter 1, theoretical discussions of users' privacy practices focus mainly on informational privacy threats, caused by organizations (government, companies) or hacker groups. In recent years, researchers have started to pay more attention to privacy protection behaviors aroused by social privacy threats such as arise from intentional or unintentional disclosure on social networking sites. Not much work has compared these two kinds of threats at once, and examined their importance in influencing users' privacy protection behavior. A few researchers have pointed out that users might not be very concerned about informational privacy threats because "those data are only accessible by government officials or computer programs" (Solove, 2007), or because they feel "lost in the crowd" (Nguyen et al., 2008). One study shows that university-

student Facebook users are more concerned about social privacy than informational privacy, and that they take more actions to protect their social privacy than to restrict companies or third parties from accessing their information (Young & Quan-Haase, 2013).

6.1.2 Theories of attitudes, intentions and behavior

To understand the link between perceived social and informational privacy risks and privacy protection behavior, I draw from two theories that address how attitudes and intentions are connected to behavior: the theory of planned behavior and fear appeals theory. The theory of planned behavior (Ajzen, 1991) predicts that attitudes towards a behavior (whether the person has a favorable or unfavorable evaluation of the behavior), subjective norms (perceived social pressure to perform or not to perform), and perceived behavioral control (similar to the concept of self-efficacy) predict a person's intentions to perform a behavior. Actual behavior is predicted by both the intention to perform the behavior and perceived behavior control. Non-motivational factors such as skills, time and money are assumed to affect perceived control (Ajzen, 1985). The theory differs from other attitude-behavior theories in its emphasis on people's concrete beliefs and intentions. For example, whether a student cheats on a test, according to the theory, depends more on the student's attitudes about cheating on a particular test in a particular course rather than on the student's general attitude about cheating.

Fear appeals theory emphasizes the role of emotions in predicting behavior. For instance, fear and anger affect people's estimations of risks and policy preferences (Lerner, Gonzalez, Small, & Fischhoff, 2003). Researchers have used the theory of fear appeals to elicit a sense of threat in messages intended to persuade people to adopt certain healthy behaviors (Witte, 1994). The theory has been employed in the field of information security (LaRose, Rifon, & Enbody, 2008), in studies on spyware (Johnston & Warkentin, 2010) and passwords (Vance, Eargle, Ouimet, & Straub, 2013). Fear appeals theory proposes that people go through two stages of information processing when they receive a fear-appeal message. First, they evaluate the threat severity (e.g., "It would be a serious problem if my computer were infected by a virus") and their own susceptibility to the danger (e.g., "It is possible that my computer will be infected"). Then, if they perceive the threat to be high, they will execute the second step, evaluating their own efficacy in averting the threat (e.g., "I can use anti-virus software to protect my computer") and response efficacy (e.g., "I believe anti-virus software can protect my computer"). When the threat and efficacy are high, people will activate a danger-control process: they will believe the message and perform actions to mitigate the threat. When the threat is high but efficacy is low, they will be afraid but, instead of taking action, they will use emotional coping strategies such as avoidance and denial.

Taken together, these theories suggest that perceived privacy threats might act together with perceived efficacy to influence people's behavioral intentions to protect themselves from privacy threats and to change their behavior by taking threat-mitigating actions.

6.2 Hypotheses

An experiment was designed to test the main argument that providing people with more structural knowledge of the Internet and Internet threats would change their perceptions of risk, their intentions, and their actual behavior.

H1a. People who are shown visualizations of how the Internet works (structural knowledge) will gain a greater awareness of social and informational privacy risks (i.e., others' access to their personal data) than those not shown such visualizations.

H1b. People who are shown visualizations of privacy threats on the Internet will gain a greater awareness of social and informational privacy risks (i.e., others' access to their personal data) than those not shown such visualizations.

H2a. People who are shown visualizations of how the Internet works (structural knowledge) will intend to protect against privacy threats more than those not shown such visualizations.

H2b. People who are shown visualizations of privacy threats on the Internet will intend to protect against privacy threats more than those not shown such visualizations.

H3a. People who are shown visualizations of how the Internet works (structural knowledge) will be more likely to make decisions to protect their privacy more than those not shown such visualizations.

H3b. People who are shown visualizations of privacy threats on the Internet will be more likely to make decisions to protect their privacy more than those not shown such visualizations.

6.3 Method

I designed and conducted a between-subject experiment. Its purpose was to examine the effects on people's privacy perceptions and actions of showing them visual models of the Internet. The models were presented as simple or articulated visualizations of the flow of information when someone connects to a website for a job search. In some conditions, the visualizations also depicted sources of threat to privacy. All visualizations were based on the results of the study of people's mental models described in the previous chapter.

The experiments were instrumented in online surveys built on Qualtrics (https://www.qualtrics.com/). Participants were told that they would answer questions about using the Internet in a hypothetical scenario. Using a hypothetical scenario to induce participants' privacy decisions has been used in prior research (Ackerman et al., 1999; Malhotra, Kim, & Agarwal, 2004). In the current work, participants were asked to read a scenario in which they imagined themselves looking for a job. They were to imagine using the Internet in a public coffee shop to conduct a job search, and to assume they had found a job search website http://www.idealjobs.com (a fictional web address). Then they were asked about their likelihood of submitting personal information in the registration form of that website to see available jobs. The description of the scenario is attached in Appendix II.

Participants were exposed to different experimental manipulations embedded in their surveys, and they answered a series of survey questions measuring their estimated likelihood to disclose information, awareness of information access, privacy concerns, self-efficacy, risk perception, estimated likelihood of using privacy protection strategies, privacy knowledge, and demographic information. I also added questions to measure people's own tendency to disclose personal information online and to learn about methods to protect their privacy and security. The survey flow is summarized in section 6.3.3.

6.3.1 Experimental design and participants

The experiment was designed as a 2 (*Threat viz* [With vs. Without]) × 3 (*Structure viz* [Articulated vs. Simple vs. No structure]) factorial. The *control* condition was the *no threat, no structure* condition, in which participants saw no visualization at all. In the *simple viz condition*, participants only saw a very simple illustration of the Internet's structure, showing their computer connecting to the Internet and then connecting to the website server (Figure 14). In the *articulated viz condition*, participants saw a visualization with the addition of a router, other computers, and multiple levels of the ISPs in between their computer and the website server (Figure 15). In the *simple threat viz condition*, participants saw the simple visualization of the Internet with a list of potential threats showing who could see their data placed next to the simple Internet structure (Figure 16). In the *articulated threat viz condition*, participants saw the articulated visualization with the same list of threats placed next to the articulated Internet structure (Figure 17). In the *threat only viz condition* (Figure 18), participants saw only a list of threats but no visualization of the Internet's structure.

Each participant was randomly assigned to one of the six conditions. In the control condition, participants were first presented with an introduction page, then an attention check question used in (Egelman & Peer, 2015) to ensure they pay full attention to the survey instructions, then the Internet scenario, and then survey questions measuring the dependent variables (described in the following section). In the visualization conditions,

participants were instructed to help evaluate a new browser plug-in that tells users how they connect to the webpage through the Internet. These participants were first presented with the introduction page, the attention check question, followed by the Internet scenario, and then they saw one of the four visualizations in the form of a browser plug-in prototype, and then survey questions measuring the dependent variables.

I recruited 271 valid responses from Amazon Mechanical Turk. The survey was described as a "survey about Internet use." Each participant was paid $2 for completing the survey (a typical compensation rate for MTurk workers is $6/hour). I required MTurk workers to be located in the U.S., have at least a 95% approval rate, and have at least 100 approved HITs. The average time participants spent on the survey was 24 minutes. To examine behavioral changes, the experiment included two surveys. One week after each participant completed the first survey, the participant received an email containing a link to a new survey. The follow-up survey paid $1, and the average time participants spent on the survey was 6 minutes. The response rate of the follow-up survey was 73% (199 out of the 271 participants). Data from the first survey was collected from 6/24/15 to 7/1/15, and the follow-up survey was collected from 07/01/15 to 07/07/15. Table 16 shows the demographic characteristics of the participants.

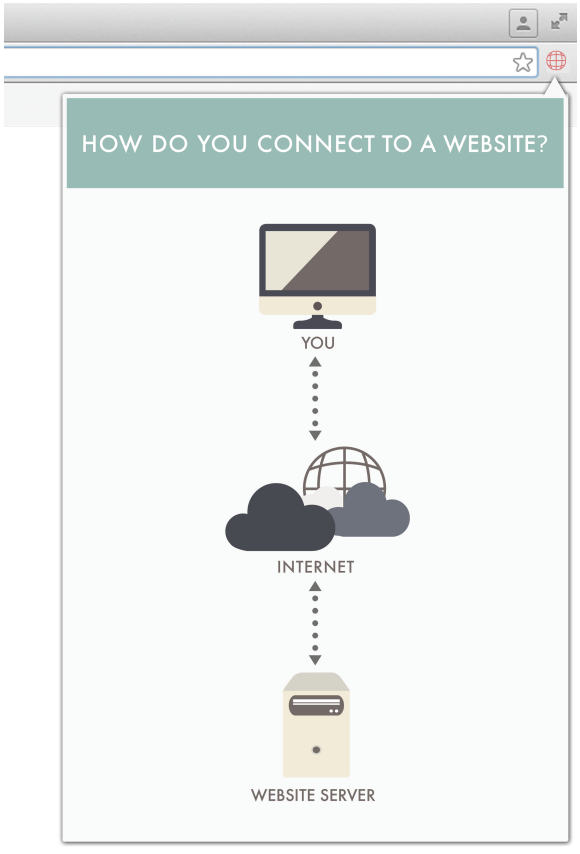


Figure 14. Simple visualization

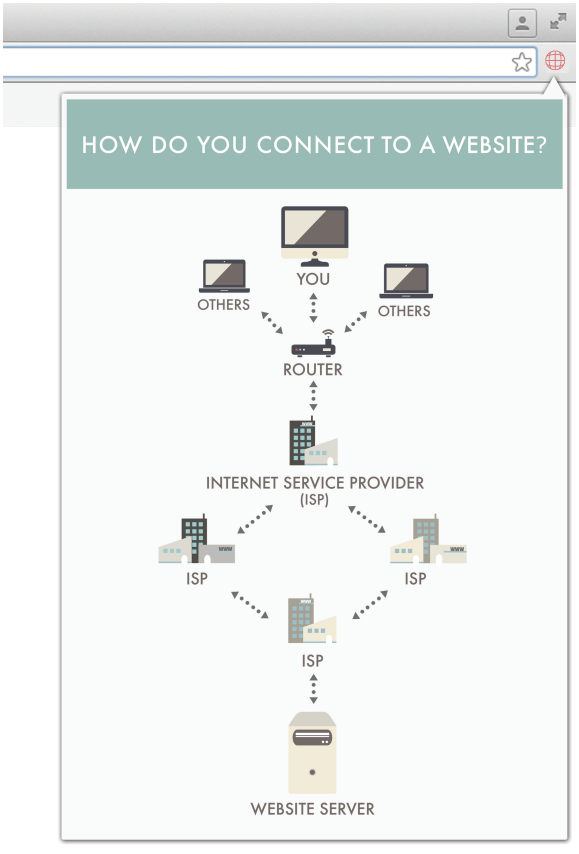


Figure 15. Articulated visualization

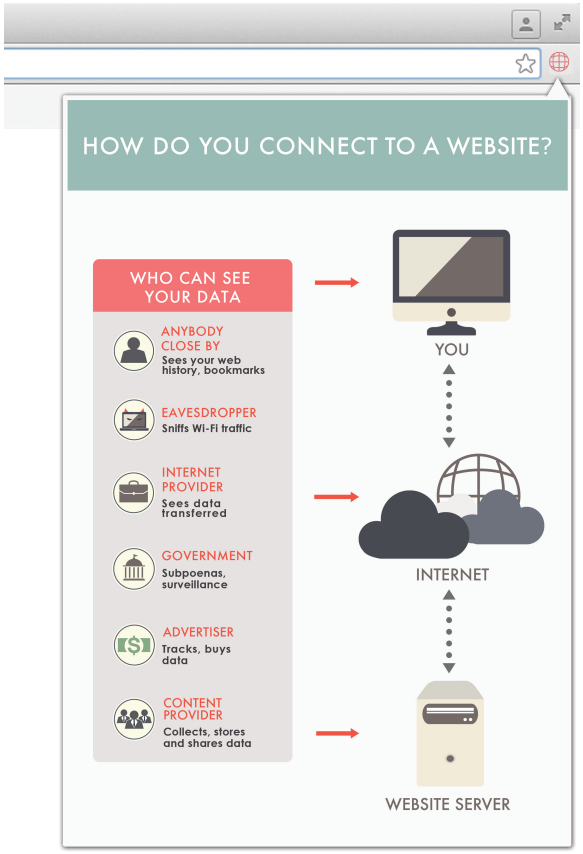


Figure 16. Simple threat visualization

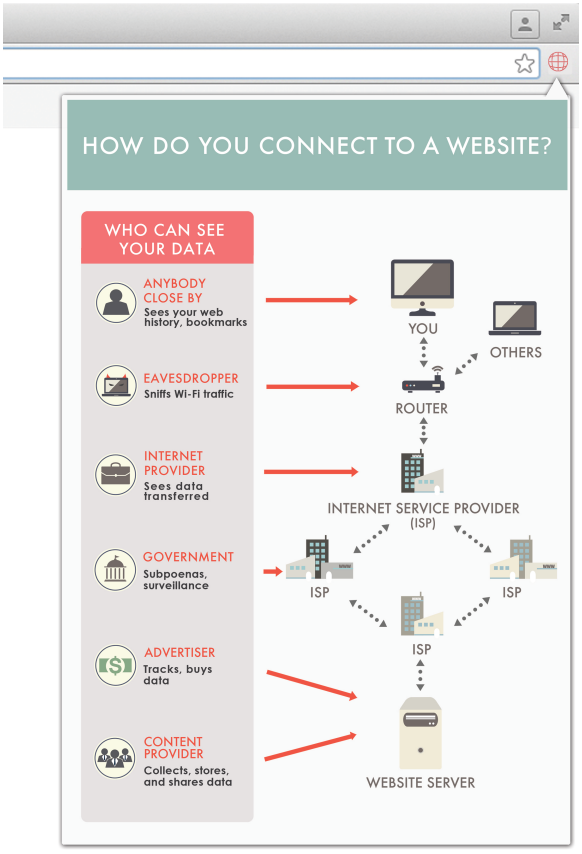


Figure 17.Articulated threat visualization

Figure 18. Threat only visualization

Number of participants in each condition	
Control condition	47
Simple viz condition	45
Simple threat viz condition	46
Articulated viz condition	43
Articulated threat viz condition	46
Threat only viz condition	44
Demographic characteristics	
% of participants with CS background	17%
% of social media user	91%
Mean age [min, max]	35 [19, 70]
Gender	
Female	50%
Male	50%
Education	
High school or less	15%
Some college	42%
College and more	42%
Marital status	
Married	34%
Living with Partner	11%
Divorced, separated or widowed	9%
Never married	44%
Decline to answer	2%
Employment status	
Employed full time	61%
Employed part time	17%
Not employed	20%
Decline to answer	2%
Political view	
Liberal	44%
Moderate	24%
Conservative	19%
Libertarian	9%
Decline to answer	2%
Income	
Under $40,000	34%
$40,000 to $75,000	40%
$75,000 or more	17%
Decline to answer	8%

Table 16. Number of participants in each condition and demographic characteristics. (N = 271) (The percentage of "Decline to answer" lower than 1% is omitted in this table.)

6.3.2 Variable definitions

The survey presented a series of questions to measure participants' evaluation of the visualizations, their behavioral intentions, their awareness of potential data access by others, their privacy protection behaviors, their knowledge of the Internet, their privacy concerns, their risk perceptions, their policy preferences, actual disclosure behavior and secure decision measures, and demographic questions. A complete version of the survey is shown in Appendix IV. In the following paragraphs I explain the main dependent variables and predictor variables used in analysis. Table 17 below presents the descriptive statistics of all the measures used in the experiment.

6.3.2.1 Visualization evaluation

Participants were asked to evaluate the visualization they saw based on its similarity to their understanding of the Internet, on how helpful the information in the visualization was for them to learn about how the Internet works, and how clearly the visualization communicated how the Internet works. In analysis, I combined the helpfulness rating and clearness rating into an "informative" scale (α = 0.80; Table 17).

6.3.2.2 Behavioral intentions

Estimated likelihood of disclosing information: Participants answered three questions measuring their behavioral intention to disclose information in the main survey and the follow-up survey. Similar approaches have been used in previous research (Ackerman et al., 1999; Leon et al., 2013; Malhotra et al., 2004). In our survey, I asked participants their likelihood of conducting a job search in a coffee shop, their likelihood of visiting the www.idealjobs.com website, and their likelihood of revealing ten types of personal information in the registration form. The ten types of personal information included nonsensitive information, such as hobbies, and sensitive information, such as current financial status. (These questions were drawn from actual job application forms.) As shown in Table 17, participants' average estimated likelihood of revealing their digital traces depended on the type of trace. Their likelihood of visiting the provided website was significantly higher than their likelihood of conducting a job search in a public coffee shop (t [267] = 4.23, $p < .0001$) and the average likelihood of disclosing personal information (t [267] = 6.22, $p < .0001$). The likelihood of conducting a job search was also higher than the likelihood of disclosing personal information, but the comparison is not significant (t [267] = 1.49, $p = .14$). Because these items were highly correlated with each other, and the Cronbach's α for a scale with all 12 items is 0.92, I combined the items into one construct in the analysis: *estimated likelihood of disclosing information.*

Estimated likelihood of using protective strategies: In the fear appeals literature, behavioral intentions are usually measured by survey questions such as "I plan to use anti-spyware software in the next 3 months" (Johnston & Warkentin, 2010). In this survey, I adapted the items from the survey items used in Chapter 4 and asked

participants to estimate how likely they were to use each of a list of tools if they were to use public WiFi in a coffee shop in the future. The Cronbach's α for one scale including nine items is 0.84, so I combined these items into one construct: *estimated likelihood of using protective strategies.*

6.3.2.3 Perceived data access

Participants were asked to estimate the likelihood that each of ten different groups or organizations could access their search history, that they had visited the job website, and the personal information they submitted to the website. Principle component analysis and exploratory factor analysis of the ten types of different audience groups generated two types of audience: individuals (family and friends, employer, other people in the same network, and other users on the site); and organizations (advertisers, government or law enforcement, hackers, the ISP, and browser). We combined participants' perceived access of other individuals or organizations to the three types of digital traces. As a result, we have two awareness measures: *perceived individuals' access to data* (α = .76), and *perceived organizations' access to data* (α = .75). Participants' awareness of the visited website's access to their data was not included because of its low reliability within the scale (α = .55). This set of questions was asked in both the main survey and the follow-up survey.

6.3.2.4 Actual behavior

Tendency to learn about protective strategies: Participants were presented with a video selection task, in which they were asked to choose one of two videos to evaluate. One video was about protecting privacy and security online; the other video was about conducting an effective job search online.

Number of disclosure questions answered: In the experiment, on the next page following the video selection task, participants were presented with four disclosure questions adapted from Brandimarte, Acquisti, & Loewenstein (2013). Similar self-disclosure questions were also used in Joinson, Reips, Buchanan, Schofield, & Carina (2010). Two questions asked for sensitive information such as whether or not participants had used drugs, and the other two questions asked non-sensitive questions such as whether or not participants had flown in an airplane. In the end of the follow-up survey, participants saw another four disclosure questions including two sensitive questions and two non-sensitive questions.

6.3.2.5 Perceived self-efficacy

According to Rhee et al (2009), self-efficacy is positively correlated to behavioral intention and security practices. We adapted their questions and modified the description to evaluate people's self-efficacy to hide their information, such as "I feel confident that I can mask my IP address"; "I feel confident that I can prevent others from seeing which websites I visited"; "I feel confident that I can communicate with others anonymously online, without revealing my real identity at all"; "I feel confident that I can prevent unwanted access to my personal information online"; "I feel confident

that I can delete my digital traces (e.g. social network account, something I've posted in the past)"; "I feel confident protecting my privacy online."

6.3.2.6 General privacy concern

It is commonly known that privacy concern is associated with people's privacy protection behavior and attitudes. We adapted six items from the IUIPC scale (Malhotra et al., 2004) to measure people's general privacy concern. This variable was used as a predictor variable in our analysis.

6.3.2.7 Technical knowledge

I used the same knowledge survey as used in the studies in Chapter 4 and Chapter 5. The knowledge scale was calculated based on the standardized mean score of the two scales (familiarity rating of technical terms and T/F questions). Participants' performance in the knowledge scale is used as a proxy for their technical background in the analysis. Participants with self-reported CS background had significantly higher knowledge scores than those without CS background (t [269] = 6.28, $p < .001$).

	Measure	**M**	**SD**	**α**
Visualization evaluation				
How informative the visualization is to communicate how the Internet works		3.65	0.99	0.76
Helpfulness of the visualization	1 item	3.52	1.13	
Clearness of the visualization	1 item	3.77	1.06	
How similar the visualization is to their understanding of the Internet	1 item	3.64	1.00	N/A
Behavioral Intentions				
Estimated likelihood of disclosing information		2.94	0.93	0.92
conduct job search in the scenario	1 item			
visit the job search website	1 item			
disclose personal information	10 items			
Estimated likelihood of using protective strategies	9 items adapted from (Rainie et al., 2013)	3.39	0.82	0.84
Actual disclosure questions	Adapted from (Brandimarte et al., 2013)			
Sensitive questions	Do drugs	39%	49%	N/A
	Download pirated material	61%	49%	
	Gain access to other's email	28%	45%	
	Cheat on partner	22%	41%	
Non-sensitive questions	Lie about age	44%	49%	
	Have flown on airplane	79%	41%	
	Turn lights out	88%	32%	
	Donate to NGO	84%	37%	
Perceived data access				
Other people can see (employer, family, other people, users on site)		2.54	0.87	0.76
your web history	4 items			
you visited the site	4 items			
information you submitted	4 items			

Other organizations can see (advertiser, government, hacker, ISP, browser)		4.16	0.66	0.75
your web history	5 items			
you visited the site	5 items			
information you submitted	5 items			
General privacy concern	6 items adapted from IUIPC (Malhotra et al., 2004)	3.57	0.92	0.90
General risk perception*				
Risk belief	3 items adapted from (Malhotra et al., 2004)	3.68	0.80	0.84
Estimated likelihood to experience bad events	10 items adapted from (Rainie et al., 2013)	14%	13%	0.88
Perceived self-efficacy of hiding	6 items revised from (Rhee et al., 2009)	2.84	0.92	0.89
Technical knowledge	Adapted from (Kang, Dabbish, Fruchter, & Kiesler, 2015)	0	0.88	0.93
Technical term familiarity	14 items [5-point Likert]	3.23	0.82	
True and False questions	6 items [# of correct items: 0~6]	2.37	1.80	
Policy preferences **	Adapted from (Rainie et al., 2013) and (Madden, 2014)			
Perceptions	"It's possible to be anonymous"	2.76	1.41	N/A
	"It's difficult to remove inaccurate information"	4.38	0.90	N/A
Attitudes	"People should be able to be anonymous"	4.10	1.20	N/A
	"The 'right to be forgotten' law would be useful to me"	3.39	1.59	N/A

*Risk perception measures are not used in the analysis to avoid multicollinearity, because of its high correlation with general privacy concern. **Policy preferences are not shown in the analysis because there is no difference between conditions.

Table 17. Descriptive statistics of survey measures. (All measures used 5-point Likert scale, except the estimated likelihood to experience bad events and number of T/F questions in knowledge measures.)

6.3.3 Survey flow

The order in which participants saw the survey questions is represented in Figure 19 below. The main dependent variables I'm interested in are the perceived data access and behavioral measures (intentions and actual behavior). In order to test the main effect of visualization manipulations on behavioral intentions, I placed the disclosure intention questions (the main behavior DV) directly after the visualization manipulations. The measure of perceived data access includes 30 questions, which may require a higher cognitive load, so I placed that item to be after the disclosure intention questions. Other questions were placed afterward to reduce the experimenter demand effects, because the wording of those questions is more closely related to privacy. Actual behavioral measures were placed in the end of the survey to avoid early dropouts, because some

disclosure questions were intrusive in nature, which might stop participants from finishing the survey. Questions that participants answered earlier will prime their answer to other questions later, and this is noted as a limitation of this study.

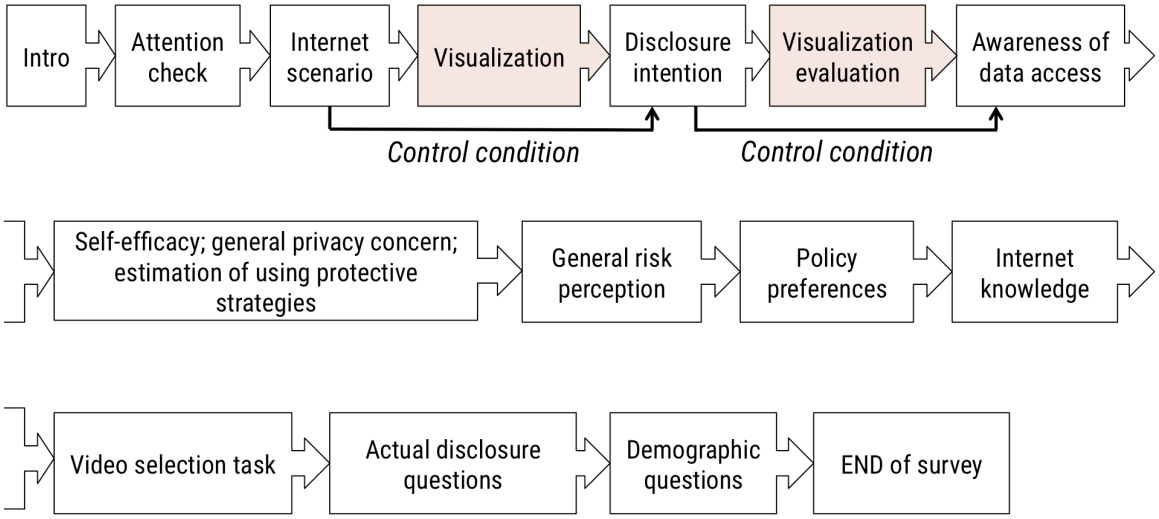


Figure 19. Survey flow.

6.4 Results

The results presented here are based on 271 participants who completed the first survey.

To test H1, H2, and H3, I first conducted analysis of variance. To test the potential additional effects of perceived privacy risks and perceived self-efficacy on behavioral intentions and actual behavior (H4, H5 and H6 explained later), I used the approach of Structural Equation Modeling (SEM) to analyze the relationships among measures.

6.4.1 Evaluation of the visualizations

To verify the finding described in Chapter 5 that technical participants' mental models of the Internet are articulated and nontechnical participants' mental models are simple, I compared participants' own evaluations of how similar the visualizations were to their understanding of the Internet. I conducted a two-way ANOVA analysis using the bivariate knowledge variable and the visualization conditions. (The no visualization condition was not included because those participants did not see any visualization and did not answer this question.) The results are a significant effect of knowledge background on participants' similarity ratings (F [1,200] = 17.40, $p < 0.01$) and a significant interaction effect of knowledge background and the visualization conditions (F [4, 200] = 3.54, $p < .01$). Figure 20 plots the similarity rating from participants with higher or lower knowledge background in different conditions. Results of student's t-tests are shown in the figure using different letters to represent statistical differences at the .05 level. The articulated visualization was evaluated as most similar to the

understanding of more knowledgeable participants, and the simple visualization was evaluated as most similar to the understanding of participants with less technical knowledge.

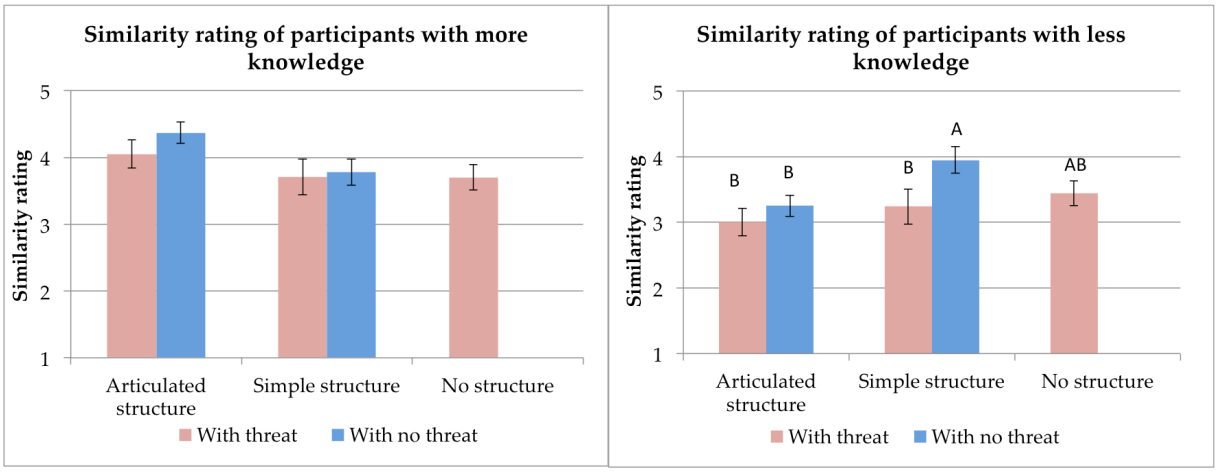


Figure 20. Participants' rating of how similar the visualization was to their own understanding of the Internet. Means with different letters are significantly different ($p < .05$)

I then examined participants' ratings of how informative the visualizations were. The analysis shows a marginally significant effect of visualization conditions (F [4,214] = 2.04, $p < .10$) and there is no significant effect of participants' knowledge background or an interaction effect. Figure 21 plots participants' ratings of informativeness against their ratings of how informative the visualizations were. Simple visualization had the lowest ratings of all conditions.

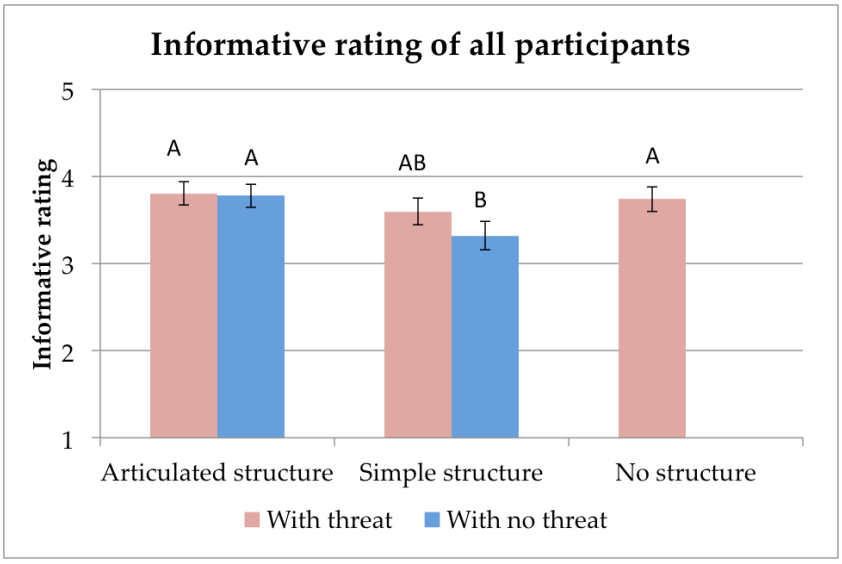


Figure 21. Participants' rating of how informative the visualization was. (Mean of the clearness rating and the helpfulness rating.)

6.4.2 The effects of the visualization manipulations on DVs

I conducted two-way ANOVA analyses to test the effects of the visualization manipulations (structure viz and threat viz) on the main dependent variables of perceptions and behavior: perceived threats (perceived data access of individuals and organizations), behavioral intentions (estimated likelihood of disclosing information and estimated likelihood of using protective strategies), and actual behavior (observed information disclosures and whether or not they selected the privacy video).

6.4.2.1 The effect of manipulations on perceived privacy risks

The result shows a marginal effect of structure visualization (F [2,264] = 2.45, p = .09) and a significant main effect of the threat visualization (F [1,264] = 10.56, p < .01) on the perceived individuals' access to data. Participants who saw an articulated Internet structure had marginally lower awareness of other individuals' access than those who saw a simple structure and those who saw no structure. Participants who saw visualizations of privacy threats had higher awareness of other individuals' access to their data than those who did not saw threat visualizations. The interaction effect between threat viz and structure viz is not significant. The structure viz does not show any significant main effect or interaction effect on perceived organizations' access, but the threat viz has a significant effect on the perceived access of organizations (F [1,265] = 7.31, p < .01). To summarize, **H1a** is not supported, because the articulated structure visualization shows an opposite effect than what I hypothesized on perceived social risks, and does not show any effect on perceived informational risks. The findings support **H1b**: people who are shown visualizations of privacy threats will have a higher awareness of social and informational privacy risks than those not shown such visualizations (Figure 22).

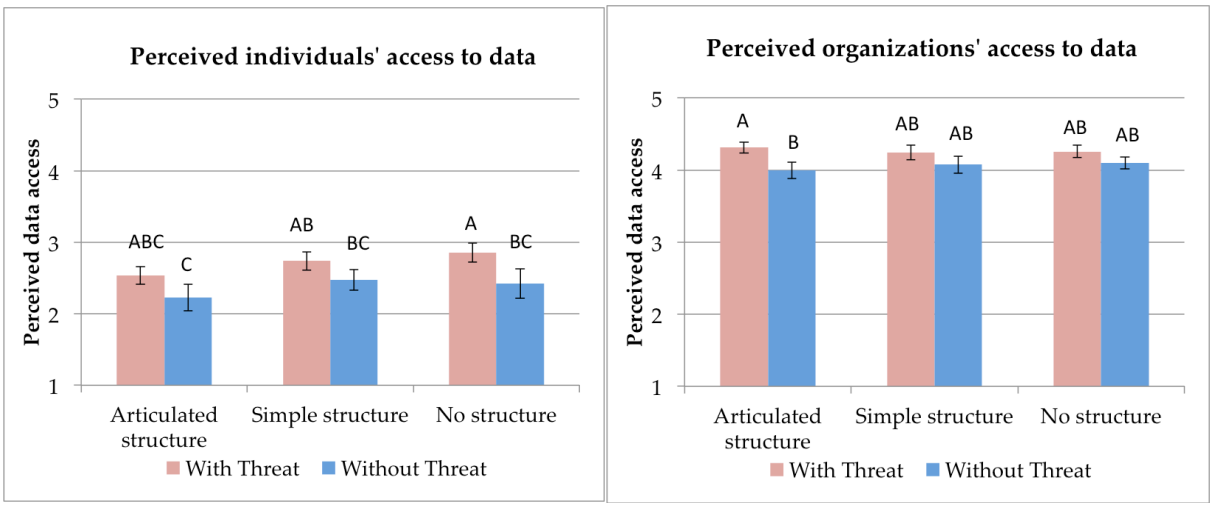


Figure 22. Perceived data access of other individuals and organizations. Means with different letters are significantly different (p < .05).

6.4.2.2 The effect of the manipulations on behavioral intentions and actual behavior

To test H2a and H2b (people who are shown visualizations of the Internet structure and privacy threats will intend to protect their privacy in the future more than those not shown such visualizations), I measured two types of behavioral intentions: people's estimated likelihood of disclosing information, and estimated likelihood of using protective strategies. The analysis of variance shows a significant main effect of the threat visualizations on the mean estimated likelihood of disclosing information (F [1,265] = 14.72, $p < .001$), but there is no significant effect of the structure visualization or the interaction effect. However, when I test the effect of manipulations on the estimated likelihood of using protective strategies, none of the effects is significant. **H2a** about the effect of structure visualizations on behavioral intentions is not supported. **H2b** is partly supported: people who saw visualizations with privacy threats had a lower estimation of disclosing personal information than those who did not see such visualizations (Figure 23).

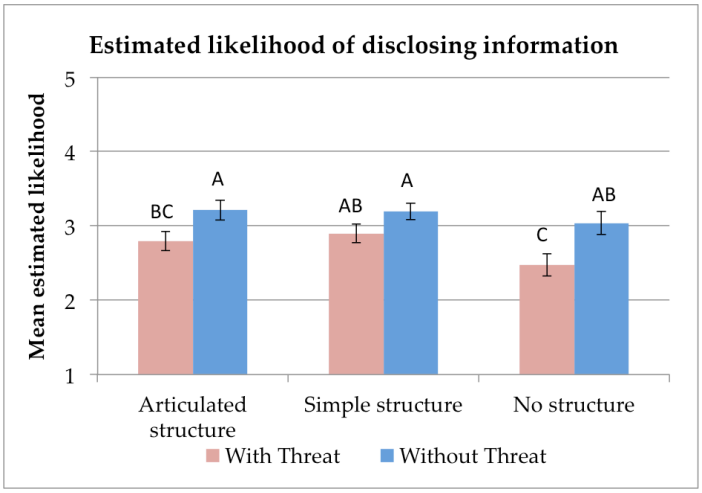


Figure 23. Participants' mean estimated likelihood of disclosing information. Means with different letters are significantly different (p < .05).

To test H3, I inserted two measures of actual behavior: the number of questions people answered in four disclosure questions, and their decision of whether or not to learn about privacy-protection tools. There is no significant effect of either structure visualizations or threat visualizations on both variables. **H3a** and **H3b** are not supported.

6.4.3 A structural model

The analyses of variance show that the visualization manipulations had significant effects on the participants' perceptions of others' access to their data and on participants' behavioral intentions, but not on their actual behavior. Based on the theoretical model of this thesis (Figure 1) and the behavioral theories introduced in the section 6.1.2, there might be some complex effects of the perceived risks (measured by perceived data

access in this study) and perceived efficacy on people's behavior. In order to test this possibility, I used an SEM approach to test the hypothesized model below. SEM requires hypotheses for expected relationships. The additional hypotheses (H4, H5, and H6) are presented below and in Figure 24.

H4. People with a greater awareness of privacy risks will intend to protect their privacy in the future more than those with less awareness of privacy risks.

H5. People with higher self-efficacy will intend to protect their privacy in the future more than those with lower self-efficacy.

H6. People who intend to protect their privacy in the future will be more likely to protect their privacy.

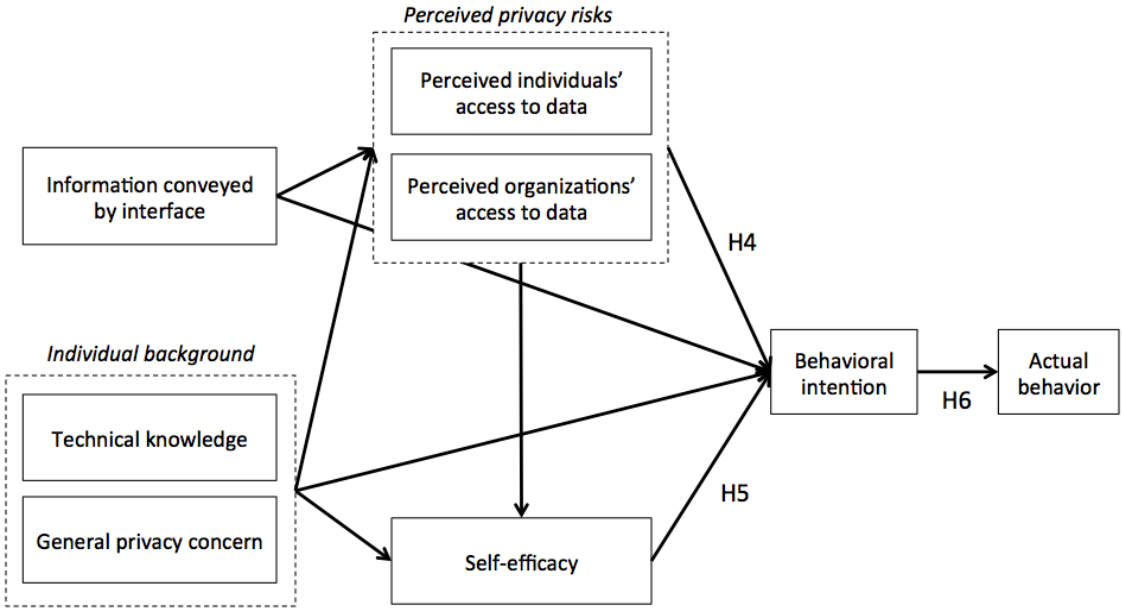


Figure 24. Hypothetical research model.

Table 18 below shows the correlations among measures used in SEM analysis. Because most of the survey measures I used were adapted from previous papers, I only present inter-construct correlations in this table.

	1	2	3	4	5	6	7	8
1. General Privacy Concern								
2. Technical Knowledge	-0.02							
3. Self Efficacy of Hiding	**-0.14***	**0.27*****						
4. Perceived Individuals' Access to Data	**0.18****	**-0.12***	**-.11**†					
5. Perceived Organizations' Access to Data	**0.27*****	-0.08	**-.29*****	**0.38*****				
6. Estimated Likelihood of Disclosing	**-0.30*****	-0.02	**.21*****	**-0.24*****	**-0.14***			
7. Estimated Likelihood of Using	**0.38*****	**0.24*****	**.21*****	0.09	0.10	**-0.25*****		

Protective Strategies								
8. Observed Information Disclosures	-0.06	0.09	.09	**-0.11†**	-0.06	0.04	0.01	
9. Selected Privacy Video	**0.10†**	0.07	**-.13***	-0.03	**0.14***	-0.07	.06	-0.02

Table 18. Correlations among measures.

The result of the SEM analysis is shown in Figure 25. The goodness of fit indices indicate that the model fits the data moderately well: $X^2(270) = 58.48$, $X^2/df = 2.25$, $p = .00$, GFI = .96, AGFI = .91, CFI = .87, RMSEA = .068. GFI, AGFI, and CFI values higher than 0.90 reflects an excellent fit (Bentler, 1989). RMSEA that is less than 0.08 reflects reasonably good model fit (Browne & Cudeck, 1993).

The findings support the relationships between perceived privacy risks, perceived self-efficacy and behavioral intentions (H4 and H5), but the relationship between behavioral intentions and actual behavior (H6) is not supported.

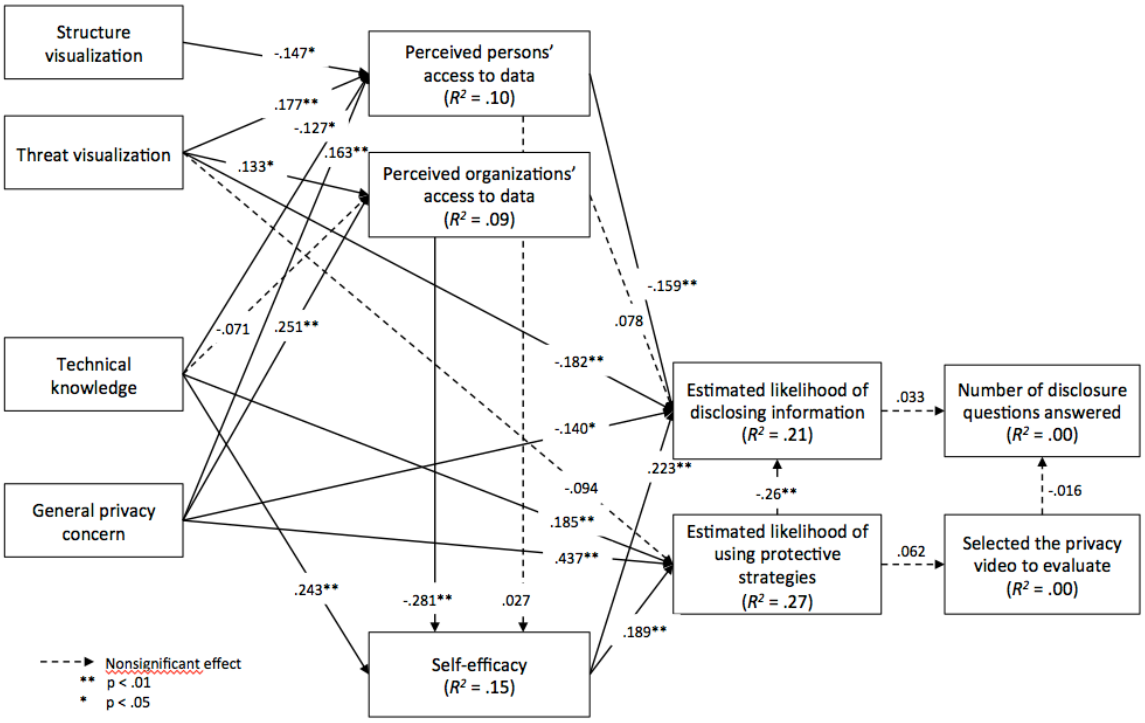


Figure 25. Results of SEM model testing. (Numbers on the lines are standardized beta path loadings. Non-significant paths are not shown in the figure.)

First, as shown in Figure 25, threat visualization and participants' general privacy concerns are both positively and significantly associated with participants' perceived access to their personal information by other individuals and organizations (supports H1b as in the previous analysis in section 6.4.2.1). In addition, structure visualization (the effect is mainly from articulated visualization vs. the other two visualizations) is negatively associated with perceiving other individuals' access to data. This result was

not expected, and suggests the complexity of the Internet might make some people perceive fewer risks and feel safer. Structure visualization does not predict perceived organizations' access to data, so I do not include that link in the model. Technical knowledge, surprisingly, is negatively associated with perceived individuals' access to data, suggesting that those who had higher technical knowledge had lower awareness of other individuals' access to their online data. The relationship between technical knowledge and the perceived organizations' access to data is not significant.

Self-efficacy of hiding information is positively predicted by technical knowledge and negatively predicted by perceived organizations' access to data. Those who were more likely to think institutions such as government and companies had access to their data were less confident about hiding their information online. However, we do not see a parallel connection between perceived individuals' access and self-efficacy.

Lastly, I examined the effects of perceived data access, self-efficacy, the manipulations and individual background on behavioral intentions and actual behaviors. Perceived individuals' access to data is negatively associated with the estimated likelihood of disclosing information. Those who think other people have access to their data are less likely to estimate they will disclose information in our Internet scenario. However we do not see the same effect of perceived organizations' access. **H4** (people with higher awareness of privacy risks will intend to protect their privacy in the future more than those with lower awareness of privacy risks) is supported for the perceived access of individuals, not for the perceived access of organizations.

Self-efficacy positively predicts both the estimated likelihood of disclosing information and the estimated likelihood of using protective strategies. However, the effect on the estimated likelihood of disclosing information is counter-intuitive. Probably when people are more confident of hiding their information, they are more likely to disclose. **H5** (people with higher self-efficacy will intend to protect their privacy in the future more than those with lower self-efficacy) is only supported for the estimated likelihood of using protective strategies.

As shown in the ANOVA results earlier, threat visualization has a significant negative association with the estimated likelihood of disclosing information: those who saw the depiction of threats were less likely to estimate disclosing their personal information (supports H2b). Threat visualization does not show the same effect on estimated likelihood of using protective strategies. General privacy concern negatively predicts the estimated likelihood to disclose and positively predicts the estimated likelihood to use protective strategies. Technical knowledge also positively predicts the estimated likelihood to use protective strategies. In addition, the estimated likelihood of using protective strategies negatively predicts the likelihood of disclosing information.

As has been found in previous research (Norberg, Horne & Horne, 2007), I did not find any association between participants' behavioral intentions and their actual behavior in

the path model. The intention to disclose information and the number of questions they actually disclosed had almost no correlation (r = 0.04). Because the privacy video was described as "An educational video about how to protect your privacy and security online", it should have been associated with participants' estimated likelihood of using privacy protective strategies. However, the correlation between the estimated likelihood to use protective strategies and whether or not they selected the privacy video is only 0.06. **H6** (people who intend to protect their privacy in the future will be more likely to protect their privacy) is not supported.

6.4.4 The effect of the manipulations on follow-up measures

To test whether the effect of visualizations actually educated users or was a result of experimenter demand, I sent a follow-up survey to experimental participants. The follow-up survey included the same measure of perceived access of other individuals or organizations to their data, behavioral intentions, and four disclosure questions. I also conducted two-way ANOVAs to analyze the effect of visualization manipulations on the follow up measures.

Perceived data access: The results still showed a significant main effect of threat visualization (F [1, 193] = 4.84, p < .05), and a marginally significant effect of structure visualization (F [2, 193] = 4.55, p = .05) on perceived individuals' access to data. There is also a significant main effect of threat visualization on perceived organizations' access to data (F [1, 193] = 6.06, p < .05). These findings suggest that the effect of the manipulations on participants' perceived data access by others lasted for at least one week. Those who saw visualizations with a depiction of threats had higher awareness of individuals and organizations' access to their data than those who did not see the list of threats. Those who saw an articulated Internet structure had marginally lower awareness of other individuals' access than those who saw a simple structure and those who saw no structure.

Estimated likelihood of disclosing information: I did not find a significant effect of the threat visualization on participants' likelihood of disclosing information in the second survey. However, the structure visualization showed a significant effect in the follow-up survey: those who saw an articulated or simple visualization in the first survey were more likely to estimate they would disclose information in the follow-up survey than those who saw no structure visualization in the first survey (F [2, 193] = 5.43, p < .01).

Actual disclosure: I did not find any effect of the manipulations on the number of disclosure questions participants answered in the follow-up survey.

6.4.5 Summary of the results

In sum, the analysis of variance showed that threat visualizations influenced participants' perceived access of other individuals and organizations' to their data and

their estimated likelihood of disclosing personal information. The influence on perceived data access lasted after one week, but the other effects disappeared. Strangely, participants in the articulated and simple visualization conditions had a higher estimated likelihood of disclosing personal information in the follow-up survey than those who did not see these visualizations.

The SEM analysis validated the hypothetical research model that perceived risks and perceived efficacy work together to predict behavioral intentions, but intentions were not predictive of actual behavior. It reveals that the estimated likelihood of disclosing personal information is mainly influenced by the perception of individuals' access to data, the threat visualization, general privacy concern, and self-efficacy. The estimated likelihood of using protective strategies is mainly influenced by participants' knowledge, general privacy concern, and self-efficacy. I did not observe any direct influence from perceived data access or the visualization manipulations on the estimated likelihood of using protective strategies.

6.5 Discussion

This chapter analyzes an online experiment investigating the effect of visualizing Internet structure and privacy threats on people's perceptions and behaviors. The findings of the experiment has implications for how we should educate users about the Internet, and what is effective or not effective in nudging people towards carrying out privacy protection behavior.

6.5.1 The level of transparency in educating people about privacy risks

Among the five different visualizations given to different groups in the experiments, the visualization with just a simple structure of the Internet (Figure 14) was considered the least informative. The other four visualizations (articulated viz, articulated viz with threat, simple viz with threat, threat only viz), although were very different in the amount of information conveyed, did not differ from each other in the informativeness ratings (Figure 21). This result indicates that the information of threat and the information of articulated Internet structure are probably considered as equally helpful in helping people understand the Internet. The effects of these visualizations on people's perception of privacy risks, however, were different. The information of threat was significantly more effective in increasing people's awareness of both informational and social privacy risks. The information of articulated structure, on the other hand, lowered people's perceived social risks (but this effect disappeared for the combined dataset), and did not influence their perceived informational risks. Moreover, in the follow-up survey, we saw a significant effect of structure visualization in increasing people's tendency to disclose information, suggesting that showing people the structure alone might make them less aware of the risks and be more careless in their actions. One possible reason is that seeing the complexity of the Internet made people feel more

comfortable, because they might feel protected by various technical components. These findings confirm expert reviewers' opinion in Chapter 5: the awareness of Internet entities and organizations is important knowledge that even lay people should know.

However, system designers need to be careful when telling people about the threats on the Internet, considering the potential tradeoff of "scaring people off." The new information conveyed by our visualizations might shake people's beliefs about their own knowledge. After the visualization evaluation questions, we asked in the survey: "At this time, how good is your understanding of how the Internet works?" People in four visualizations conditions (articulated viz, articulated viz with threat, simple viz with threat, threat only viz) rated their understanding significantly lower than those in the control condition (F [5, 260] = 2.45, $p < .05$). In the open-ended questions about the visualization, one respondent in the articulated threat viz condition wrote: *"It's pretty scary to see the Internet broke down like that. It makes me pretty uncomfortable about being on it."* And another respondent in the same condition noted: "*It opened my eyes… [The Internet] was more there than I thought."* This effect might result in the afore-mentioned "learned helplessness" phenomena: people might feel scared and feel less confident about what they can do, therefore give up on protecting their information. Although we find no direct association between the visualizations and our self-efficacy measures, there could be some negative effects on emotions or other aspects that we did not measure. Future work could consider measuring people's emotional responses to different visualizations and examine how to balance the purpose of educating user and the negative emotions aroused by these visualizations.

6.5.2 The influence of social and informational privacy threats

The present study revealed that the perceived individuals' access to data (a proxy of social privacy threat) significantly lowered people's intention to disclose information, but we did not see perceived organizations' access to data (a proxy of informational privacy threat) have the same effect. This finding echoes previous research that people are more likely to protect their social privacy rather than informational privacy (Young & Quan-Haase, 2013) and my study asked a more generalizable population instead of university Facebook users about their general internet use. It indicates that the concern about social privacy threats is probably a main motivator for people to restrict their own disclosure behavior. This disclosure behavior includes not only what information they post on websites, but also whether or not they conduct a web search, and whether or not they decide to visit a website.

The perceived informational privacy threats, on the other hand, lowered people's self-efficacy rating, which might result in their lower likelihood of using privacy protection strategies. This evidence of "learned-helplessness" seems to be mainly caused by the heightened concerns of informational privacy threats, rather than social privacy threats. The more people feel they can be traced by governments, hackers, or advertisers, the less

confident they feel in hiding their information online. Thus they are reluctant to adopt protective strategies probably because they think no method will work.

This finding provides important and novel insights about how social and informational privacy threats affect people's behavioral intention in protecting their privacy. An increased concern about social privacy threats could restrict people's own disclosure behavior, but an increased concern about informational privacy threats may work in the opposite direction – it could reduce people's perceived self-efficacy of hiding their information therefore lower their likelihood of adopting external tools to protect their information.

6.5.3 The lack of connection between behavioral intention and actual behavior

Unfortunately, our experiment failed to find a connection between users' behavioral intention and their actual privacy protection behavior (measured by information disclosure and tendency to learn about tools). One possible reason is that the behavioral theories, based on which this study was designed, cannot predict actual privacy behavior very well. Although these behavioral theories have been applied in various domains such as advertising, healthcare, and consumer behavior, these theories are mostly rational models that assume individuals' behavior can be mainly predicted by their motivations, or a rational calculation of risks and benefits. However, there is much evidence in privacy research showing that stated intentions, in many cases, do not predict actual behavior.

Another possible explanation is that participants might have different threats in mind when they answer the behavioral intention questions versus when they answer the actual disclosure questions. All behavioral intention questions were phrased in the Internet scenario (using Internet in a public coffee shop), but the actual disclosure questions were phrased as: "We [researchers] are interested in your general risk-taking behaviors. Please answer the following questions." Although our visualizations educated them about all the possible entities that may see their data on the Internet, the survey was built on a survey website affiliated with well-known institutions (the url started with: http://cmu.qualtrics.com/). This association might remove the threat from people's mental model and this is noted as a limitation of our study.

6.5.4 Limitations

Our visualizations might emphasize social threats more than informational threats because "anybody close by" and "others" were shown on top of the visualizations. People read from top to bottom so the first items come to their attention might have a stronger effect on their perception of threats. This was not designed by intention, but the sequence of how threats are shown might change the results. Future work should test

different version of the threat visualizations and examine if the effects on social and informational threats are the same.

Another limitation of this study is that the order of the survey questions may influence the effect of our manipulations. The first dependent measure participants answered was the disclosure intention questions. Therefore the effect of threat manipulation might be stronger on this DV rather than on the estimated likelihood of using protective strategies. The actual disclosure questions were inserted in the end of the survey, so it could also be that the order of this question removed the effect of manipulations on actual behavior.

Lastly, I did not test whether or not people's perceptions of threat are accurate. I found knowledge negatively correlated with perceived individuals' access, suggesting that people with higher technical background think fewer individuals would have access to their data. But because my experiment used a hypothetical scenario instead of asking people to do actual Internet surfing, I did not have an absolutely accurate model of who have access to people's data (actual data access may vary depending on specific computer settings). Previous research suggests that users sometimes misunderstand data access by others. For example, the majority of participants in Ur et al.'s (2012) study believed that advertisers collect personally identifiable information through the use of cookies, whereas this practice is explicitly prohibited by industry guidelines. An inaccurate perception of threats might result in unnecessary concerns and drive people's intentions toward ineffective actions.

7 Conclusion

7.1 Summary of findings

This thesis research shows that many people have tried to keep their identity or online information private using a broad range of strategies. Their motivations include social privacy (managing social boundaries in their lives) and informational privacy (keeping personal information private from hackers, companies, and government). I found people who were younger, used social media, had higher education, and who were more oriented towards segmenting different parts of their lives were more likely to control their information online. Many Internet users have limited technical knowledge, a vague understanding of the Internet, and do not know who has access to their data online and what can be done with identifying information and data about their activities online. Thus, their actions to protect their privacy may have limited effectiveness. News stories are common about people whose activities online have been unintentionally revealed. Yet even those with more knowledge of the Internet from formal education or from experience and system interfaces, who were more likely to perceive privacy risks and to feel they could handle their own privacy, did not take notably more privacy protective actions. In the following paragraphs I will review the research questions raised in Chapter 1 and discuss how my findings address these questions.

7.1.1 Why do people hide their information online?

The desire to manage boundaries among different social groups or environments is a main social reason that people hide their identity or information online (Chapter 2, 3, and 4). This finding echoes previous research that defines privacy as a boundary-regulation process (Altman, 1975; Petronio 2002) and overlaps with the empirical evidence found in studies of social media sites (Litt et al. 2014; Vitak et al. 2014; Marwick & boyd, 2011). My studies showed that this motivation not only influences how people manage their information on social media sites, but also influences their general Internet use and how they manage information across different activities and platforms (i.e., switching to the anonymous communication applications studied in Chapter 3).

Another important reason for managing social boundaries is prior negative Internet experience (Chapter 2). Having experienced privacy invasion, online harassment, or unpleasant communication with other users seems to alarm and motivate people to hide from future informational or social threats online (Chapter 4). The APCO model cited in Chapter 1 (Smith et al., 2011) predicts this effect, and research on information privacy (Culnan, 1993) describes the effect of previous invasion experiences on people's privacy concerns. Negative experiences also have been found to change how people manage their information on social media sites (Litt & Hargittai 2014). Some of our interviewees described in Chapter 5 told stories about how negative experiences influenced their behavior. A woman (C06) said she kept monitoring her bank accounts for a few months after the Target retail company's data breach happened, and did so until she was certain her accounts were not affected (perhaps prematurely stopping this surveillance, as hackers sometimes wait many months before using their stolen information). Another participant (C02), after his mother's computer was infected by virus, became the designated "technical support" person in his family.

In our interviews and the previous literature, I often heard people mentioning a reason for not acting on privacy concerns because they did not know how to do so. Nevertheless, knowing more about the Internet was not associated with taking more privacy protective actions overall, and who took action did not necessarily have more knowledge (see p.87, quote from N 11 in Chapter 5). Thus the link between knowledge and action seems weak and still remains to be discovered.

7.1.2 How do people hide their information online?

According to the sample surveys described in Chapter 4, more than half of the U.S. Internet users and an even higher percentage of the MTurk workers we surveyed have sought to hide information from at least one group of people or organizations. A significant minority of these samples took further steps to hide their identity.

According to the interviews described in Chapter 2, to protect their online information, people adopt both behavioral strategies (e.g., falsifying identities; editing previous posts) and technical strategies (e.g., clearing browser cookies; using a proxy server, Tor, or encryption). Behavioral strategies and comparatively easy-to-use technical strategies, such as deleting cookies, were used by a large majority, regardless of their technical background. More advanced technical strategies, such as using a proxy server, had a lower adoption rate, and were strongly associated with users' formal technical education or technical skills (Chapter 4, Chapter 5).

The research described in Chapter 4 suggested that the strategies involving editing content ("deleted or edited something you posted in the past"; "asked someone to remove something that was posted about you online") were mainly used to hide from social threats, whereas technical methods ("used a proxy server, Tor software, or a virtual personal network"; "encrypted your communications") were mainly used to hide

from informational threats (Figure 4). The findings of the experiment (Chapter 6) also showed that perceived social threats were more associated with people restricting their own disclosure behavior (behavioral strategies) rather than using technical tools or methods. Previous research on users' privacy protection behavior (Paine et al 2007) and security behavior (Egelman & Peer, 2015) did not distinguish behavioral strategies from technical strategies, but my results suggest there may be different antecedents motivating people to take behavioral and technical privacy protection strategies. Understanding these differences can inform future design of interfaces that aims at influencing people to adopt different protection strategies.

7.1.3 How do people understand different privacy threats?

This thesis examined people's perceptions of both informational and social privacy threats and their actions towards these threats. People seem to perceive informational threats to be far more prevalent than social threats (Figure 22, Chapter 6). This perception of differential data access might result from people's prior experience, general privacy concerns, technical knowledge, and the information they learned from interfaces. For instance, a technical participant in the Internet mental model study noted about receiving personalized advertisement by advertisers: *"They are totally telling you that they know what you're viewing, because they recommend videos for you"* (T06, Chapter 5). As shown in the results presented in Chapter 4 (Figure 5), having experienced negative events online prompted technical people to think it is not possible to be anonymous online.

Although most interviewees in the anonymity study described in Chapter 2 hid their identity with a specific threat in mind, many still suffered from a sense of uncertainty about the source of threat. Seventeen interviewees in that study expressed concern about unknown threats; although they did not know whom they were afraid of, they were hiding from uncertain threats. Some people did not know how anonymous they were when they tried to hide their identity, who they might be anonymous to, and how their activities or identities across platforms could be connected to identify them. Several interviewees in multiple studies said that nothing on the Internet is private – "The Internet never forgets" (p.23, Chapter 2); "I think everyone has access" (p.79, Chapter 5). Surprisingly, others, including most users of the anonymous communication applications we talked to (Chapter 3) were confident about their anonymity. Perhaps they were mainly thinking about social threats, and the absence of usernames on these applications felt anonymous, even if they were not fully anonymous.

7.1.4 How do people's understanding of privacy threats affect their decisions to hide their information online?

My findings suggest that whether or not people change their own behavior (e.g., not visit a website, or edit the content they posted online) and whether or not people adopt

tools (e.g., use proxy, Tor) might be caused by different concerns. Changing behavior online and using anonymous communication apps seems to be motivated by social privacy concerns whereas using more advanced technical tools such as Tor, seems to be motivated by informational privacy concerns.

The effects of technical knowledge are not straightforward. Although people with more technical knowledge are more likely to use advanced tools, results from the experiment (Chapter 6) showed that, after seeing a list of privacy threats, participants were more aware of being tracked and followed by institutions or companies (informational threats), but these concerns did not result in more privacy protective actions. Perceived social threats increased people's intentions to disclose less, but this intention did not translate into their actual behavior either.

The mismatch between people's behavioral intentions and actual behavior has been documented in previous literature (Acquisti et al 2004; Norberg & Horne 2007). The different effect of social and informational threats on people's behavior, however, has not been studied before. This distinction is important in informing the design of future systems and tools that aim to improve users' own privacy protection actions, and future research should further examine this difference.

7.2 Future work

The findings of this thesis suggest implications for future research, design and policies.

7.2.1 Examine other aspects of knowledge

Our studies specifically examined participants' knowledge of the underpinnings of the Internet, how they tried to control data access by others. It is shown that knowledge of the technical structure of the Internet may not be helpful in guiding more secure behavior, but a simple list of threats had more effect on people's behavioral intention to protect their information. We did not measure participants' knowledge of how attacks occur or how they understood different privacy and security risks in detail. There is much more to learn about these and other dimensions of Internet knowledge. Future work could use the same approach as in the mental model study to examine people's understanding of privacy or security attacks. This will also help us understand how negative Internet experiences influence people's perception and behavior. We also need to understand better how people understand security versus privacy—or whether they even need such a distinction.

7.2.2 Develop better measures of actual privacy protection behavior

Prior research has used various methods to measure people's privacy protection behavior such as the amount of personal information disclosure (Brandimarte et al., 2013; Norberg et al., 2007), online purchase decisions (Tsai, Egelman, Cranor, & Acquisti,

2011), whether or not users adjust privacy settings (Almuhimedi et al., 2015; Knijnenburg et al., 2013), and whether or not users edit content they posted (Wang et al., 2014). I adapted the measure of personal information disclosure from previous research and created a measure to capture people's tendency to learn about protection methods. However, none of these measures showed an effect in my experiment. It is hard to explain whether it's because the manipulations did not change people's actual behavior, or because the measures did not work. Creating a realistic measure of people's privacy protection behavior in an experimental setting is difficult, because participants tend to trust the researcher more than other entities they encounter in daily use of the Internet. Therefore, my thesis argues for more research efforts on testing and developing accurate and generalizable measures of users' actual privacy protection behavior.

7.2.3 Implications for design

People rely on observable cues to understand how their information is accessed, used, or protected online. These cues include interface cues (e.g., lock sign, dots replacing password), dynamic information (e.g., tailored advertisements), and social information (e.g., comments on a post). Most of our participants were aware of personalized services or advertisements, which spoke to their high awareness of informational privacy threats. Social cues on sites like YouTube and Facebook (e.g., user profiles, number of views, and uploader's profile) indicated the presence of other users, which rendered participants' activity on those sites more public. Regardless of their technical knowledge, participants seem to have made most of their privacy-related decisions based on these experiences and cues.

Most observable cues inform users about their privacy and security in the application layer and mainly deal with informational threats. Other limited cues educate users about social threats from other people, such as supervisors, or security risks at other layers of the network. However, it can be easy to miss these limited cues. One design implication is to provide a "privacy indicator" for people's Internet activities, showing them who can see what information. Bernstein et al. (2015) proposed that visualizing the size of one's audience on social media would help users understand the exposure of one's posts. Visualizing one's audience across applications and different network layers might help to increase users' awareness of privacy and security risks. At a minimum, applications could inform users about what control they have over their data, if any, once they put it online. Data access was the most important aspect of privacy emphasized by expert reviewers, but it was also the most difficult for participants to grasp. The challenge, as one expert reviewer in Chapter 5 noted, is in which data or security risk to surface or prioritize for user attention.

7.2.4 Implications for technology and policy

Security threat models in previous research emphasize the methods attackers use to exploit weaknesses in existing systems. From a user perspective, we emphasize how

variations in individual background and interface cues shape users' perceptions and their actions to mitigate different threats. Current attempts to build network-level security can be informed by these behavioral models. For example, our studies showed that people with different characteristics care about privacy for some sessions in which they use the Internet and not others. Their concerns are diverse (e.g., cross platform or not). For any Internet communication, a threat could be located at the source, destination, or different points in the end-to-end path. Internet users do not have a clear picture of who gets access to information at each point as a result of their specific actions. Although people have learned how to use application-layer tactics, they do not know how these methods work and what threats these methods address and do not address. For instance, deleting cookies may prevent third parties from accessing one's web history but do not prevent authorities from obtaining your data from the companies you visit. Having network-level mechanisms to deal with various threats might help solve the problem if these mechanisms were designed based on what users do and choices they make (e.g., to encrypt or not, to "go public" or not).

Although the expert reviewers and a few technical participants in Chapter 5 suggested that users should take more responsibility of their own privacy rather than putting too much trust in the system or the software, the findings of my thesis indicate, even when people have higher knowledge of the Internet and know what tools they can use, they still don't take more actual actions to protect their information. Our work suggests a need for more research into privacy protections (in the aspects of both technology and policy) that reduce the responsibility on users to make myriads of privacy protection decisions on their own (Holdren et al., 2015).

References

Ackerman, M. S., Cranor, L. F., & Reagle, J. (1999). Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM conference on Electronic commerce* (pp. 1–8).

Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce* (pp. 21–29). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=988777

Acquisti, A., Brandimarte, L., & George. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514.

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Privacy enhancing technologies* (pp. 36–58).

Ajzen, I. (1985). *From intentions to actions: A theory of planned behavior*. Springer.

Ajzen, I. (1991). The theory of planned behavior. *Orgnizational Behavior and Human Decision Processes*, *50*, 179–211. doi:10.1016/0749-5978(91)90020-T

Albrechtsen, E. (2007). A qualitative study of users' view on information security. *Computers & Security*, *26*(4), 276–289.

Almuhimedi, H., Schaub, F., Sadeh, N., Adjerid, I., Acquisti, A., Gluck, J., … Agarwal, Y. (2015). Your Location has been Shared 5398 Times ! A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 787–796). doi:10.1145/2702123.2702210

Altman, I. (1975). *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding.* Monterey, California: Brooks/Cole Publishing Company.

Angulo, J., Erik, W., & Johan, H. (2014). What Would It Take for You to Tell Your Secrets to a Cloud? Studying Decision Factors When Disclosing Information to Cloud Services. *Secure IT Systems*, 129–145. doi:10.1007/978-3-642-34210-3

Ashforth, B. E., Kreiner, G. E., & Fugate, M. (2000). All in a day's work: Boundaries and micro role transitions. *Academy of Management Review*, *25*(3), 472–491. doi:10.5465/AMR.2000.3363315

Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, *54*(15), 2787–2805. doi:10.1016/j.comnet.2010.05.010

Balebako, R., Jung, J., Lu, W., Cranor, L. F., & Nguyen, C. (2013). "Little brothers watching you": raising awareness of data leaks on smartphones. *Proceedings of the Ninth Symposium on Usable Privacy and Security - SOUPS '13*. doi:10.1145/2501604.2501616

Bargh, J. A., McKenna, K. Y. A., & Fitzsimons, G. M. (2002). Can You See the Real Me ? Activation and Expression of the " True Self " on the Internet. *Journal of Social Issues*, *58*(1), 33–48. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1540-4560.00247/full

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin*, *91*(1), 3.

Bentler, P. M. (1989). *EQS structural equations program manual*. Los Angeles: Multivariate Software.

Berendt, B. (2005). Privacy in e-commerce: Stated preferences vs. actual behavior. *Communications of the ACM*, *48*(4), 101–106. Retrieved from http://dl.acm.org/citation.cfm?id=1053295

Bernstein, M., Bakshy, E., Burke, M., & Karrer, B. (2013). Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 21–30).

Bernstein, M. S., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., & Vargas, G. (2011). 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *5th International Conference on Weblogs and Social Media (ICWSM), Barcelona, Spain*.

Biddle, R., Patrick, A. S., & Sobey, J. (2009). Browser Interfaces and Extended Validation SSL Certificates : An Empirical Study Categories and Subject Descriptors. In *CCSW '09*. doi:http://doi.acm.org/10.1145/1655008.1655012

Birnholtz, J., Aaron, N., Merola, R., & Paul, A. (2015). " Is it Weird to Still Be a Virgin ?:" Anonymous , Locally Targeted Questions on Facebook Confession Boards.

boyd, b, & Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday*, *15*(8). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086/2589

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2013). Misplaced confidences privacy and the control paradox. *Social Psychological and Personality Science*, *4*(3), 340–347.

Bravo-Lillo, C., Cranor, L. F., Downs, J., & Komanduri, S. (2011). Bridging the gap in computer security warnings: A mental model approach. *IEEE Security and Privacy*, *9*(2), 18–26. doi:10.1109/MSP.2010.198

Brewer, M. B., & Chen, Y.-R. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological Review*, *114*(1), 133–151. doi:10.1037/0033-295X.114.1.133

Brown, A., Mortier, R., & Rodden, T. (2013). MultiNet: reducing interaction overhead in domestic wireless networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1569–1578).

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage Publications.

Chen, K., & Rea, A. (2004). Protecting personal information online: A survey of user privacy concerns and control techniques. *Journal of Computer Information Systems*, *44*(4), 85–92. Retrieved from ftp://140.130.109.23/upload/CFIP/References/2004/JCIS-2004-44-04-085-Protecting Personal Information Online A Survey of User Privacy Concerns and .pdf

Christopherson, K. M. (2007). The positive and negative implications of anonymity in Internet social interactions. *Computers in Human Behavior*, *23*(6), 3038–3056.

Coleman, E. G., & Golub, A. (2008). Hacker practice. *Anthropological Theory*, *8*(3), 255–277.

Constine, J. (2015). Secret Shuts Down. Retrieved from http://techcrunch.com/2015/04/29/psst/

Conti, G., & Sobiesk, E. (2007). An honest man has nothing to fear: user perceptions on web-based information disclosure. In *SOUPS* (Vol. 7, pp. 112–121).

Corbin, J. M., & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications, Inc.

Correa, D., Silva, L. A., Mondal, M., Benevenuto, F., & Gummadi, K. P. (2015). The Many Shades of Anonymity : Characterizing Anonymous Social Media Content. In *ICWSM* (pp. 71–80). AAAI.

Culnan, M. (1993). " How Did They Get My Name ?": An Exploratory Investigation of Consumer Attitudes Toward Secondary Information Use. *Mis Quarterly*, *17*(3), 341–363. Retrieved from http://dl.acm.org/citation.cfm?id=2017108

Daley, E. (2014). Trial of Pittsburgh researcher accused of poisoning wife goes to jury. *Reuters*. Retrieved from http://www.reuters.com/article/2014/11/06/us-usa-pennsylvania-cyanide-idUSKBN0IQ2MY20141106

Das, S., Kim, T., Dabbish, L., & Hong, J. (2014). The Effect of Social Influence on Security Sensitivity. In *SOUPS 2014* (pp. 143–157). USENIX. Retrieved from http://cmuchimps.org/uploads/publication/paper/147/the_effect_of_social_influence_on_security_sensitivity.pdf

Das, S., & Kramer, A. (2013). Self-Censorship on Facebook. In *ICWSM* (pp. 120–127). AAAI.

Derlega, V. J., Grzelak, J., & others. (1979). Appropriateness of self-disclosure. *Self-Disclosure: Origins, Patterns, and Implications of Openness in Interpersonal Relationships*, 151–176.

Dimicco, J. M., & Millen, D. R. (2007). Identity Management : Multiple Presentations of Self in Facebook. In *Group'07* (pp. 383–386). ACM. doi:10.1145/1316624.1316682

Dinev, T., Hart, P., & Mullen, M. R. (2008). Internet privacy concerns and beliefs about government surveillance--An empirical investigation. *The Journal of Strategic Information Systems*, *17*(3), 214–233.

Dommeyer, C. J., & Gross, B. L. (2003). What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy proteciton strategies. *Journal of Interactive Marketing*, *17*(2), 34–51. doi:10.1002/dir.10053

Donath, J. S. (1999). Identity and deception in the virtual community. *Communities in Cyberspace*, *1996*(London), 29–59. Retrieved from http://smg.media.mit.edu/people/Judith/Identity/IdentityDeception.html

Dryden, A. (2014). Social Networking as Peer Surveillance. Retrieved from https://modelviewculture.com/pieces/social-networking-as-peer-surveillance

Egelman, S., & Peer, E. (2015). Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *CHI '15* (pp. 2873–2882). Seoul, Republic of Korea: ACM.

Ellison, N., Heino, R., & Gibbs, J. (2006). Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment. *Journal of Computer-Mediated Communication*, *11*(2), 415–441.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215.

Farnham, S. D., & Churchill, E. F. (2011). Faceted identity, faceted lives: social and technical issues with being yourself online. In *CSCW 2011* (pp. 359–368). ACM Press. doi:10.1145/1958824.1958880

Farrall, K. (2012). Online collectivism, individualism and anonymity in East Asia. *Surveillance and Society*, *9*(4), 424–440.

Friedman, B., Hurley, D., Howe, D. C., Nissenbaum, H., & Felten, E. (2002). Users' conceptions of risks and harms on the web: A comparative study. *CHI '02 Extended Abstracts on Human Factors in Computing Systems - CHI '02*, 614. doi:10.1145/506443.506510

Furnell, S. M., Bryant, P., & Phippen, a. D. (2007). Assessing the security perceptions of personal Internet users. *Computers and Security*, *26*(5), 410–417. doi:10.1016/j.cose.2007.03.001

Gino, F., Sharek, Z., & Moore, D. a. (2011). Keeping the illusion of control under control: Ceilings, floors, and imperfect calibration. *Organizational Behavior and Human Decision Processes*, *114*(2), 104–114. doi:10.1016/j.obhdp.2010.10.002

Goffman, E. (1959). The presentation of self in everyday life. Garden City, NY: Doubleday Anchor Books.

Greenberger, D. B., Miceli, M. P., & Cohen, D. J. (1987). Oppositionists and group norms: The reciprocal influence of whistle-blowers and co-workers. *Journal of Business Ethics*, *6*(7), 527–542.

Greenwald, G. (2013). NSA collecting phone records of millions of Verizon customers daily. Retrieved from http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order

Gross, E. F. (2004). Adolescent internet use: What we expect, what teens report. *Journal of Applied Developmental Psychology*, *25*(6 SPEC. ISS.), 633–649. doi:10.1016/j.appdev.2004.09.005

Hargittai, E., & Litt, E. (2013). New Strategies for Employment ? Internet skills and online privacy practices during people's job search, *11*.

Harris, P. (1996). Sufficient grounds for optimism? The relationship between perceived controllability and optimistic bias, *15*(1), 9–52. doi:10.1521/jscp.1996.15.1.9

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, *28*(1), 127–138. doi:10.1016/S0364-0213(03)00065-X

Hofstede, G. (1983). Dimensions of national cultures in fifty countries and three regions. In J.Deregowski, S. Dzuirawiec, & R. Annis (Eds.), *Explications in Cross-Cultural Psychology*.

Holdren, J., Lander, E., Press, W., Savitz, M., Austin, W., Chyba, C., & ... (2015). *Report to the President and Congress Ensuring Leadership in Federally Funded Research and Development in Information Technology*. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/nitrd_report_aug_2015.pdf

Iachello, G., & Hong, J. (2007). End-User Privacy in Human-Computer Interaction. *Foundations and Trends in Human-Computer Interaction*, *1*(1), 1–137. doi:10.1561/1100000004

Jacobson, M. J. (2001). Problem solving, cognition, and complex systems: Differences between experts and novices. *Complexity*, *6*(3), 41–49.

Jensen, C., & Potts, C. (2005). Privacy practices of Internet users: self-reports versus observed behavior. *International Journal of Human-Computer Studies*, *63*(1), 203–227. doi:10.1016/j.ijhcs.2005.04.019

John, L. K., Acquisti, A., & Loewenstein, G. (2011). Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information. *The Journal of Consumer Research*, *37*(5), 858–873. doi:10.1086/656423

Johnston, A. C., & Warkentin, M. (2010). Fear Appeals and Information Security Behaviors: An Empirical Study. *MIS Quarterly*, *34*(3), 549–A4. doi:Article

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, *31*(2), 177–192. doi:10.1002/ejsp.36

Joinson, A. N., Reips, U.-D., Buchanan, T., Schofield, P., & Carina, B. (2010). Privacy, trust, and self-disclosure online. *Human-Computer Interaction*, *25*, 1–24. doi:10.1080/07370020903586662

Jonassen, D., & Cho, Y. H. (2008). Externalizing Mental Models with Mindtools. In *Understanding Models for Learning and Instruction* (pp. 145–159). Springer.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, 193–206.

Kang, R., Brown, S., & Kiesler, S. (2013). Why Do People Seek Anonymity on the Internet?: Informing Policy and Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2657–2666). New York, NY, USA: ACM. doi:10.1145/2470654.2481368

Kang, R., Dabbish, L., Fruchter, N., & Kiesler, S. (2015). " My Data Just Goes Everywhere :" User Mental Models of the Internet and Implications for Privacy and Security. In *SOUPS 2015* (pp. 39–52). USENIX Association.

Kang, R., Dabbish, L., Kiesler, S., & Sutton, K. (2016). Strangers on your phone: Why people use anonymous communication applications. In *CSCW '16*. ACM.

Kelley, P. (2010). Conducting usable privacy & security studies with amazon's mechanical turk. In *Symposium on Usable Privacy and Security (SOUPS)*. Redmond, WA. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.208.1736&rep=rep1&type=pdf

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*(10), 1123–1134. doi:10.1037/0003-066X.39.10.1123

Kim, T. H.-J., Stuart, H. C., Hsiao, H.-C., Lin, Y.-H., Zhang, L., Dabbish, L., & Kiesler, S. (2014). YourPassword: applying feedback loops to improve security behavior of managing multiple passwords. In *Proceedings of the 9th ACM symposium on Information, computer and communications security* (pp. 513–518).

Klasnja, Predrag, E. A. (2009). " When I am on Wi-Fi , I am Fearless :" Privacy Concerns & Practices in Everyday Wi-Fi Use. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM.*, 1993–2002. doi:10.1145/1518701.1519004

Knijnenburg, B. P., Kobsa, A., & Jin, H. (2013). Preference-based Location Sharing: Are More Privacy Options Really Better? In *CHI '13* (pp. 2667–2676).

Kowitz, B., & Cranor, L. (2005). Peripheral privacy notifications for wireless networks. In *ACM workshop on Privacy in the electronic* (pp. 90–96). doi:10.1145/1102199.1102217

Kraut, R. E., & Resnick, P. (2011). *Building Successful Online Communities: Evidence-Based Social Design*. Cambridge, Massachusetts London, England: The MIT Press.

Krishnamurthy, B., & Wills, C. E. (2008). Characterizing privacy in online social networks. In *Proceedings of the first workshop on Online social networks* (pp. 37–42).

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 905–914).

Lampinen, A., Lehtinen, V., Lehmuskallio, A., & Tamminen, S. (2011). We're in It Together: Interpersonal Management of Disclosure in Social Network Services. In *CHI 2011* (pp. 3217–3226). ACM. doi:10.1145/1978942.1979420

LaRose, R., Rifon, N. J., & Enbody, R. (2008). Promoting personal responsibility for internet safety. *Communications of the ACM*, *51*(3), 71–76. doi:10.1145/1325555.1325569

Leavitt, A. (2015). " This is a Throwaway Account ": Temporary Technical Identities and Perceptions of Anonymity in a Massive Online Community.

Leon, P. G., Ur, B., Wang, Y., Sleeper, M., Balebako, R., Shay, R., … Cranor, L. F. (2013). What matters to users?: factors that affect users' willingness to share information with online advertisers. *Proceedings of the Ninth Symposium on Usable Privacy and Security - SOUPS '13*. doi:10.1145/2501604.2501611

Lerner, J. S., Gonzalez, R. M., Small, D. a., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science*, *14*(2), 144–150. doi:10.1111/1467-9280.01433

Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., & Zhang, J. (2012). Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy Through Crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 501–510). New York, NY, USA: ACM. doi:10.1145/2370216.2370290

Litt, E. (2013). Measuring users' internet skills: A review of past assessments and a look toward the future. *New Media & Society*, *15*(4), 612–630. doi:10.1177/1461444813475424

Litt, E., Spottswood, E., Birnholtz, J., Hancock, J., Smith, M. E., & Reynolds, L. (2014). Awkward Encounters of an "Other" Kind: Collective Self-Presentation and Face Threat on Facebook. *ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*, 449–460. doi:10.1145/2531602.2531646

Madden, M. (2014). *Public Perceptions of Privacy and Security in the Post-Snowden Era*. Retrieved from http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/

Madden, M., & Smith, A. (2010). *Reputation management and social media*. Retrieved from http://www.pewinternet.org/Reports/2010/Reputation-Management.aspx.

Madden, M., Fox, S., Smith, A. (2007). *Digital footprints*. Retrieved from http://www.pewinternet.org/Reports/2007/Digital-Footprints.aspx

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, *15*(4), 336–355.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. doi:10.1037/0033-295X.98.2.224

Marwick, A. E., & danah boyd. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133.

Marx, G. T. (1999). What's in a Name ? Some Reflections on the Sociology of Anonymity. *The Information Society*, *15*(2), 99–112. doi:10.1080/019722499128565

Mazurek, M. L., Arsenault, J. P., Bresee, J., Gupta, N., Ion, I., Johns, C., … Reiter, M. K. (2010). Access Control for Home Data Sharing: Attitudes, Needs and Practices. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 645–654). New York, NY, USA: ACM. doi:10.1145/1753326.1753421

McKenna, K. Y. A., & Bargh, J. A. (2000). Plan 9 From Cyberspace: The Implications of the Internet for Personality and Social Psychology. *Personality and Social Psychology Review*, *4*(1), 57–75. doi:10.1207/S15327957PSPR0401_6

Mohamed, N., & Ahmad, I. H. (2012). Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior*, *28*(6), 2366–2375. doi:10.1016/j.chb.2012.07.008

Morris, M. R., Inkpen, K., & Venolia, G. (2014). Remote Shopping Advice : Enhancing In-Store Shopping with Social Technologies. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, 662–673. doi:10.1145/2531602.2531707

Nguyen, D. H., Kobsa, A., & Hayes, G. R. (2008). An empirical investigation of concerns of everyday tracking and recording technologies. *Proceedings of the 10th International Conference on Ubiquitous Computing - UbiComp '08*, 182. doi:10.1145/1409635.1409661

Norberg, P. a, Horne, D. R., & Horne, D. (2007). The Privacy Paradox : Personal Information Disclosure Intentions vers us Behaviors, *41*(I), 100–126. doi:10.1111/j.1083-6101.2009.01494.x

Norman, D. A. (1988). *The psychology of everyday things*. Basic books.

Nunamaker, J. F., Applegate, L. M., & Konsynski, B. R. (1988). Computer-Aided Deliberation: Model Management and Group Decision Support: Special Focus Article. *Operations Research*, *36*(6), 826–848.

Odom, W., & Sellen, A. (2012). Lost in translation: Understanding the possession of digital things in the cloud. In *CHI'12: Proceedings of the …*. Retrieved from http://willodom.com/publications/paper1202-odom.pdf

Omarzu, J. (2000). A Disclosure Decision Model: Determining How and When Individuals Will Self-Disclose. *Personality and Social Psychology Review*, *4*(2), 174–185. doi:10.1207/S15327957PSPR0402_05

Page, K., & Uncles, M. (2004). Consumer knowledge of the World Wide Web: Conceptualization and measurement. *Psychology and Marketing*, *21*(8), 573–591. doi:10.1002/mar.20023

Paine, C., Reips, U.-D., Stieger, S., Joinson, A., & Buchanan, T. (2007). Internet users' perceptions of "privacy concerns" and "privacy actions." *International Journal of Human-Computer Studies*, *65*(6), 526–536. doi:10.1016/j.ijhcs.2006.12.001

Palen, L., & Dourish, P. (2003). Unpacking privacy for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 129–136).

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1626226

Park, Y. J. (2011). Digital Literacy and Privacy Behavior Online. *Communication Research*. doi:10.1177/0093650211418338

Peddinti, S. T., Korolova, A., Bursztein, E., & Sampemane, G. (2014). Cloak and Swagger: Understanding Data Sensitivity Through the Lens of User Anonymity. *S&P'14*, 493–508. doi:10.1109/SP.2014.38

Pedersen, D. M. (1982). Personality correlates of privacy. *The Journal of Psychology*, *112*(1), 11–14.

Pennebaker, J. W., Kiecolt-Glaser, J. K., & Glaser, R. (1988). Disclosure of traumas and immune function: health implications for psychotherapy. *Journal of Consulting and Clinical Psychology*, *56*(2), 239.

Petronio, S. (2002). *Boundaries of privacy*. *State University of New York Press, Albany, NY*.

Pew Internet Project. (2010). Reputation management and social media: Our digital footprints. Retrieved from http://www.pewinternet.org/Infographics/2010/Reputation-Management.aspx

Pew Research Center. (2013). Majority Views NSA Phone Tracking as Acceptable Anti-terror Tactic. Retrieved from http://www.people-press.org/2013/06/10/majority-views-nsa-phone-tracking-as-acceptable-anti-terror-tactic/

Pew Research Center. (2014). What Internet Users Know about Technology and the Web. Retrieved from http://www.pewinternet.org/2014/11/25/web-iq/

Poole, E. S., Chetty, M., Grinter, R. E., & Edwards, W. K. (2008). More than meets the eye: transforming the user experience of home network management. In *Proceedings of the 7th ACM conference on Designing interactive systems* (pp. 455–464).

Potosky, D. (2007). The Internet knowledge (iKnow) measure. *Computers in Human Behavior*, *23*(6), 2760–2777. doi:10.1016/j.chb.2006.05.003

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, *20*(2), 201–223.

Qian, H., & Scott, C. R. (2007). Anonymity and self-disclosure on weblogs. *Journal of Computer-Mediated Communication*, *12*(4), 1428–1451. doi:10.1111/j.1083-6101.2007.00380.x

Rader, E. (2014). Awareness of behavioral tracking and information privacy concern in facebook and google. In *Proc. of Symposium on Usable Privacy and Security (SOUPS), Menlo Park, CA, USA*.

Rainie, L., Kiesler, S., Kang, R., & Madden, M. (2013). Anonymity, Privacy, and Security Online. *Pew Research Center*, 35. Retrieved from http://www.pewinternet.org/Reports/2013/Anonymity-online.aspx

Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., & Dabbish, L. (2013). *Anonymity, privacy, and security online. Pew Research Center*.

Raja, F., Hawkey, K., & Beznosov, K. (2009). Revealing Hidden Context: Improving Mental Models of Personal Firewall Users. In *SOUPS 2009*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Revealing+Hidden+Context:+Improving+Mental+Models+of+Personal+Firewall+Users#0

Raynes-Goldie, K. (2010). Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday*, *15*(1). doi:10.5210/fm.v15i1.2775

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, *28*(3), 377–408.

Resnick, M., & Wilensky, U. (1998). Diving Into Complexity: Developing Probabilistic Decentralized Thinking Through Role-Playing Activities. *Journal of the Learning Sciences*, *7*(2), 153–172. doi:10.1207/s15327809jls0702_1

Rhee, H. S., Kim, C., & Ryu, Y. U. (2009). Self-efficacy in information security: Its influence on end users' information security practice behavior. *Computers and Security*, *28*(8), 816–826. doi:10.1016/j.cose.2009.05.008

Rubin, Z. (1975). Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology*, *11*(3), 233–260. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022103175800254

Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., & Rao, J. (2009). Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing*, *13*(6), 401–412.

Schechter, S. E., Dhamija, R., Ozment, A., & Fischer, I. (2007). The Emperor's New Security Indicators An evaluation of website authentication and the effect of role playing on usability studies. In *IEEE Symposium on Security and Privacy*.

Seligman, M. E. P. (1972). Learned helplessness. *Annual Review of Medicine*, *23*(1), 407–412.

Sen, S., Joe-Wong, C., Ha, S., & Chiang, M. (2013). A survey of smart data pricing: Past proposals, current plans, and future trends. *ACM Computing Surveys (CSUR)*, *46*(2), 15.

Señor, I. C., Fernández-Alemán, J. L., & Toval, A. (2012). Are Personal Health Records Safe? A Review of Free Web-Accessible Personal Health Record Privacy Policies. *Journal of Medical Internet Research*, *14*(4), e114.

Shay, R., Ion, I., Reeder, R. W., & Consolvo, S. (2014). My religious aunt asked why i was trying to sell her viagra: experiences with account hijacking. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 2657–2666).

Shklovski, I., & Kotamraju, N. (2011). Online contribution practices in countries that engage in internet blocking and censorship. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11* (p. 1109). New York, New York, USA: ACM Press. doi:10.1145/1978942.1979108

Shklovski, I., Mainwaring, S. D., Skúladóttir, H. H., & Borgthorsson, H. (2014). Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 2347–2356).

Sibona, C., & Walczak, S. (2011). Unfriending on facebook: Friend request and online/offline behavior analysis. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 1–10. doi:10.1109/HICSS.2011.467

Sleeper, M., Balebako, R., Das, S., McConahy, A. L., Wiese, J., & Cranor, L. F. (2013). The post that wasn't: exploring self-censorship on facebook. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 793. doi:10.1145/2441776.2441865

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS Quarterly*, *35*(4), 989–1015. doi:10.1126/science.1103618

Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, *51*(1), 125–139. doi:10.1037/0022-3514.51.1.125

Software engineer hooked on child porn jailed for three years. (2008). *The Sligo Champion*. Retrieved from

http://www.independent.ie/regionals/sligochampion/news/software-engineer-hooked-on-child-porn-jailed-for-three-years-27558635.html

Solove, D. J. (2007). "I"ve Got Nothing to Hide' and Other Misunderstandings of Privacy. *San Diego Law Review*, *44*(May), 1–23. doi:10.2139/ssrn.998565

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior. *EC '01 Third ACM Conference on Electronic Commerce*, 38–47. doi:10.1145/501158.501163

Stern, S. (2007). Producing Sites, Exploring Identities: Youth Online Authorship. In D. Buckingham (Ed.), *Youth, Identity, and Digital Media* (Vol. -, pp. 95–117). Cambridge, MA: The MIT Press. doi:i: 10.1162/dmal.9780262524834.095</p>

Stuart, H. C., Dabbish, L., Kiesler, S., Kinnaird, P., & Kang, R. (2012). Social Transparency in Networked Information Exchange : A Framework and Research Question. *Network Computing*, 451–460. doi:10.1145/2145204.2145275

Stutzman, F., Capra, R., & Thompson, J. (2011). Factors mediating disclosure in social network sites. *Computers in Human Behavior*, *27*(1), 590–598. doi:10.1016/j.chb.2010.10.017

Stutzman, F., Gross, R., & Acquisti, A. (2013). Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. *Journal of Privacy and …*, (2), 7–41. Retrieved from http://repository.cmu.edu/jpc/vol4/iss2/2/

Stutzman, F., & Hartzog, W. (2012). Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 769–778). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1566904

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *7*(3), 321–6. doi:10.1089/1094931041291295

Suler, J. R. (2002). Identity management in cyberspace. *Journal of Applied Psychoanalytic Studies*, *4*(4), 455–459.

Tbahriti, S.-E., Ghedira, C., Medjahed, B., & Mrissa, M. (2014). Privacy-Enhanced Web Service Composition. *Services Computing, IEEE Transactions on*, *7*(2), 210–222.

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, *96*(3), 506–520. doi:10.1037/0033-295X.96.3.506

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, *22*(2), 254–268. doi:10.1287/isre.1090.0260

Turkle, S. (1995). Life on the Screen. Identity in the Age of the Age of Internet. *Touchstone, NY, NY, USA*.

Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. Basic books.

Turner, E., & Dasgupta, S. (2003). Privacy on the Web: An examination of user concerns, technology, and implications for business organizations and individuals. *Information Systems Management*, 8–19. Retrieved from http://www.tandfonline.com/doi/abs/10.1201/1078/43203.20.1.20031201/40079.2

Turow, J. (2003). Americans and Online Privacy - The System is Broken, 1 – 36.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Twenge, J. M., Campbell, W. K., & Carter, N. T. (2014). Declines in Trust in Others and Confidence in Institutions Among American Adults and Late Adolescents, 1972-2012. *Psychological Science*. doi:10.1177/0956797614545133

Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y. (2012). Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security - SOUPS '12* (p. 15 pages). doi:10.1145/2335356.2335362

Vance, A., Eargle, D., Ouimet, K., & Straub, D. (2013). Enhancing password security through interactive fear appeals: A web-based field experiment. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2988–2997. doi:10.1109/HICSS.2013.196

Vaniea, K. E., Rader, E., & Wash, R. (2014). Betrayed by updates: How negative experiences affect future security. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*, 2671–2674. doi:10.1145/2556288.2557275

Vitak, J., & Kim, J. (2014). "You can't block people offline": examining how facebook's affordances shape the disclosure process. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, 461–474. doi:10.1145/2531602.2531672

Walker, K. (2000). "It's Difficult to Hide It": The Presentation of Self on Internet Home Pages. *Qualitative Sociology*, *23*(1), 99–120.

Wang, G., Wang, B., Wang, T., Nika, A., Zheng, H., & Zhao, B. Y. (2014). Whispers in the Dark : Analysis of an Anonymous Social Network Categories and Subject Descriptors. In *IMC '14* (pp. 137–150). ACM.

Wang, Y., Leon, P. G., Acquisti, A., Cranor, L. F., Forget, A., & Sadeh, N. (2014). A Field Trial of Privacy Nudges for Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2367–2376). New York, NY, USA: ACM. doi:10.1145/2556288.2557413

Wang, Y., Norice, G., & Cranor, L. (2011). Who Is Concerned about What ? A Study of American , Chinese and Indian Users ' Privacy Concerns on Social Network Sites. *Trust and Trustworthy Computing*. Retrieved from http://www.springerlink.com/index/LU75324R42234601.pdf

Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, 193–220.

Wash, R. (2010). Folk models of home computer security. *Proceedings of the Sixth Symposium on Usable Privacy and Security - SOUPS '10*, 1–16. doi:10.1145/1837110.1837125

Weinstein, N. D. (1989). Optimistic Biases about Personal Risks. *Science*, *246*(4935), 1232–1233.

Weise, E. (2014). JP Morgan reveals data breach affected 76 million households. Retrieved from http://www.usatoday.com/story/tech/2014/10/02/jp-morgan-security-breach/16590689/

Wellman, B., & Gulia, M. (1998). Net surfers don't ride alone: Virtual communities as communities. *Communities in Cyberspace*, 167–193. doi:10.2307/2655574

Westin, A. F. (1967). *Privacy and freedom*. Atheneum New York.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, *90*(2), 245.

Wisniewski, P., Lipford, H., & Wilson, D. (2012). Fighting for My Space: Coping Mechanisms for SNS Boundary Regulation. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12*, 609. doi:10.1145/2207676.2207761

Witte, K. (1994). Fear control and danger control: A test of the extended parallel process model (EPPM). *Communication Monographs*, *61*(2), 113–134. doi:10.1080/03637759409376328

Woodruff, A. (2014). Necessary , Unpleasant , and Disempowering : Reputation Management in the Internet Age, 149–158.

Woodruff, A., Pihur, V., & Consolvo, S. (2014). Would a privacy fundamentalist sell their DNA for $ 1000 ... if nothing bad happened as a result ? The Westin categories , behavioral intentions , and consequences. In *Symposium on Usable Privacy and Security* (pp. 1–18). Retrieved from https://www.usenix.org/system/files/conference/soups2014/soups14-paper-woodruff.pdf

Young, A. L., & Quan-Haase, A. (2013). Privacy Protection Strategies on Facebook. *Information, Communication and Society*, *16*(4), 1–22. doi:10.1080/1369118X.2013.777757

Yurchisin, J., Watchravesringkan, K., & Mccabe, D. B. (2005). AN EXPLORATION OF IDENTITY RE-CREATION IN THE CONTEXT OF INTERNET DATING. *Social Behavior and Personality: An International Journal*, *33*(8), 735–750. doi:doi:10.2224/sbp.2005.33.8.735

Zhang, X. (2005). What do consumers really know about spyware? *Communications of the ACM*, *48*(8), 44–48.

Zhao, X., Salehi, N., Naranjit, S., Alwaalan, S., Voida, S., & Cosley, D. (2013). The many faces of Facebook: Experiencing social media as performance, exhibition, and personal archive. In *CHI '13* (pp. 1–10). ACM. doi:10.1145/2470654.2470656

Zheng, J., Simplot-Ryl, D., Bisdikian, C., & Mouftah, H. (2011). The Internet of Things. *IEEE Communications Magazine*, 30–31.

Zwerdling, D. (2013). Easily obtained subpoenas turn your personal information against you. Retrieved from http://cironline.org/reports/easily-obtained-subpoenas-turn-your-personal-information-against-you-5104

Appendix I: Survey questions used in Chapter 4

Note: We only show the questions analyzed in this thesis. Questions that were the same in the two surveys are numbered only (without any letters preceding the numbers). Questions that were different in the two surveys are marked using letters before the number (e.g., Pew survey items are designated "PEW", MTurk items are marked as "MTURK").

MTURK 1. Do you ever use a site like Twitter, Facebook, LinkedIn, Google Plus, or another social networking site? □Yes □No

PEW 1. Please tell me if you ever use the Internet to do any of the following things. Do you ever use the Internet to_____?

	Yes	No
Use a social networking site like Facebook, LinkedIn or Google Plus	□	□
Use Twitter	□	□

2. Is any of the following information about you available on the Internet for others to see? It doesn't matter if you put it there yourself or someone else did so.

	Yes, it's online	No, it's not online	Not sure	Does not apply
Your email address	□	□	□	□
Your home address	□	□	□	□
Your home phone number	□	□	□	□
Your cell phone number	□	□	□	□
Your employer or a company you work for	□	□	□	□
Your political party or political affiliation	□	□	□	□
Something you've written that has your name on it	□	□	□	□
A photo of you	□	□	□	□
Video of you	□	□	□	□
Which groups or organizations you belong to	□	□	□	□
Your birth date	□	□	□	□
Other information (please specify)				

3. Do you ever worry about how much information is available about you on the Internet, or is that not something you worry about? □Yes, worry about it. □No, don't worry about it. □Not sure

4. Considering everything you know and have heard about the Internet, do you think it is possible for someone to use the Internet completely anonymously – so that none of their online activities can be easily traced back to them? □Yes □No □Not sure

5. Have you ever tried to use the Internet in a way that hides or masks your identity from certain people or organizations?

□Yes □No □Not sure

[Measure of Prior bad experiences]

6. As far as you know, have you ever had any of these bad experiences as a result of your online activities?

	Yes	No	Not sure
Had important personal information stolen such as your Social Security Number, your credit card, or bank account information	□	□	□
Had an email or social networking account of yours compromised or taken over without your permission by someone else	□	□	□
Been the victim of an online scam and lost money	□	□	□
Been stalked or harassed online	□	□	□
Lost a job opportunity or educational opportunity because of something you posted online or someone posted about you online	□	□	□
Experienced trouble in a relationship between you and a family member or a friend because of something you posted online	□	□	□
Had your reputation damaged because of something that happened online	□	□	□
Something happened online that led you into physical danger	□	□	□
Something else bad happened (please explain: _____)	□	□	□

MTURK7. Do you ever post comments, questions, or information on the Internet using the following types of names?

	Yes	No	Not sure
Your real name	□	□	□
A username or screenname that people associate with you	□	□	□
A username or screen name that people do not associate with you	□	□	□
No name at all	□	□	□

PEW7. Do you ever post comments, questions, or information on the Internet _____?

	Yes	No
Using your real name	□	□
Using a username or screen name that people associate with you	□	□
Without revealing who you are	□	□

MTurk 8. Have you ever tried to use the Internet in such a way that <u>your family members, a romantic partner, certain friends, coworkers</u> would be unable to see what you have read, watched, or posted online? □Yes, I've done this. □No, I haven't done this.

MTurk 9. Have you ever tried to use the Internet in such a way that <u>an employer, supervisor, or companies you work for</u> would be unable to see what you have read, watched, or posted

online? □Yes, I've done this. □No, I haven't done this.

MTurk 10. Have you ever tried to use the Internet in such a way that people from your past, or people who might criticize, harass, or target you would be unable to see what you have read, watched, or posted online? □Yes, I've done this. □No, I haven't done this.

MTurk 11. Have you ever tried to use the Internet in such a way that law enforcement, the government, or companies or people that might want payment for the files you download such as songs, movies, or games would be unable to see what you have read, watched, or posted online? □Yes, I've done this. □No, I haven't done this.

MTurk 12. Have you ever tried to use the Internet in such a way that hackers, criminals, or advertisers would be unable to see what you have read, watched, or posted online? □Yes, I've done this. □No, I haven't done this.

PEW 8. Have you ever tried to use the Internet in ways that keep _____ from being able to see what you have read, watched or posted online?

	Yes, did this	No, did not
Family members or a romantic partner	□	□
Certain friends	□	□
An employer, supervisor, or coworkers	□	□
The companies or people who run the website you visited	□	□
Hackers or criminals	□	□
Law enforcement	□	□
People who might criticize, harass, or target you	□	□
Companies or people that might want payment for the files you download such as songs, movies, or games	□	□
People from your past	□	□
Advertisers	□	□
The government	□	□

13. Thinking about current laws, do you think the laws provide reasonable protections of people's privacy about their online activities? □ Yes, they provide reasonable protection □ No, they're not good enough □ Not sure

14. Do you think that people should have the ability to use the Internet completely anonymously for certain kinds of online activities? □ Yes, should have the ability □ No, should not have the ability □ Not sure

MTurk 15. Do you think the government should be able to monitor everyone's email and other online activities if officials say this might prevent future terrorist attacks? □ Yes, should monitor □ No, should not monitor □ Not sure

[Knowledge questions used in MTurk Survey in chapter 4, the interview study in chapter 5 and the experiment in chapter 6]

MTurk 16. How would you evaluate your computer literacy level? □ Very low □ Low □

Neither high nor low □ High □ Very high

MTurk 17. How would you evaluate your Internet literacy level? □ Very low □ Low □ Neither high nor low □ High □ Very high

MTurk 18. How would you rate your familiarity with the following concepts or tools?

	I've never heard of this.	I've heard of this but I don't know what it is.	I know what this is but I don't know how it works.	I know generally how this works.	I know very well how this works.
Cookie	□	□	□	□	□
Incognito mode/private browsing mode in browsers	□	□	□	□	□
IP address	□	□	□	□	□
Tor	□	□	□	□	□
Virtual Private Network (VPN)	□	□	□	□	□
Encryption	□	□	□	□	□
Secure Sockets Layer (SSL)	□	□	□	□	□
Proxy server	□	□	□	□	□
Privacy settings	□	□	□	□	□

MTurk 19. Please indicate whether you think each statement is true or false. Please select "I'm not sure" if you don't know the answer.

- Incognito mode / private browsing mode in browsers prevents websites from collecting information about you.
- Tor can be used to hide the source of a network request from the destination.
- A VPN is the same as a Proxy server.
- IP addresses can always uniquely identify your computer.
- HTTPS is standard HTTP with SSL to preserve the confidentiality of network traffic
- A proxy server can not be tracked to the original source.
- Website cookies can store users' logins and passwords in your web browser. *[*This question was removed in studies in chapter 5 and 6 due to ambiguity.]*
- No one, except for the sender and intended receiver, can reveal the content of an encrypted email. *[*This question was removed in studies in chapter 5 and 6 due to ambiguity.]*

[Social orientation measures in MTurk Survey]

MTurk 20. Do you agree or disagree with each of the following statements? □ Disagree strongly □ Disagree somewhat □ Neither disagree nor agree □ Agree somewhat □ Agree strongly

Collective identity

In general, belonging to social groups is an important part of my self- image.

The social groups I belong to are an important reflection of who I am.
To me, pleasure is spending time with others.
My happiness depends very much on the happiness of those around me.
Individual identity
I often do "my own thing".
I enjoy being unique and different from others in many ways.
Segmented identity
In different situations, I often act like very different persons.
I'm not always the person I appear to be.
I guess I put on a show to impress or entertain others.
I have parts of my life that are really very different from each other.
I would probably make a good actor.
I prefer to keep different parts of my life separate.
Other measures (not used in the analysis)
I am reading this question, not randomly selecting.
I generally have faith in humanity.
It is important to closely follow instructions and procedures.
Rules and regulations are important because they inform me of what is expected of me.
Standardized work procedures are helpful.
I generally trust other people unless they give me reason not to.
I tend to count upon other people.

These following questions are for statistical purposes only.

21. What is your gender? □ Male □ Female □ Other

22. How old are you (years)? _____

23. What is the highest level of school you have completed or the highest degree you have received?

□ Less than high school (Grades 1-8 or no formal schooling)
□ High school incomplete (Grades 9-11 or Grade 12 with NO diploma)
□ High school graduate (Grade 12 with diploma or GED certificate)
□ Some college, no degree (includes some community college)
□ Two year associate degree from a college or university
□ Four year college or university degree/Bachelor's degree (e.g., BS, BA, AB)
□ Some postgraduate or professional schooling, no postgraduate degree
□ Postgraduate or professional degree, including master's, doctorate, medical or law degree (e.g., MA, MS, PhD, MD, JD)
□ Not sure

MTurk 24. Where were you born?

□ China
□ India
□ United Kindom

□ United States
□ Other (please specify)_____

MTurk 25. Do you usually access the Internet from these locations?

	True	False	I'm not sure
China	□	□	□
India	□	□	□
United Kingdom	□	□	□
United States	□	□	□

Other (please specify)_____

Appendix II: Prescreen survey used in Chapter 5

This survey was given to the technical participants in our study as a prescreen test of their technical knowledge about networking. It was also given to students in a graduate level computer networking class. We computed the scale reliability by combining these two datasets together (the participants in our interview study and the students in the networking class). The 8-item survey had a Cronbach's alpha of 0.61. Question 5 and Question 7 marked with an asterisk had item-total correlations lower than 0.50. After we removed those two items from the scale, Cronbach's alpha for the scale was 0.79 (N = 33). Note: The correct answers are marked in black boxes.

Technical Network Knowledge Scale

1. What is a three-way handshake in TCP/IP?
☐ Three or more computers connected and communicating together
■ A method to establish a connection between two computers
☐ Three computers on the same LAN or WLAN
☐ A deal made between an ISP and a customer regarding Internet service
☐ I'm not sure

2. Which of the following protocols work on the Data-Link layer of the OSI Model?
☐ SMTP
☐ HTTP
☐ UDP
■ ARP
☐ I'm not sure

3. Which of the following is the correct order for the OSI model layers?
☐ Physical, Data Link, Transport, Network, Presentation, Session, Application
☐ Physical, Data Link, Network, Transport, Presentation, Session, Application
■ Physical, Data Link, Network, Transport, Session, Presentation, Application
☐ Physical, Data Link, Transport, Network, Session, Presentation, Application
☐ I'm not sure

4. Which numbers below represent an IP address?
☐ 2042.1.6.227
☐ 125.120.255
☐ 72.1380.12.86
■ 138.5.221.113
☐ I'm not sure

*5. Which of the following capabilities does Tor software have?

☐ Obscures your data even if someone is monitoring your network
■ Hides the source of a network request
☐ Can only be used by domain experts
☐ Acts as a VPN
☐ I'm not sure

6. Which of these statements about SSL/CAs is NOT correct?
☐ CAs can be compromised by attackers
☐ A CA is a third party organization
☐ A CA issues digital certificates
■ Using trusted certificates from a CA always guarantees the owner's identity
☐ I'm not sure

*7. What does the wireless network encryption tool WEP stand for?
■ Wired Equivalent Privacy
☐ Wireless Equivalent Privacy
☐ Wireless Equivalent Protocol
☐ None of the above
☐ I'm not sure

8. Of the following choices, what is the best choice for a device to filter and cache content from web pages?
☐ Web security gateway
☐ VPN concentrator
■ Proxy server
☐ MAC filtering
☐ I'm not sure

Appendix III: Interview script used in Chapter 5

Below is the text of our interviewer script along with our primary interview questions. Interviewers read this script to each participant prior to the drawing exercise and then went through the questions prompting the participant to illustrate their thoughts on paper while simultaneously explaining their diagram and thought process. Question 5, 6, and 7 marked with an asterisk were asked for each of the following activities: sending an email; making a payment online; receiving an online advertisement; browsing a website.

Interviewer:

I'm going to ask you to explain your perceptions and ideas about how the Internet works—keeping in mind how things work "behind the scenes"—when you are doing certain activities online. This is a drawing exercise. I'm going to ask you to draw how you think the Internet works on these papers (hand over pen and papers). Please talk aloud and explain your thought processes while you are drawing.

Please keep in mind that there is no correct answer to these questions—just answer these questions based on your own knowledge and experiences.

1. First off, we'd like to get a picture of how you envision the Internet. Can you draw on this paper and explain for me how you think the Internet works, or how you connect to the Internet?

2. Where do you think your data on the Internet goes? How does your data flow on the Internet?

3. Are there any other people, organizations or companies that can see your connections and activities?

4. Do you do anything to prevent others from seeing your connections and activities?

**5. Please recall an instance when you [watch a YouTube video] on your laptop (or computer). Can you draw and explain for me how you think that works.*

**6. Do you do this same activity on a smartphone? How do you think it works when you are connecting through your smart phone? Is there any difference?*

**7. Is there any example of this system didn't work? Why? Did there anything surprising or unexpected happened? What do you think happened?*

Appendix IV: Survey questions used in Chapter 6

[Scenario]

Please imagine you are using the Internet in the following scenario:

You are not satisfied with your current job and want to change jobs. You plan to search for other jobs online. Your home Internet broke down, so you decide to use the Internet in a public coffee shop. You found a new job search website: http://www.idealjobs.com

During registration, the website asks for some personal information such as your age, gender, current occupation, and current financial status. You are required to fill out the registration form before you can see available jobs.

[Survey measures]

Q1. Imagining yourself in the hypothetical scenario, how likely are you to conduct a job search using the Internet in the coffee shop? [Extremely unlikely/ Unlikely/ Neither likely nor unlikely/ Likely/ Extremely likely]

Q2. How likely are you to visit the http://www.idealjobs.com website to browse job opportunities?

Q3. The job website, http://www.idealjobs.com, asks you to enter the following information in the registration form. Given the hypothetical scenario, specify the extent to which you would reveal each of the following pieces of information through the Internet:

- Name
- Phone number
- Mailing address
- Education history
- Hobbies
- Employment status / current position
- Current financial status (e.g., annual income)
- Social network account (e.g., Facebook, LinkedIn)
- Personality test answers
- Criminal background or any past/present legal problems

Q4. Can you explain why you make the above choices?:_____

Q5. Click here to see the plug-in diagram again. Before you started this task, was your understanding of the Internet similar to what the diagram presents? [Not similar at all/ Slightly similar / Somewhat similar / Moderately similar / Extremely similar]

Q6. How clearly does the diagram communicate how the Internet works? [Not clear at all / Slightly clear / Somewhat clear / Moderately clear / Extremely clear]

Q7. How helpful is the information in this diagram for you to learn about how the internet works? [Not helpful at all/ Slightly helpful / Somewhat helpful / Moderately helpful / Extremely helpful]

Q8. How would you explain the diagram to a friend? (no fewer than 50 words):

Q9. Which of the following items are present in the browser plug-in diagram you just saw? Please select all that apply. [ISP/ Website server/ Router/ Your computer/ Others/ Internet/ DNS/ Network switch/ Advertisers/ Firewall/ Eavesdropper/ Government/ Content provider]

Q10. Do you have any other comments about the plug-in diagram?

Q11. At this time, how good is your understanding of how the Internet works? <poor, fair, good, very good, excellent>

Please imagine yourself in the previously stated scenario (conducting a job search in a public coffee shop), and rate how likely it is that each of the following persons or groups would be able to see some of your Internet activities.

Q12. How likely is it that your employer or supervisor would be able to see … ?
[Extremely unlikely/ Unlikely/ Neither likely nor unlikely/ Likely/ Extremely likely]

- your search history
- that you have visited the www.idealjobs.com website
- the personal information you submitted to the website in the registration form

Q13. How likely is it that advertisers would be able to see … ?
[Extremely unlikely/ Unlikely/ Neither likely nor unlikely/ Likely/ Extremely likely]

- your search history
- that you have visited the www.idealjobs.com website
- the personal information you submitted to the website in the registration form

[Ask the same question for:]

- Government or law enforcement
- Hackers
- Your family and friends, or other people you know
- The Internet service provider
- Other people who use the same network
- Company who owns the browser
- Company who owns the website
- Other users on the website

Q14. We are interested in any privacy concerns you might have when you are online. Please answer every question using the full scale provided. [Not at all/ Slightly/ Somewhat/ Moderately/ Very much]

- In general, how concerned are you about your privacy while you are using the Internet?
- Are you concerned about online organizations not being who they claim they are?
- Are you concerned about online identity theft?
- Are you concerned about people you do not know obtaining personal information about you from your online activities?
- Are you concerned that if you use your credit card to buy something on the internet your credit card number will be obtained/intercepted by someone else?
- Are you concerned that an email you send may be read by someone else besides the person you sent it to?

Q15. Please select the extent to which you agree with each of the following statements. [Strongly disagree/ Disagree/ Neither agree nor disagree/ Agree/ Strongly agree]

- I feel confident that I can mask my IP address.
- I feel confident that I can prevent others from seeing which websites I visited.
- I feel confident that I can communicate with others anonymously online, without revealing my real identity at all.
- I feel confident that I can prevent unwanted access to my personal information online.
- I feel confident that I can delete my digital traces (e.g. social network account, something I've posted in the past).
- I feel confident protecting my privacy online.

Q16. Please imagine yourself in the previously stated scenario (conducting a job search in a public coffee shop). How likely is it that you will do any of the following things when you are in the coffee shop? [Extremely unlikely/ Unlikely/ Neither likely nor unlikely/ Likely/ Extremely likely]

- Use a password that nobody else knows to activate your device
- Use a temporary username or email address
- Use a fake name or untraceable username
- Give inaccurate or misleading information about yourself
- Set your browser to disable or turn off cookies
- Clear cookies and browser history
- Use incognito mode or private browsing mode on your browser
- Use a service that allows you to browse the web anonymously, such as a proxy server, Tor software, or a virtual personal network
- Encrypt your communications
- Other_____

Q17. Please select the extent to which you agree with each of the following statements. [Strongly disagree/ Disagree/ Neither agree nor disagree/ Agree/ Strongly agree]

- In general, it is risky to reveal my personal information through the Internet.
- There is too much uncertainty associated with revealing my personal information through the Internet.
- Revealing my personal information through the Internet involves many unexpected problems.

Q18. Within the next 12 months, please estimate the probability that you will experience the following events, from 0% (the event is impossible) to 100% (the event is certain to happen).

- Have an email or social networking account of yours compromised or taken over without your permission by someone else
- Be the victim of an online scam or lose money
- Be stalked or harassed online
- Have important personal information stolen such as your Social Security Number, your credit card, or bank account information
- Lose a job opportunity or educational opportunity because of something you post online or someone posts about you online
- Experience trouble in a relationship between you and a family member or a friend because of something you post online
- Have your reputation damaged because of something that happens online
- Something happens online that leads you into physical danger
- Get into trouble with local authorities or government because of your online activities
- Have your personal information leaked by a company

Q19. Do you agree that people should have the ability to use the Internet completely anonymously for certain kinds of online activities? [Strongly disagree/ Disagree/ Neither agree nor disagree/ Agree/ Strongly agree/ I'm not sure]

Q20. Considering everything you know and have heard about the Internet, do you agree it is possible for someone to use the internet completely anonymously – so that none of their online activities can be easily traced back to them? [Strongly disagree/ Disagree/ Neither agree nor disagree/ Agree/ Strongly agree/ I'm not sure]

Q21. In 2014, Google has launched the "right to be forgotten" practice in Europe. Europeans can send requests to Google to have their personal data removed from search result if the information is inaccurate, inadequate, irrelevant or outdated. Some people criticize that allowing the "right to be forgotten" will violate the freedom of speech. Do you think the "right to be forgotten" practice would be useful for you or not? [Not useful at all/ Slightly useful/ Somewhat useful/ Moderately useful/ Extremely useful / I'm not sure]

Q22. If inaccurate information about you got posted online, do you agree that it would be very difficult to get it removed? [Strongly disagree/ Disagree/ Neither agree nor disagree/ Agree/ Strongly agree/ I'm not sure]

[Knowledge survey same as in Appendix I, omitted]

[Video evaluation task]

In this section of this survey, we would like to get your feedback on some educational videos we are developing for future study. You can choose one from the two YouTube videos below to evaluate. After you make the selection, you will see the video, and answer several short questions about that video.

Please select one of the following videos to evaluate:

1. An educational video about how to protect your privacy and security online https://www.youtube.com/watch?v=_p-LNLv49Ug
2. An educational video about how to conduct effective job search online https://www.youtube.com/watch?v=usJMn1l3F2I

[Actual disclosure question]

In this section, we are interested in your general risk-taking behaviors. Please answer the following questions: [Decline to answer/Never/Once or Twice/Often/Always]

- Have you ever used drugs of any kind (e.g.: weed, heroin, crack)?
- Have you ever downloaded pirated material (e.g., songs, videos, software) from the Internet?
- Have you ever lied about your age?
- Have you ever flown on an airplane?

[Demographic question same as in Appendix I, omitted]

Follow-up survey:

In this survey, you will answer questions about a hypothetical scenario. You will first read the scenario, and then you will answer some survey questions related to how you use the Internet. Your participation is completely voluntary and confidential.

Please imagine you are using the Internet in the following scenario:

You are not satisfied with your current job and want to change jobs. You plan to search for other jobs online. Your home Internet broke down, so you decide to use the Internet in a public coffee shop. You found a new job search website: http://www.idealjobs.com

During registration, the website asks for some personal information such as your age, gender, current occupation, and current financial status. You are required to fill out the registration form before you can see available jobs.

Q1. Imagining yourself in the hypothetical scenario, how likely are you to conduct a job search using the Internet in the coffee shop? [Extremely unlikely/ Unlikely/ Neither likely nor unlikely/ Likely/ Extremely likely]

Q2. How likely are you to visit the http://www.idealjobs.com website to browse job opportunities?

Q3. The job website, http://www.idealjobs.com, asks you to enter the following information in the registration form. Given the hypothetical scenario, specify the extent to which you would reveal each of the following pieces of information through the Internet

- Name
- Phone number
- Mailing address

- Education history
- Hobbies
- Employment status / current position
- Current financial status (e.g., annual income)
- Social network account (e.g., Facebook, LinkedIn)
- Personality test answers
- Criminal background or any past/present legal problems

Q4. At this time, how good is your understanding of how the Internet works? [poor/fair/good/very good/excellent]

Q5. How likely is it that your employer or supervisor would be able to see … ? [Extremely unlikely/ Unlikely/ Neither likely nor unlikely/ Likely/ Extremely likely]

- your search history
- that you have visited the www.idealjobs.com website
- the personal information you submitted to the website in the registration form

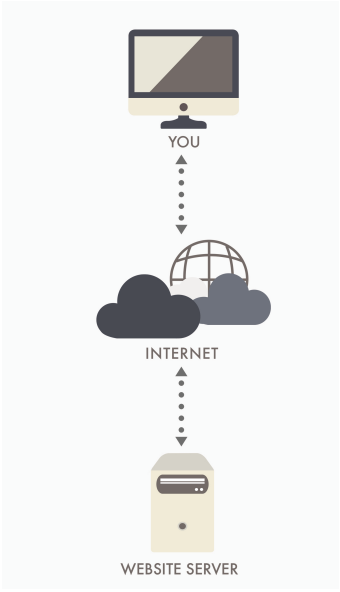
Q6. How likely is it that advertisers would be able to see … ? [Extremely unlikely/ Unlikely/ Neither likely nor unlikely/ Likely/ Extremely likely]

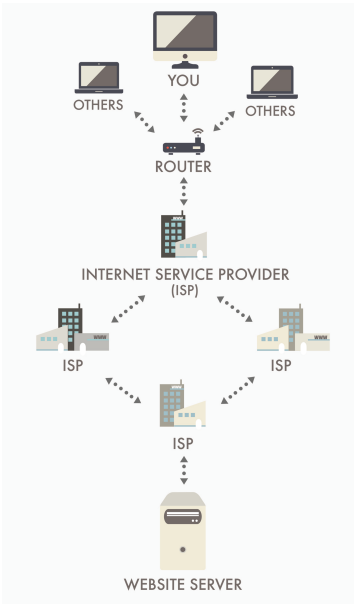
- your search history
- that you have visited the www.idealjobs.com website
- the personal information you submitted to the website in the registration form

[Ask the same question for:]

- Government or law enforcement
- Hackers
- Your family and friends, or other people you know
- The Internet service provider
- Other people who use the same network
- Company who owns the browser
- Company who owns the website
- Other users on the website

Q7. Which of the following two visualizations is more similar to your own understanding of the Internet?

A. 

B. 

Q8. In this section, we are interested in your general risk-taking behaviors. Please answer the following questions [Decline to answer/Never/Once or Twice/Often/Always]

- Have you ever had a sexual relationship with somebody other than your partner without their knowledge or consent?
- Have you ever tried to gain access to someone else's email account (e.g., a partner's, friend's, colleague's) without their knowledge or consent?
- Have you ever made a donation to a non-profit organization?
- Do you always turn the lights out at home and work, even if you're feeling lazy?