Incognito Online: Why and How People Hide their Digital Traces

Ruogu Kang
Ph.D. Thesis proposal
Jan 8th, 2014

Thesis committee:
Sara Kiesler (Chair, HCII, CMU)
Laura Dabbish (HCII & Heinz, CMU)
Lorrie Cranor (ISR & EPP, CMU)
Alessandro Acquisti (Heinz, CMU)

Summary
The Internet contains much personal information, and many people are worried about the possible consequences of exposure of this information to others. Some people have actually experienced emotional or tangible damage from exposure of their personal information. My prior work shows that most people who use the Internet sometimes want to hide their identity or their online interactions or content.

The first part of this dissertation explores the motivations and strategies of those who have sought anonymity or who have tried to hide some aspect of their identity, interactions, or content online. People's reasons for seeking anonymity range widely, from protecting family from unpleasant gossip (a relationship threat) to hiding from hackers or government surveillance (information threats). I have also examined how individual differences in social orientation, past negative experiences, and people's technical knowledge shape their perceptions of different privacy threats, and how those perceptions motivate different strategies to hide from those threats. According to these studies, technical knowledge and past negative experiences are two major filters determining perceptions and strategies for mitigating privacy threat.

The second part of this dissertation examines how people understand the Internet and how they envision information transmission over the Internet. A think aloud study during which people were asked to draw how information passes over the Internet suggests that nontechnical users lack awareness of the complex structure and important entities in the network, and that technical users tend to be overconfident, leading them to potentially overlook or misplace some privacy threats to their personal information.

My research so far suggests that many people have the desire to hide certain online information but do not have sufficient knowledge or the correct tools to do so. My proposed work will explore how levels of threat (manipulated and measured via the proxy variable: negative online experience), and simple versus complex knowledge of the Internet, influence people's decisions to seek anonymity or use information-hiding tools. The findings of my dissertation will inform the design of future Internet architecture and applications to help Internet users better protect their information, and will contribute to the body of research in privacy and HCI.

Table of contents

Chapter 1. Introduction **3**

Chapter 2. Background **5**
2.1 Social privacy threat **5**
2.2 Information privacy threat **7**
2.2.1 Threat from government 7
2.2.2 Threat from companies and other third parties 8
2.3 Coping with social and information privacy threat **9**
2.3.1 Strategies people use to cope with privacy threat 10
2.3.2 Barriers in adopting effective strategies 12
2.4 Summary **14**

Chapter 3. Why and how do people seek anonymity on the Internet? **15**
3.1 Introduction **15**
3.2 Results **16**
3.2.1 Anonymous activities 16
3.2.2 Reasons for seeking anonymity 16
3.2.3 Strategies people use to attain anonymity 18
3.2.4 People are uncertain about how anonymous they are 19
3.3 Summary **19**

Chapter 4. Users' perception and strategies of hiding their online information **21**
4.1 Introduction **21**
Factors affecting how people hide information online 21
4.2 Results **23**
4.2.1 Hiding identity and hiding interactions from specific groups 23
4.2.2 Strategies people use to hide content and interactions 24
4.2.3 Individual differences factors affecting how people manage their information 25
4.3 Summary **29**

Chapter 5. Users' mental model of the Internet **31**
5.1 Introduction **31**
Mental models 31
Technical vs. nontechnical users 32
5.2 Results **33**
5.2.1 View of the Internet structure: Black box, simple chain, or complex system 33
5.2.2 Infrastructural awareness 35
5.3 Summary **38**

Chapter 6. Proposed work: The effect of privacy threat and Internet knowledge on users' decisions to adopt privacy strategies **39**
6.1 Introduction **39**
6.2 Study plan **42**
Manipulations 43
Dependent variables 45
Other measures 46
Procedure 46
6.3 Timeline **47**

References **48**

Appendix: Survey questions **55**

Chapter 1. Introduction

The Internet contains large amounts of personal information, but most Internet users do not understand what information about them is revealed to others, where their information is held, and who has access to this information (Rainie et al, 2013; Berstein, 2013; Lin, 2012). Widespread news has raised people's concerns about government surveillance[1], company data leakage[2], and various tracking techniques launched by websites, apps, and even mobile service providers[3] (Tene & Polonetsky, 2012). People's own activities on social media, if revealed in unintended ways, can endanger their social relationships (Litt et al, 2014). These phenomena have made people feel increasingly worried or threatened, and some who have unintentionally revealed personal information have experienced emotional or tangible damage (Woodruff, 2014; Kang, 2013; Shay 2014). Because of these concerns and experiences, some people try various strategies to hide their digital traces – their identity, content, or interactions on the Internet. The main goals of my dissertation are to understand why people try to hide their digital traces, what strategies they use to do so, and explore the factors that influence their perceptions of threat and their decisions to hide from sources of threat. The findings of my dissertation will have implications for the design of future Internet architecture and applications that protect people's privacy.

The first part of this dissertation explores the motivations and strategies of those who have sought anonymity or who have tried to hide some aspect of their identity, content, or interactions online. I have conducted an interview study of anonymity-seekers in the U.S., China, and Europe (Kang et al., 2013; described in Chapter 3). Reasons for seeking anonymity range widely, from protecting family from unpleasant gossip (a relationship threat) to self-protection from hackers or government (information threats). I have also conducted two surveys (one U.S. only random sample, and one international sample; described in Chapter 4) to examine how individual differences in social orientation, past negative experience, and technical knowledge shape people's perceptions of different threats, and how those perceptions motivate different strategies to hide from those threats. This part of my work aims at answering the first research question:

Q1. *Why and how do people try to hide their identity, interactions and content online?*

My findings show that most people who use the Internet at times, and for some purposes, want to hide their identity, content, or interactions. However, only very few people with advanced technical knowledge and with a specific threat in mind employ sophisticated technical strategies to attain anonymity or anonymize information. The second part of this dissertation examines how people understand the Internet and how they envision various privacy threats that may happen to information transmission over the Internet. I conducted a qualitative study asking technical and nontechnical users to draw their mental models of the Internet, and asked experts to evaluate these models to identify areas of misinformation or lack of knowledge that might put user security and privacy at risk (Chapter 5). The second research question is:

[1] http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order
[2] http://www.usatoday.com/story/tech/2014/10/02/jp-morgan-security-breach/16590689/
[3] http://www.wired.com/2014/10/verizons-perma-cookie/

Q2. *What do people know about how the Internet works and how others may gain access to their information?*

People's mental models reveal that nontechnical users lack awareness of the complex structure and important entities in the network, which could lead them to overlook threats to their personal information. Technical users without experiencing any negative events online tend to be overconfident about their knowledge and have a false sense of security. It isn't yet known whether we can update people's perception of threat and motivate them to make better choices to protect their information from different kinds of threats. In my proposed work, I will conduct a controlled experiment that induces different threat models by priming people with a prior negative event and depict simple or complex models of the Internet. I will measure their tendency to try to protect themselves using anonymization tools (Chapter 6). The goal of this experiment is to answer this research question:

Q3. *How do people's perception of threat and knowledge of the Internet affect their decisions to hide their information?*

The findings of my dissertation will make the following contributions:

- **Design contribution:** This dissertation will provide design implications for designing future privacy management technologies and tools. By studying what strategies people are currently using to hide their online information, this work can point out potential barriers for people to adopt more effective strategies. It will also provide directions for improving system transparency and user education in the design of future Internet systems.

- **Theoretical contribution:** Most previous work investigates how users manage their privacy on one social platform (e.g., Facebook) or focuses on the evaluation of certain security tools (e.g., firewalls). Some research deals with privacy problems related to one's social life, and other work examines problems with government and organization surveillance. My work looks at these various aspects together and provides more understanding of how users hide their information from different privacy threats (including social and information threats). I have conducted a series of studies to provide both qualitative and quantitative evidence showing how individuals' background knowledge and Internet experience influence their behavior and perceptions.

Chapter 2. Background

Over years of research, the concept of privacy is still considered hard to define by many scholars in different domains. Law researchers define privacy as "the right to be let alone" (Warren and Brandeis, 1890). Westin (1967) defines: "Privacy is the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." Petronio (2002) explains the management of privacy boundaries as "how people manage the relationship between revealing and concealing." Altman (1975) suggests privacy as a dynamic, and context-dependent "boundary regulation" process. Solove (2007) synthesizes a wide range of discussions around the conceptualization of privacy, and defines it as a "plurality of different things" rather than one single concept. The taxonomy he developed covers the collection, processing, dissemination, and invasion of personal information. All these definitions and properties can be used to analyze how individuals deal with their relationships with other people, institutions or organizations both online and offline. This dissertation specifically focuses on privacy problems related to personal information, not physical privacy (e.g., an individual's physical property being intruded by others). 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.

In today's Internet, people are not only concerned about how companies or government collect and use their personal data, but also what can be seen by their friends and families, or a random stranger who comes across their Facebook profile through a friend of their friends. Sometimes the latter concern can be even more threatening. Palen and Dourish (2003) address both concerns in their paper, using examples of surveillance, personal identity theft, and interpersonal privacy matters. Rader (2014) distinguishes these two types of concerns into information privacy and social privacy, and my dissertation will address issues around people's motivation and behaviors hiding from both types of privacy threats.

In this section I first review literature related to the social aspect of privacy threats, including previous literature on self-presentation, context collapse and boundary regulation; then I review literature about informational privacy threats, and people's privacy concerns and actions toward their information being collected and used by companies and institutions. In the last part I summarize privacy protection strategies listed in previous work, and then discuss some reasons why people feel helpless or are unable to take effective strategies to manage their information towards different threats.

2.1 Social privacy threat

Rader defines social privacy as "how we manage self-disclosure, availability, and access to information about ourselves by other *people*" (p52). This concept is similar to Iachello and Hong's (2007) definition of personal privacy -- "how people manage their privacy with respect to other individuals." Both concepts reflect the idea that social privacy problems are essential to how people regulate their social boundaries (Ashforth 2000; Altman 1975; Petronio 2002).

One motivation for controlling or hiding certain information from others is to manage one's self-presentation. The presentation of self, as defined by Goffman (1959), is how people express themselves in the presence of others. Baumeister (1982) suggests the goal of self-presentation is to gain social approval (to be liked or accepted) from others. 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 et al 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, Suler (2002) proposes that even the communication channels people choose can reflect their identity.

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). And sometimes these roles are incompatible with each other. For example, I interviewed a fan fiction writer and also a school teacher who use multiple Facebook accounts to maintain separate identities (Kang et al. 2013). 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, 1986). Farnham and Churchill (2011) suggest that how those 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.

Social boundaries are often intertwined, and difficult to manage. Prior studies show that people have inaccurate understanding of who can see the information they share online. They cannot accurately 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, people's privacy sometimes can be violated by what others share about them. Litt and colleagues (2014) conducted a survey with Facebook users and showed evidence that many people have experienced their self-presentation goals being violated by the content shared by their friends. An article used the term "peer surveillance" [4] 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. In the next section I summarize prior literature about users' perceptive on their information being collected and used by government and other institutions.

2.2 Information privacy threat

Information privacy, according to Rader (2014), 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). This set of problems deal with two types of threats: authorities (government or employers), and companies (businesses people directly interact with and third parties). Smith (1996) examined people's concerns about organizational privacy practices and categorized those concerns into collection, unauthorized secondary use, errors, and improper access to personal information.

2.2.1 Threat from government

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. Shklovki and Kotamraju (2011) interviewed people who experience government blocking and censorship in their daily use of the Internet. They found that 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. Dinev et al (2008)'s survey study shows that Internet users who support government surveillance are more willing to provide personal information online, and have lower privacy concerns. Those who are concerned about privacy are also more concerned about government monitoring, and are less willing to provide personal information over the Internet. 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, et al 2008).

Is the public ok with government surveillance? A 2006 survey[5] conducted after Bush administration's surveillance program got publicized shows 53% of the U.S. public accept government surveillance when "this was necessary to reduce the threat of

[4] https://modelviewculture.com/pieces/social-networking-as-peer-surveillance
[5] http://www.nytimes.com/2006/01/27/politics/27poll.html

terrorism" is stated in the survey question, but this percentage dropped to 46% when the above sentence is removed. A Pew survey conducted in June 2013[6] (one week after Snowden revelation) shows 56% of the Americans they sampled agree that NSA tracking civilians' telephone records is acceptable to investigate terrorism, and 41% disagree. These surveys demonstrate that overall public opinion towards government surveillance is almost unchanged from 2006 to 2013. On the other hand, another research study (Twenge et al. 2014) shows a declining trend in American public's trust and confidence in large institutions such as business and congress from 1972 to 2012. These articles and surveys seem to show mixed evidence about users' opinion about their personal information being accessed by authorities. It is still unclear what factors influence these perceptions and opinions, such as one's cultural background, political environment or personal experience.

2.2.2 Threat from companies and other third parties

Recent advances in technology and "big data" analytics make information collection and processing by companies and other third parties more visible to users. Users' opinion about their information being used and analyzed by companies or other third parties is more important today than ever. 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 et al 2010). It is now much easier to identify and track individual users through their mobile phones [7] [8], Internet activities [9], and other ubiquitous devices or sensors [10]. A group of researchers analyzed students' internet traffic flow collected by the campus network and found that certain Internet usage are associated with higher depression scores such as high email usage and high amounts of file sharing (Katikalapudi et al, 2012). 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. When asked about specific sources of threat, the participants were more concerned about RFID data found out by thieves and strangers, less concerned about government and companies. People's incorrect or incomplete understanding of how these technologies work contributes to their underestimation of the potential threats.

Personalization techniques and behavioral profiling have been widely used by companies in recommendation services and behavioral advertising. Before the proliferation of online advertising, Culnan (1993) examined what influences people's attitudes about secondary information use in direct mail marketing. Her analysis shows that those who perceive more benefits of shopping by mail have lower privacy concerns about the loss of control,

[6] http://www.people-press.org/2013/06/10/majority-views-nsa-phone-tracking-as-acceptable-anti-terror-tactic/

[7] http://www.huffingtonpost.com/2011/04/20/apple-iphones-tracking-users_n_851532.html

[8] http://www.informationweek.com/software/social/facebook-ads-invade-apps-user-locations/d/d-id/1316563

[9] http://www.wired.com/2014/10/verizons-perma-cookie/

[10] http://www.forbes.com/sites/parmyolson/2014/06/24/google-nest-smart-home-internet-of-things/

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. 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. 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. The authors argue that this is because the perceived risk associated with advertising is more salient. Ur et al (2012)'s study shows that more than half of their interviewees are aware that the ads they see online are personalized, and people perceive OBA as both "useful and privacy invasive", and they are "scared about being tracked and monitored". 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. Leon et al (2013) found users' willingness to disclose information for OBA varies significantly depending on information types (personal identifiable information are considered most sensitive), data-retention policies (more willing to disclose if data is only retained for one day), and the scope of data use. They also showed that giving user better control over their information increased their willingness to share. Contrary to Ur et al (2012), their experiment shows that people's opinions do not differ for well-known website versus unfamiliar website.

In addition, people may not be aware that their activities on different sites can be linked together to identify them [11]. Companies, governments, and individuals are collecting and using others' personal data for a myriad of purposes. 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 and Wills, 2001).

2.3 Coping with social and information privacy threat

These prior work cited above tells us that people are concerned about their online information for social and nonsocial reasons, but most of them do not have enough expertise or knowledge to clearly understand what happens to their data, and only have vague ideas of what to do about it. In 2013, a University of Pittsburgh researcher murdered his wife using cyanide, and the key evidence for this case is that his Google search history contains searches for cyanide multiple times. He also searched for how to remove his computer search history, but obviously he did not succeed in hiding his search

[11] http://cironline.org/reports/easily-obtained-subpoenas-turn-your-personal-information-against-you-5104

traces[12]. Although in this case the failure to use the correct strategy to hide works to the advantage of law enforcement, it still demonstrates the technical barrier for lay people to choose the appropriate strategy to hide their digital traces. In this section, I will first review users' coping strategies summarized by previous literature, and then discuss why people are not able to take effective strategies.

2.3.1 Strategies people use to cope with privacy threat

There have been decades of research on practices to protect 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). Albrechtsen (2006) surveyed people's security actions in organizations, including cautious use of email, mobile devices and Internet, lock computers, and manage passwords. 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). Turner and Dusgupta (2003) have reviewed a list of 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. In Milne et al (2005)'s paper, they summarized several online privacy protection behaviors suggested by Center for Democracy and Technology and FTC: install firewall; opt out of third party information sharing; read online privacy policy; clear cache; reject unnecessary cookies; use anonymous emailers; encrypt sensitive data; use anonymizers while browsing, and other methods. A list of the strategies reviewed in prior literature is shown in Table 1.

Many of these strategies mentioned above focus on security problems, such as preventing malicious attacks using antivirus software and using secure passwords. The focus of my dissertation – hiding one's identity, communication and content – is more about concealing information from others. Technical strategies that deal with this specific need should include using anonymizers, encrypting data, deleting cookies or cache and using anonymous emailer.

People also use a number of strategies to control and manage their social privacy threat. 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

[12] http://www.reuters.com/article/2014/11/06/us-usa-pennsylvania-cyanide-idUSKBN0IQ2MY20141106

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). Lampinen et al (2011) summarized a list of strategies that people use to collaboratively manage privacy concerns in social network sites, such as sharing with separate groups, deleting comments or tags, or asking others to delete unwanted content.

Strategies to cope with:
Information privacy • Read privacy policy • Use secure passwords • Use anti-virus software and other security software (firewall) • Keep email hygine (e.g., don't open attachments from unknown sender) • Block websites, avoid phishing websites • Delete or manage cookies • Clear cache • Use anonymizers (proxy, SSL) • Encrypt data • Use anonymous emailer
Social privacy • Use privacy settings • Use multiple social networking profiles • Use different sites or different communication channels • Selectively share information to different audiences, share with separate groups • Limit the amount of information shared (e.g., do not share real name or identifiable information) • Use encoded languages in their posts • Deactivate account • Delete comments or ask others to delete • Self-censor posts

Table 1. Strategies people use to cope with information and social privacy threat found in literature

2.3.2 Barriers in adopting effective strategies

Although people can choose from this long list of strategies, there are several barriers that 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 et al, 2005; Woodruff et al 2014). Prior work has documented a phenomenon of "privacy paradox" – individuals' actual behaviors do not align with their concerns (Spiekermann, 2001) and people often disclose more than they intended to (Norberg, 2007).

Unstable risk perception

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 et al 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 et al 2010).

Referring to literature in social psychology and behavioral economics, we learn that surface level cues or the framing of a question can change people's behavior from their desired intentions. 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. 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 et al. (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. 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.

Lack of knowledge and awareness

Another barrier is that most 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. People with more computer-related technical knowledge 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 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)

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. Dommeyer and Gross (2003) found that consumers are aware of privacy protection strategies, but do not use them. They also found younger men are more likely to be aware of those strategies, but those who are more likely to use those strategies are young people and those who have negative attitudes toward direct marketing. Rhee et al (2009) examined people's self-efficacy in information security (self-efficacy is individual's self-evaluation of their behavior). Their study shows that more experience with Internet and computers increases people's self-efficacy level, and higher self-efficacy level is associated with more use of security software and features. However, contrary to this finding, Jensen et al (2005) found that many people's self-reported knowledge is usually inaccurate, and higher than their actual knowledge about privacy technology, suggesting that people may be overconfident about their security on the Internet.

Many lay person's perception of managing their Internet activities stays 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 et al. (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; removing the authentication image prevented 3% from entering their passwords; and an explicit warning page was most effective – 47% did not enter their passwords.

2.4 Summary

Overall, previous literature suggests that many people who are concerned about privacy have a specific source of threat in mind. There are social and relational reasons for people to conceal or limit access to their information online, such as maintaining their self-presentation online or managing different boundaries of their lives. People are also concerned about being tracked, monitored, and analyzed by government or companies. However, just having the fear (threat) in mind may not motivate people to act, because they don't know what to do. The literature on fear appeals suggests that making people aware of danger is only one part of getting them to take action. They also have to feel they can do something that will work (Witte 1995). To deal with these concerns, people can choose from a wide range of strategies with different levels of technical sophistication, but prior work reveals many challenges to the effective prevention of these privacy threats, including biases in correctly estimating the risks, the lack of awareness of where the threats are, and lack of knowledge of protection strategies.

Despite the rich literature on privacy and security, there is not much research about Internet users' own experiences *hiding* their online information. My dissertation specifically examines what people do to hide their information across different Internet activities and platforms, and who they are concerned about during everyday use of the Internet. My proposed work will explore intervention techniques to modify their risk perception, improve user awareness and knowledge, and nudge people toward better privacy decisions.

Chapter 3. Why and how do people seek anonymity on the Internet?[13]

3.1 Introduction

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 used 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. The main purpose of this chapter is to examine how people think about online anonymity, and why they seek it.

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. What we know about these reasons is derived mainly from studies of particular activities or groups who intentionally seek anonymity, including whistle blowers (Greenberger et al 1987), members of stigmatized groups (Mckenna et al. 2000), people conducting sensitive searches (Conti et al 2007), hackers (Coleman et al 2008), and lurkers (Preece et al 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 et al 2004).

Another purpose of this work is to investigate the strategies people use in trying to achieve anonymity online, 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. 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). Most tools available to achieve online anonymity are poorly understood. We wanted to discover how users try to achieve anonymity, and whether they are confident that they have achieved it.

To study these problems, I conducted one-hour semi-structured remote interviews with 44 Internet users (23 woman, and 21 men) who said they had done something anonymously online in the past, and who volunteered for the study. Interviewees were

[13] This chapter is based on my paper: Kang, R., Brown, S., and Kiesler, S. Why do people seek anonymity on the Internet?: Informing policy and design. In *Proc. of CHI 2013,* ACM (2013), 2657-2666.

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.

3.2 Results

3.2.1 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. Many anonymous social activities, however, were associated with groups. We categorized seven categories of social activities that people participate in anonymously, including participating in special interest groups (mostly hobby groups such as fiction, music, pets), social networking, sharing art or work, exchanging help and support, buying and selling, discussing or being involved in politics, and reviewing and recommending products or services.

3.2.2 Reasons for seeking anonymity

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.

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".

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. 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.

Personal threat models

Interviewees' reasons for seeking anonymity reflected a personal "threat model" of persons 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)

3.2.3 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.

3.2.4 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.

Under Marx's definition of anonymity, 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.

3.3 Summary

This chapter summarizes an interview study I conducted with an international sample. The findings contribute to the first research question, by revealing rich real life examples of people trying to hide their online identity. Main highlights of the findings include:

- People participate in a wide range of anonymous activities online, including not only malicious or illegal activities, but also pro-social and altruistic activities.

- Many people consider a "personal threat model" of persons and organizations when seeking anonymity online. Some factors that affect the construction of this personal threat model include prior negative experience and the desire to manage boundaries in their lives.
- Interviewees use both behavioral and technical strategies to attain anonymity online. Technical strategies require some more advanced technical skills, whereas behavioral methods are used by almost everyone regardless of their background knowledge.
- Although most people hide their identity with a specific threat in mind, many still suffer from the sense of uncertainty. Some expressed concern about unknown threats – although they don't know whom they are afraid of, they are hiding from uncertain threats. Most people do not have accurate perception of how anonymous they are, who they may be anonymous to and do not know their activities or identities across platforms can be connected to identify them.

Although this chapter provides rich qualitative evidence for online anonymity-seeking, the interview study used a self-selected, anonymity-seeking sample. In the next chapter, I present results from two surveys with representative samples of the Internet users. The goals are to: 1) find out how representative the needs for anonymity are; 2) quantitatively examine the relationships between people's individual differences (technical knowledge of the Internet, boundary regulation needs and prior negative experiences) and their perceptions and actions of hiding online traces.

Chapter 4. Users' perception and strategies of hiding their online information[14]

4.1 Introduction

Most Internet users try to control access to their online personal information in numerous ways. In this chapter, I report findings from two surveys to examine the relationship between individual difference factors and people's behavior hiding identity and selectively hiding interactions online. Also, I look at 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.

The questions used in the two surveys were developed based on the interview questions about anonymity in Kang et al (2013) and questions on privacy that the Pew Research Center fielded in its previous surveys. Survey questions are attached in the Appendix. The first survey was conducted by Pew Reserach Center by calling people on their landline phones or cell phones in July 2013. This dataset I analyzed includes 775 U.S. Internet users. The second survey was conducted on Amazon MTurk, a crowdsourcing platform, in February 2014. The survey was hosted online on SurveyMonkey. We recruited 396 MTurk workers, including 182 respondents from the U.S., 128 from India, and 86 respondents from a variety of other countries. The analyses shown in this Chapter only include the 182 U.S. MTurk users.

Factors affecting how people hide information online

From the study findings in Chapter 3, we learned that each person's life experience and personal preferences can lead to a different "personal threat model" and influence their privacy-related concerns and actions. Prior literature also suggests that people's demographic characteristic, computer and Internet experience, prior negative Internet experience, and social orientation can shape their perception of privacy and their conceptualization of what can be considered as a threat.

The demographic characteristics of a group of people may be highly predictive of their attitudes. For instance, younger people may be more politically liberal than older people, which could lead to more concern about privacy and supports for free of surveillance. Some existing work shows that 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. Many papers mentioned in Chapter 2 show the effect of technical knowledge on people's privacy perceptions and behaviors. In the MTurk survey, we measured respondents' computer and Internet knowledge using their self-rated familiarity

[14] This chapter is based on the datasets used in these publications:

- Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., & Dabbish, L. Anonymity, Privacy, and Security Online. Pew Research Center (2013). http://www.pewinternet.org/2013/09/05/anonymity-privacy-and-security-online/
- Kang, R., Brown, S., Dabbish, L., & Kiesler, S. Privacy Attitudes of Mechanical Turk Workers and the US Public. In Proc. of SOUPS 2014,USENIX (2014),38-49.

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.") . In addition, other factors could also cause differences in perceived privacy threat, such as people's basic social orientation (which derive from personality and culture) and their own past Internet experience.

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. In the MTurk survey, we adopted existing scales to measure three types of social orientation: 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]). Questions are attached in appendix (we also used some other personality measures but did not use them in analysis).

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 3 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. We constructed a survey question based on these experiences and asked respondents in both surveys if they had experienced any of ten different harmful situations online, such

as “Had your reputation damaged because of something that happened online”. The question is attached in appendix.

4.2 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%). We asked respondents if they had experienced any of ten different harmful situations online (e.g., had your reputation damaged because of something that happened online; been the victim of an online scam and lost money; full list attached in Appendix). MTurk sample is more likely to have experienced negative experience than the Pew sample (MTurk: 49%; Pew: 36%).

4.2.1 Hiding identity and hiding interactions from specific groups

In order to get an overall sense of whether people had tried to hide their identity, we asked: “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” were coded as having tried to hide their identity. Table 2 shows the percent of people from both surveys who have tried to hide their identity online.

To examine whether people were selective in who they tried to hide from, we asked: “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?” about 11 types of persons or organizations. 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. From a three-factor solution accounting for 56% of the variance, we created three scales, using items that loaded .40 or more on the factors: threat to personal relationships (family or romantic partner, certain friends, employer or coworkers, people who might criticize or harass, people from one’s past; Cronbach’s alpha = .75); threat from authorities (law enforcement, companies wanting payment, government; Cronbach’s alpha = .61), and threat from unwanted influence or exploitation (advertisers, hackers, criminals; Cronbach’s alpha = .59). Later in the MTurk survey, we divided the first scale into three subgroups and asked respondents whether or not they have tried hide from all five groups separately, including family, friends, co-workers (“your family members, a romantic partner, certain friends, or coworkers”), employers and supervisors (“an employer, supervisor, or companies you work for”), unwanted ties (“people from your past, or people who might criticize, harass, or target you”), authorities (“Law enforcement, the government, or companies or people that might want payment for the files you download such as songs, movies, or games”), and other third-parties (“hackers, criminals, or advertisers”). In table 2 we listed five groups.

	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 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 authorities	10%	18%
Hide from other third parties	44%	28%

Table 2. Percent who have tried to hide their identity and percent who have tried to hide from different groups

As shown in the above table, the most common threat identified by the MTurk sample was family, friends and coworkers, whereas the most common threat identified by the Pew sample was other third parties including advertisers, hackers and criminals. 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.

4.2.2 Strategies people use to hide content and interactions

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 3.

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.

When analyzing the data, we categorized nine of these methods 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 3 shows the percent of respondents from each survey who

had reported using each type of strategies. The most commonly used strategy 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 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 3. 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 1 shows what kind of strategies respondents used to hide from each threat. From the figure, we see that that managing cookies (including clearing a browser history) and using alias were the approaches that respondents most commonly used to protect themselves from almost all privacy threats. Those who had tried to hide from institutions or unknown third parties (e.g., government, law enforcement, companies, advertisers) were more likely to use the technical methods, but less likely to edit content than those who hid from the three types of social privacy threats.

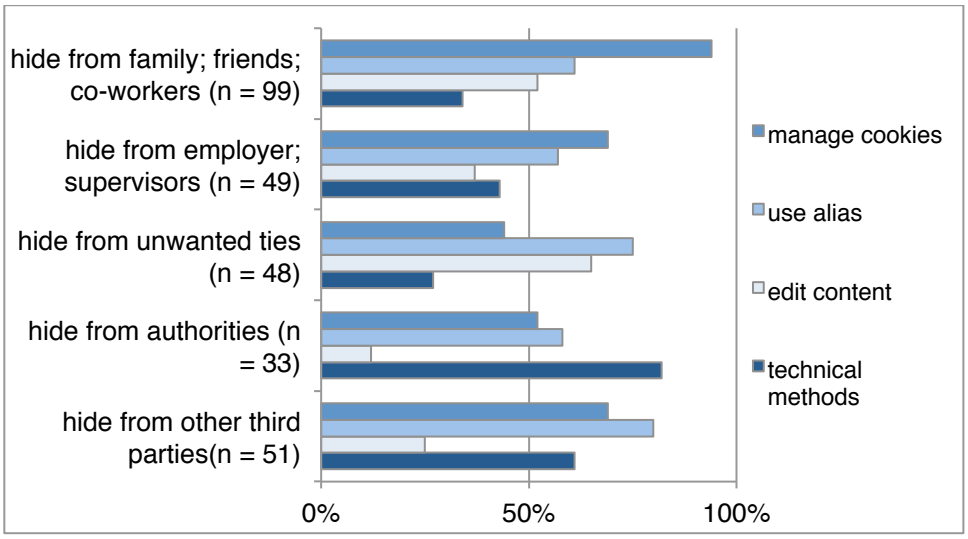


Figure 1. Percent of respondents who used each category of strategies to hide their interactions, divided by source of privacy threat. Data shown in this figure is from the U.S. MTurk sample (N = 182)

4.2.3 Individual differences factors affecting how people manage 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 age, gender and social media use affect people's hiding behaviors. 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.

Because of these sample differences, I present results from the two surveys separately in the following.

Demographics, social media use, and negative experience (Pew survey)

As shown in the following table, 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). Worry about information is used as a proxy to measure their general privacy concern. 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. Having had negative experience online consistently predicts hiding from all kinds of privacy threats. Lastly, those who said they are worried about information online are also more likely to hide from almost all threats except for family and friends.

		Hide from known groups (social privacy threats)			Hide from organizations (information 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*
Have 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 4. Factors predicing hiding identity and interactions 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 5). People who are younger, use social media, and have bad experience are significantly more likely to use all types of 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 methods**
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*
Have 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 5. Factors predicing strategies they use to hide. HS grad or less (N=203), some college/associate degree (N=243), and college or above (N = 327).

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.

Technical knowledge, prior negative experience, and social orientation (MTurk survey)
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. Prior negative experience predicts hiding from social privacy threats in this sample, but not hiding from informational threats. Technical knowledge strongly predicts hiding from informational threats, hiding from employers and marginally predicts hiding one's identity. 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. Past negative experience and technical knowledge have almost no correlation (r = -.03)

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.

			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
Have 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 6. Factors predicing hiding identity and interactions from people or organizations. Data shown in this table is from the U.S. MTurk sample (N = 182)

In addition, we examined the effect of technical knowledge, prior negative experience, and social orientations on the strategies people use to mitigate different threats (Table 7). 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). Having bad experience is associated with more use of editing content and technical methods. High segmented identity and low collective identity orientation predicts more use of editing content. High segmented identity is also associated with using technical methods.

	Manage cookies	**Use alias**	**Edit content**	**Use technical methods**
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**
Have bad experience	.05	.03	.25**	.12†
Technical knowledge	.15†	.10	-.07	.48***
R^2	.08	.12	.16	.30

Table 7. Factors predicting strategies they use. Data shown in this table is from the U.S. MTurk sample (N =182)

The effect of technical knowledge and negative experience on policy preferences and perceptions (MTurk survey)

The previous data shows that people with more technical knowledge reported hiding from more threats, and taking more technical methods to protect their information. But do they feel more secure than those without technical knowledge, or feel less secure than those without technical knowledge who benefit from the bliss of 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 than those with lower knowledge.

We asked two preferences questions related to people's opinion about anonymity in the MTurk survey and examined the effect of bad experience and technical knowledge on these questions. Age, gender, and education were added into the model as control variables. Those who had bad experience are significantly less likely to think that it is possible to be completely anonymous, but they are more likely to agree that people should have the ability to be completely anonymous. People with more technical knowledge are also more likely to agree that people should have the ability to be anonymous. We noticed a significant interaction effect between technical knowledge and bad experience on whether or not they think 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 2, Among people with higher technical knowledge, those have no prior bad experience seem to be more confident and think they can be anonymous than those who have bad experience (45% said yes vs. 16% said yes, t [93] = 3.18, $p < .01$). For nontechnical people, bad experience has no significant impact on their perceptions (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***
Have bad experience	**-.23****	**-.17***
Technical knowledge × have 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 8. Logistic regression examining factors that predict policy preferences. Data shown in this table is from the U.S. MTurk survey (N = 182). Those who answered "not sure" were treated as missing values.

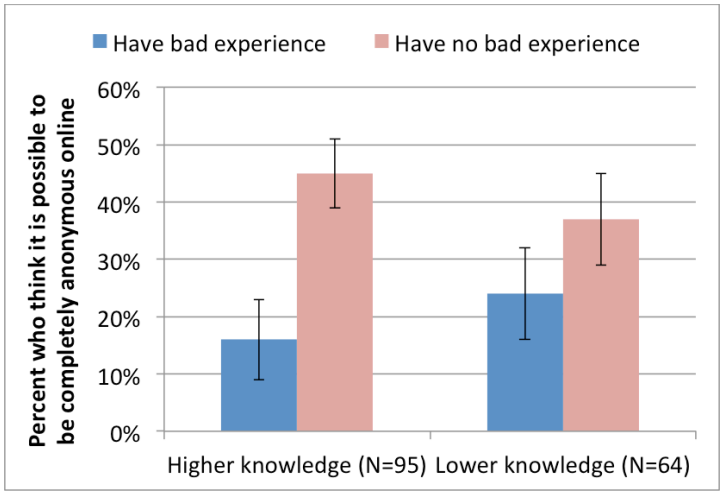


Figure 2. The Interaction effect of knowledge and bad experience on perception of whether or not anonymity is possible.

4.3 Summary

Overall, these findings provide direct empirical evidence to show the impact of individual differences and experience on how people identify and act upon different sources of privacy threats online. Main takeaways of the findings are:

- The majority of the Internet users have tried to hide some of their online footprints. A small proportion of them have explicitly tried to hide their identity.

- **The effect of demographics:** Younger people and social media users are more likely to hide from almost all threats than their counterparts. They are also more likely to use all kinds of strategies to practice hiding (supported by Pew survey).
- **The effect of technical knowledge:** Those with more technical knowledge about the Internet are more likely to hide their identity, hide from informational threats and hide from more number of threats. They are also more equipped to use technical methods to hide their traces. (supported by MTurk survey)
- **The effect of bad experience:** Pew respondents who have bad experience are more likely to hide their online interaction from all threats, and more likely to report using all kinds of strategies. MTurk respondents who had bad experience are more likely to hide from social threats and more number of threats, and they are more likely to edit their online content or use technical methods. Bad experience also elevated people's concern and made them feel less anonymous online, and this effect was more pronounced for technical people. We found that people with higher technical knowledge who did not have any prior bad experience on the Internet were more confident, and think it is more possible to achieve anonymity than technical people who had prior bad experience. They probably think the strategies they take are effective in achieving anonymity unless they have experienced a bad incidence.
- **The effect of social orientation:** Among the three social orientations measures, low collective identity and high segmented identity orientations are associated with hiding identity and hiding from family and friends, and editing their online content. In addition, high segmented identity is also associated with hiding from employers, and hiding from more number of threats in general. It also predicts using technical methods. But for other threats and threats mitigation methods, social orientation did not show significant effect. (supported by MTurk survey)

These two surveys mainly examine people's practices hiding their online information. In the next Chapter, I look more specifically into how users' understanding of the Internet can shape their use of the Internet and their privacy perceptions (Research Question 2).

Chapter 5. Users' mental model of the Internet

5.1 Introduction

"The Internet is a series of tubes!" – Former U.S. Senator Ted Stevens, 2006
"*The Internet is a big cloud you connect to."* – Expert reviewer #02 commented on a participant's drawing in the mental model study

Many people use the Internet every day yet know little about how it really works. Prior literature and many stories in the news suggest that people put their security and privacy at risk in part because they do not understand why their actions are risky. From the survey results shown in Chapter 4, we learned that people with more advanced knowledge about the Internet are more likely to hide their digital footprints and use more technical methods to protect their information than those with less knowledge. However, we measured technical knowledge by a number of true or false questions in that study, which capture factual knowledge, but not their overall understanding of the Internet and could not reveal more nuanced differences in people's understanding of the Internet. We do not know what exactly leads to the different behavior or perceptions. For example, are technical users more aware of the potential privacy and security risks than non-technical users because their more knowledge of the Internet structure? Or is it simply because technical users are more knowledgeable of more sophisticated strategies to protect their privacy?

In order to design better privacy and security technologies and educational programs for end users, we need to first understand what people know about the Internet. Towards these goals, we conducted a qualitative study asking users to describe and explain how the Internet works, both in general and while they did different common, Internet-based tasks (including watching a YouTube video, sending an email, making a payment online, receiving an online advertisement and browsing a webpage). We sampled 10 technical (with computer related college majors) and 11 non-technical users and identified patterns in their conceptual models of the network and awareness of security and privacy issues surrounding it. All participants were recruited through flyers, personal contacts and an online participant pool in Pittsburgh area. We then invited 5 networking and computer security experts to comment on users' mental models. These experts helped us identify common problems in user mental models of the Internet and the potential consequences of different models for Internet related behavior and decision-making.

Mental models

Researchers recommend diagramming exercises as a good way of capturing mental models (Jonassen 2008), and this method has been adopted to understand users' perceptions of the Internet. Poole et al (2008) used a sketching task in order to understand laypersons' knowledge of home networks. Their results suggests 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 to study how users understand and react towards Wi-Fi. Their study reveals that users have incomplete understandings of how Wi-Fi works and do not protect themselves against threats; examples include poor understanding of malicious access points and 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 an inaccurate or incomplete mental model may indicate a lack of awareness of the security risks surrounding Internet activities. Some prior work specifically examined users' perception of security systems. Wash (2010) interviewed people about how they understand security threats to their home computer, and summarized eight folk models about home computer security. 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 rely 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 and found differences in users' perceptions of firewall operation, network location, and connection ranging from incorrect, incomplete, to complete models.

Previous work on Internet mental models provides some insight into the nature of most users' understanding. For the most part, it seems mental models of the Internet are sparse and often shaped by interface cues. Much of this work, however, is task specific and dates prior to the current state of the Internet. In addition, it does not consider differences between more technically advanced and nontechnical users.

Technical vs. nontechnical users

Many studies show that more technically advanced users have a different understanding of the Internet and computer systems than more naive users. Bravo-Lillo and colleagues (2011) compared advanced and novice users' differences in their mental models about computer security warnings, finding that advanced users have much more complex models than novice users. They found several misconceptions in novice users' understandings, such as blindly assuming a trusted organization (e.g., bank) to be safe regardless of the warning message. 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. Similarly, Zhang-Kennedy et al. (2013)'s study found that correct understanding of a system can guide more secure behavior. Their study showed users had a limited understanding of passwords and did not fully understand how password attacks worked. They found users created stronger passwords after using educational infographics about how password attacks work.

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. Silver and Pfeffer (2004) compared expert and novice's conceptualization of a complex system and found that novice's understanding focuses more on

"perceptually available" components, whereas experts mention more "functional and behavioral" components. A few other studies (Resnick and Wilensky, 1998; Jacobson, 2001) found that people often assume centralized control and single causality, especially domain novices, while experts think about decentralized control and multiple causes when asked to comprehend a complex system.

5.2 Results

Our empirical data revealed a set of key differences in technical and non-technical participants' conceptions of the Internet and its inner workings. In this section I describe their perception differences in two dimensions and discuss experts reviews related to each dimension: 1) structure of the Internet, and 2) awareness of the components involved.

5.2.1 View of the Internet structure: Black box, simple chain, or complex system

Participant models varied in their representation of the Internet as simplistic (a black box – the "Internet" in Figure 3; a simple chain – Figure 4) or a highly articulated, complex system (Figure 4). Predominantly, non-technical users (8 out of 11) tended to represent the Internet as either a black box, where data came from a cloud, main computer, abstract entity ('the Internet'), or as a simple chain of access points. On the other hand, technical users had a more articulated model of the Internet as a complex system with varied hardware components and a more involved set of connections among components. Most non-technical participants connected endpoints directly (such as Figure 3), but technical participants drew more entities (either routers, ISPs) between endpoints (such as Figure 5).

Perspective: Egocentric vs. Holistic view of the network

The overall structure of participants' models fell into two categories: a holistic (13 participants - 8 technical and 5 non-technical) model and an egocentric model (8 participants - 2 technical and 6 non-technical). Participants who perceived the Internet holistically (13 out of 21) drew and explained the network as a system or web, with multiple connections, sometimes with a main center. Each drawing was comprised of multiple devices, servers, and connections between them (Figure 5 as an example). Overall, most holistic participants sketched web-like networks, connecting abstracted devices (computers, cell phones, servers, etc.). Crucially, participants who drew holistic models also recognized the presence of others within the system. Participants who perceived the Internet egocentrically (8 out of 21) conceptualized the system as branching out from their own machine. They drew and explained a direct, step-by-step process of how data was transmitted from the Internet to their own computer (Figure 3 and Figure 4). Some participants used physical items such as interfaces, computers, routers, wireless signals, and cables to explain how their computer connected to the Internet to gather information. All participants with egocentric models had idealized illustrations and explanations that abstracted away most detail and infrastructure, and did not envision the Internet as a network. Egocentric models tended to lack awareness of other parties.

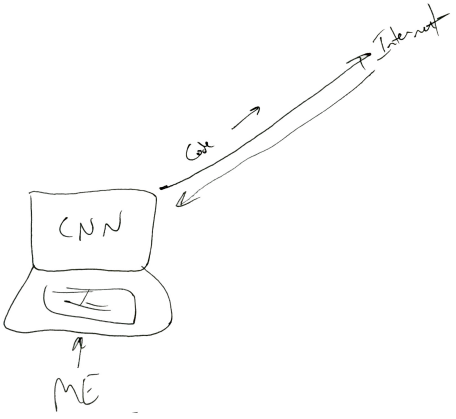


Figure 3. N06 – Egocentric view of own computer connecting to the Internet- considered a black box.

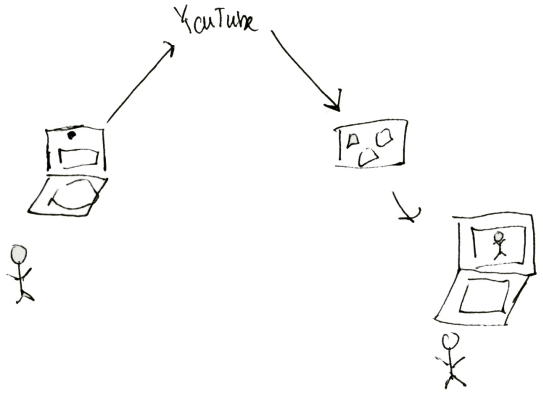


Figure 4. Drawing of watching a YouTube video. (N04, nontechnical participant)

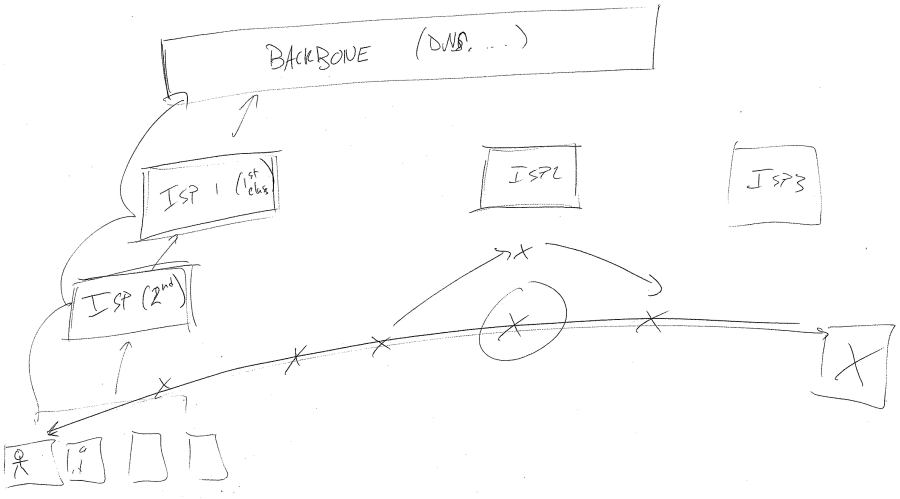


Figure 5. Drawing that reflects multiple layers of the network (T06, technical participant)

Expert evaluation: Misrepresenting structure

One major mistake expert reviewers pointed out was the inaccuracy in representing the structure of the Internet. Participants with structural inaccuracies often had an egocentric view of the Internet, and represented the structure as a simple chain or black box. Our experts highlighted entity awareness as a distinguishing factor of more the more accurate models, with E02 contrasting these two representations: one where *"the internet is composed of multiple, separate entities such as internet service providers,"* and another where "*the internet is a big cloud you connect to."*

Our experts noted that a more abstract and simplistic representation could indicate a lack of understanding depending on what elements they prioritized. When commenting on T10's drawings (Figure 6), E01 noted how "*the high-level abstraction is probably reasonable enough and [indicates] a high-level understanding of the network. But it doesn't suggest that they know any detail anywhere."* Furthermore, E01 noted how an inappropriate level of abstraction can actually "invalidate" a model beyond a certain

level; in reference to T10's explanation of cellular networking, he explained that *"[T10's explanation] actually suggests that the level of detail below is probably wrong"* due to the presence of several technical misconceptions at the high level.

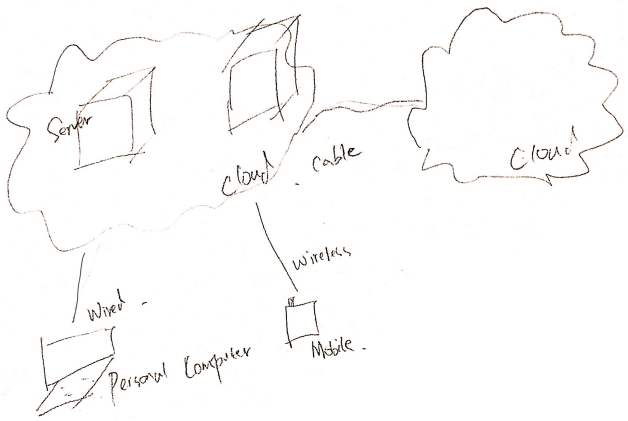


Figure 6. T10's drawing of how his personal computer and mobile phone connect to the Internet.

5.2.2 Infrastructural awareness

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, infrastructure hardware, and companies spoke to a participant's knowledge and understanding of the Internet as a complex system. Most nontechnical participants described one layer of the network, whereas some technical participants were aware of multiple layers of the network (Figure 4). 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 "US" server, or a university as a physical entity).

Most technical participants (8 out of 10) had broader awareness of entities and organizations involved in the Internet. For example, 6 technical participants noted there were many different ISPs. Furthermore, some concepts were only known by technically advanced users. Five technical participants mentioned network protocols such as "TCP/IP", "SMTP", or "IMAP", but none of the nontechnical participants mentioned these concepts. Technical participants were also more likely to mention logical elements such as "routing" or "packets". In contrast, most nontechnical participants only had awareness of organizations and services they interacted with directly without any mention of the underlying infrastructures, such as website addresses or popular websites (such as Facebook or Google). 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 of our non-technical participants were aware of physical objects that helped them connect to the Internet, such as drawing of a router; two nontechnical participants also drew mobile towers when describing a cellular network. Many non-technical participants mentioned "servers" (6 out of 11), but could not describe their role beyond data storage. Sometimes they used incorrect terminology or metaphors to describe technical

components such as "library" to describe a database (N03) or "code" to describe data being transmitted (N06). A few non-technical participants (4 out of 11) were aware of ISPs, but only the ISPs they interacted with directly to purchase Internet service, and they referred to them using the company names such as "Verizon" (N10) or "AT&T" (N05).

Awareness of data access and the protection strategies people take

After each participant described how they think the Internet works and how they complete a task on the Internet, the interviewer asked: "Is there any people, organizations or companies can see your connections and activities?" After they answered that question, the interviewer followed up by asking: "Did you do anything to prevent any others from seeing your connections or activities?" Because we did not have recordings for all participants, I show data from 7 nontechnical participants and 8 technical participants in Table 9.

Both nontechnical and technical participants were aware of companies like Facebook and Google or websites they visit having access to their data. They also had knowledge of advertising companies, data analytics companies, and companies "working with" big companies like Google or Facebook, which they were presumably referring to data providers or advertisers. Participants' responses suggested they learned about ad serving when things they had viewed in the past appeared in unexpected places during other Internet use. When probed to describe how advertisements were delivered, they recalled recent experiences browsing the web and then visiting a site such as Facebook, and seeing an ad derived from their recent browsing activity (retargeted ads). Although many participants were aware of third parties serving ads, they were not able to correctly explain how ads were generated using their data. They were aware, however, of the partnership between different organizations. For example, N11 mentioned the *"paid relationship between Google and Amazon."* These users had awareness that their data was being tracked, sold, and processed to generate advertisements but could not describe the algorithm or other technical aspects of the process. Only a few nontechnical participants mentioned other parties like government, ISP and hackers. In contrast, more technical participants mentioned hackers or "man-in-the-middle", and other people sharing the same network.

In terms of the strategies they take, we did not see too much difference between the two types of participants. Most of our participants, regardless of their technical knowledge, did not bother to use more sophisticated methods to prevent others from seeing their data because they *"haven't had the need to do that."* (T04) Both groups reported using behavioral methods such as avoiding logging in or using fake account, and technical methods such as deleting cookies and using private browsing. One difference we see in technical participants is that more of them mentioned paying attention to https and "lock icon" that are associated with secure browsing, while only one nontechnical participant mentioned using this precaution strategy. In general, most of our participants did not seem to be concerned about data access: one participant mentioned, *"I don't care who see and read my email"* although he was aware that *"hackers can act as mail servers*" (T10).

ID	Perception of who have access	Protection strategies
N01	Youtube, ads companies, Gmail, and others who Google allows access	No
N06	Websites tracking ads, Amazon and other companies	No
N07	Youtube, other companies, Google, Facebook, and companies working with them	Don't post with my name, don't have an account, use https
N08	Google ads companies, Google analytics; ISP; admin of home network	Use passwords to secure my home network
N09	Youtube, Facebook	Delete cookies; change privacy settings
N10	ISP or some specific websites; NSA share information with Google	No
N11	YouTube, Google, ads companies, analytics companies; government; hackers	A Google app to stop companies tracking me
T01	Websites I visit, third party; other people sharing the same wifi	No
T04	Service providers; NSA; anyone on the same network	Avoid using banking website in public wifi, use https
T05	Youtube, Google, Amazon; government	No
T06	Google, Facebook, any website you log on to; NSA; other people in the network; man-in-the-middle	Delete Facebook account, use fake Youtube account
T07	Google, people who own the browser	Kill cache, use private browsing, use https
T08	Remote server (companies'), plugins that track your request; man-in-the-middle	Stop using plugins, use encryption/ssl/https, not log in
T09	Google or Facebook; ISPs and a bunch other stuff; other people in the same network; man-in-the-middle	Use https and watch for lock icon
T10	Google, Youtube, corporations who provide mail service, traffic analytics tools; other ISPs; government; hacker	Disable cookies, not log in

Table 9. Participants' perception of who have access to their internet connection and activities and the strategies they take to prevent their data from being seen

Expert evaluation: Entity omissions

Experts identified entity omission as an important mistake that technical and nontechnical participants made in their drawings. They conjectured that participants in many cases were surfacing entities they had interacted with, abstracting the rest into a 'black box.' Four out of the five experts noted other entity omissions: not representing second level ISPs or advertisers, a lack of protocol awareness, or a simple lack of awareness regarding other network layers. Our experts also noted that many technical participants did not adequately represent ISP interconnections and the concept of different "tier" ISPs. Experts suggested this misconception might increase security and privacy risks, as users

who lack awareness of entities or organizations involved may have no idea about whom to blame when attacks, leaks, or other security issues occur.

Expert reviewers noted that these privacy and security risks were perhaps most problematic for non-technical users, whose views of data access and privacy online was more limited and vague. As one expert reviewer noted, the 'black box' view may engender too much trust in the system:

"*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.*" *(E02)*

5.3 Summary

The findings of this study help us breakdown the differences between technical and nontechnical users in terms of how they conceptualize the process of connecting to the Internet and how information is delivered online. Major gaps in users' mental models include the absence of Internet organizations and entities involved, and inaccurate structures connecting Internet components or layers. Non-technical users had more simplistic models and limited awareness of the technical or organizational complexities of the network. These differences could potentially explain why technical people are more likely to hide from informational threats (survey results in Chapter 4), and hide from more threats – one reason is that they are more aware of the entities that could have access to their data. When asked about what strategies they take to prevent their data from being seen by others, the interviewees in this study, however, did not report using too sophisticated strategies and we did not see a significant difference between nontechnical and technical users. Most people answered they "don't care" or "haven't had the need" to do so.

Technical knowledge can lead to heightened awareness of who can access their information, but it does not seem to increase people's threat perception. In other words, knowing the Internet structure may not significantly increase their perceived probability of a privacy invasion or a security breach happening to them. People could still suffer from the optimism bias which inhibits them from taking actions to hide their identity or information. The survey results in Chapter 4 show that previous negative experience is also associated with more hiding. Taken together, it is reasonable to hypothesize that experiencing a negative event may increase people's perception of the probability of privacy risks (they think it's less possible to be anonymous online), and for those who are equipped with technical knowledge, they are more aware of where the threats come from, and are therefore likely to take actions to protect their information from those specific threats. In the next chapter I will propose an experiment to examine this hypothesis.

Chapter 6. Proposed work: The effect of privacy threat and Internet knowledge on users' decisions to adopt privacy strategies

6.1 Introduction

Our results from the mental model study raise an important question about how much users need to know about how the Internet works. In particular, do users need to have more complex articulated models like what the technical participants have in our sample? Do they need better awareness of the entities involved in delivering and receiving a piece of online content? An appropriate mental model of the Internet can guide people towards better troubleshooting breakdowns, security and privacy management, and policy decision-making. Ordinary Internet users may not need to completely understand the technical details of how Internet works in order to surf the web, but when problem occurs (e.g., password gets compromised by a malicious website, or personal information gets leaked to unintended third parties), the knowledge of how their information get transmitted over the Internet could help people pinpoint the problem, and better understand what strategies they can use to protect themselves. The goal of this proposed study is to investigate how we can help people deal with different privacy threat by updating their mental model of the Internet, and how different models affect people's decision to hide their online information. I will first define the scope of the Internet addressed in this proposal. We will focus on the Internet components that are associated with other people, organizations or institutions (e.g., employer or company as internet service provider) and the connections between them. Users' understandings of other components of the Internet such as physical layers, Internet protocols are not discussed in this study.

In our interview study in Chapter 3, most interviewees reported hiding their online information from a specific source of threat, reflecting a personal "threat model" of persons or organizations. People's past negative experience on the Internet seems to influence their perception of this personal "threat model". From the survey results in Chapter 4, we learned that both technical knowledge and prior negative experience were associated with more hiding of their personal information online. To study the effect of privacy threat and Internet knowledge on people's privacy behavior, we can draw from the fear appeal literature. Using fear appeals to persuade people into adopting certain behavior by eliciting a sense of threat in a message has been applied in many domains, especially in communicating health-related information (Witte, 1994). It has also been used in the area of information security (LaRose et al 2008), such as a study about spyware (Johnston and Warkentin, 2010) and a study about password (Vance et al 2013). According to Witte (1994), 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 is infected by virus") and their own susceptibility to the danger (e.g., "It is possible that my computer will be infected"). If they perceive the threat to be high, they will execute the second step – evaluating the efficacy to avert the threat, which includes their self-efficacy (e.g., "I can use anti-virus software to protect my computer") and the response efficacy (e.g., "I believe anti-virus software can protect my computer"). When threat and efficacy are both high, people would activate a danger

control process -- accept the message and perform actions to mitigate the threat. When threat is high but efficacy is low, they will instead activate a fear control process -- reject the message and avoid future information about the threat. To understand the effect of fear on people's perception, our first research question is:

RQ1. How does being exposed to a negative Internet event influence people's perception of threat from the source of the negative event?

Experiencing negative events, or being exposed to examples about privacy invasion may elevate people's perception of the threat, but do not increase their efficacy. The literature on fear appeals suggests that making people aware of danger is only one part of getting them to take action. They also have to feel they can do something that will work. This feeling of "efficacy" may derive from their understanding of the Internet.

The interview study in Chapter 5 shows that currently most lay people comprehend the Internet in a more abstract way and cannot spell out the entities involved in the process of delivering or receiving a piece of content from the Internet. In his seminal book "The Design of Everyday Things", Don Norman (1988) has emphasized the importance of mapping user's mental model of how a system works in the design process. Users develop their mental models of the system by directly interacting with the system. Now the time it takes to connect to the Internet is almost negligible. The webpage users want to access usually shows up immediately after they type in the URL in the browser window, therefore it is not surprising that lay people form an abstract mental model as in Figure 3 and Figure 4. Sometimes people gain more awareness if they have had interactions with the service providers (e.g., they will know Comcast is involved in connecting them to the Internet), or if they have seen online advertisement tailored to their demographics and interests (e.g., they will know advertisers may have access to those personal information by working with other companies who have their data). Some other tools help increase user awareness of third parties who may have access to their information, such as the Ghostery app which detects and stops third party tracking when users access a webpage (https://www.ghostery.com/).

Different from novice users whose mental models are mainly trained by external tools or interfaces, technical users' mental models are probably gained from classes they took. Those models are more accurate, contain more details of the structure and have more awareness of the entities involved in the process. However, it is unclear how these different understandings of the Internet affect with people's attitudes and behaviors managing their online information. As a result, system designers need to answer the question of how much detail we should inform users about the Internet. Some researcher argues for greater transparency in telling users what others can do to their data to avoid unintended information disclosure (Solove, 2007), but others have suggested potential tradeoffs of the increased transparency (Stuart, et al 2012). The principle suggested by Norman says designers need to make system status visible, provide feedback, and map user control with the functions controlled (Norman, 1988). However when the system is too complex, greater visibility could cause information overload, increase the feeling of uncertainty, and can sometimes backfire. The second research question we hope to examine in this study is:

RQ 2: What is the effect of showing people a complex or a simple model of the Internet on how they manage their online information?

Privacy and security researchers have explored ways to provide greater transparency by increasing user awareness of information leakage over the Internet. Kowitz and Cranor (2005) proposed a large peripheral display to inform users of personal information leaks on wireless network. On a display installed in a public workplace for two weeks, they showed words from users' chat message or web searches that can be intercepted from unencrypted communications transmitted within the same network. They did not find significant differences in people's privacy perceptions before and after seeing the display, but found qualitative evidence that it made people feel 'less private' about their online chats and searches. Balebako et al (2013) conducted a qualitative study examining users misconceptions of what information is collected and shared by smartphones applications, and designed interfaces to inform users of the data leakage by two popular smartphone games. Some of their participants were not too concerned about privacy after using the interface because they considered data sharing as a tradeoff for free games, whereas some others expressed strong negative sentiment after learning about the unexpected data sharing. Wang et al (2014) suggested two design manipulations on Facebook trying to nudge people toward more careful disclosure on Facebook. One of their manipulations is to show users the audience of their posts by displaying five profile pictures of those who can see their posts. Their evaluation shows that the privacy nudges effectively attracted users' attention, but they did not find significant changes in people's behavior (e.g., edit privacy setting) after seeing the nudges.

None of this work has provided direct evidence that increased transparency leads to more secure behavior online. In our study, most technical participants have a more accurate and complete picture of the Internet, but it did not seem to motivate them to take more secure strategies to protect their information. The reason they stated was "I don't care" or "I don't have the need to do so". It seems like a threat must be present in order to motivate people to take action, and a more detailed model of the Internet can help people better identify where the threat is. As LaRose et al (2008) proposed, efficacy may interact with the perception of threat. We also hypothesize an interaction effect of their understanding and the threat stimuli. Taken together, we hypothesize:

> *H1. People who are exposed to a negative Internet event will have a higher perception of threat than those who are not.*
>
> *H2. People who are shown a detailed model of how their information is transmitted over the Internet will perceive a higher level of efficacy than those who receive an abstract model.*
>
> *H3. When a negative event is present, people's perception of threat will be higher when they are presented with a detailed model of how their information is transmitted over the Internet than when they are presented with an abstract model.*

H4. People will be more likely to hide their information when their perception of threat and perception of efficacy are both high.

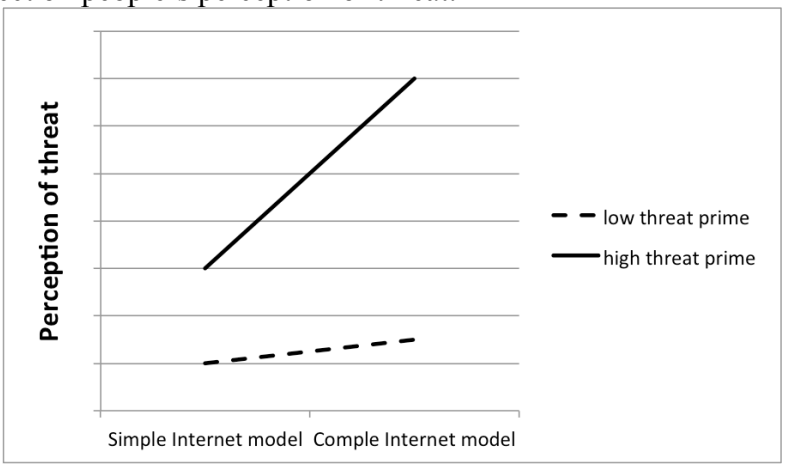
6.2 Study plan

To investigate the above hypotheses, I propose a 2 × 2 experimental design (Table 9). I plan to collect responses from two platforms: Amazon MTurk (U.S. only) and CBDR online participant pool. The purpose is to get respondents with diverse background and technical expertise.

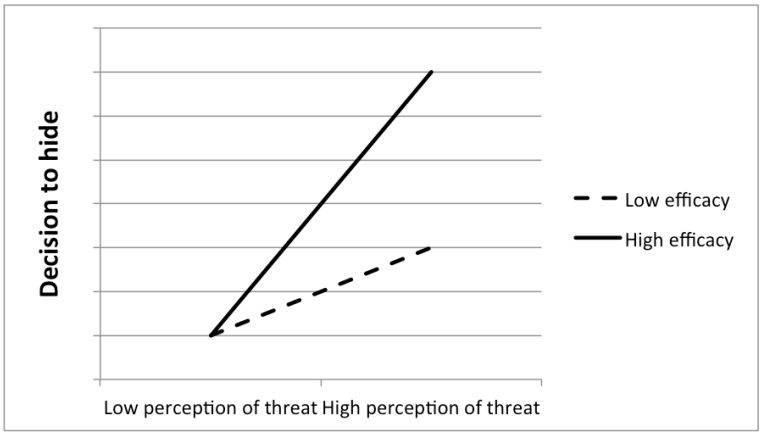
Conditions	Manipulations	
	Threat	Internet model
1	Low	Simple
2	Low	Complex
3	High	Simple
4	High	Complex

Table 10. Experiment conditions

According to my hypotheses, the threat prime and model manipulation will have an interaction effect on people's perception of threat:

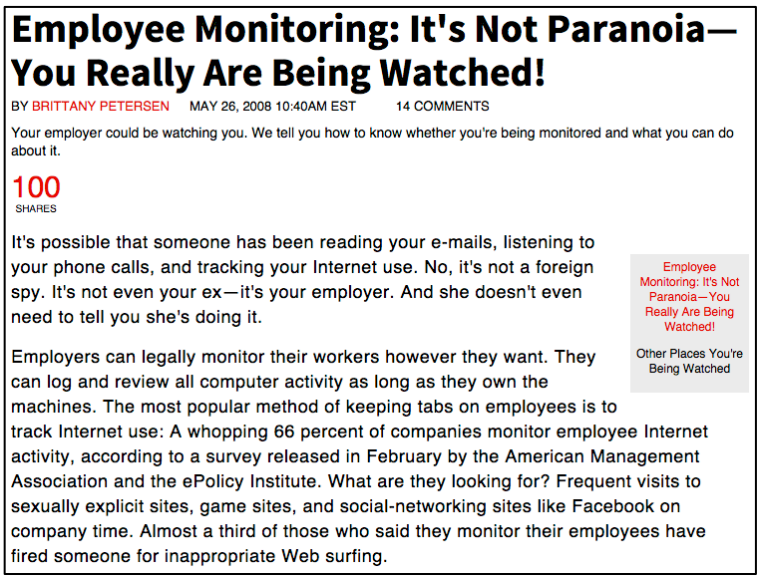


The Internet model manipulation will directly influence people's perception of efficacy, then the perception of efficacy and the perception of threat will influence the dependent measure: people's decision to hide their information:



Manipulations

To manipulate the threat, I will show half of the participants a news article that describes personal information being accessed or monitored by other people or organizations. To examine the effect of social and information privacy threat, I will select several articles emphasizing threats from employers, family and friends, government, Internet service providers and other third parties. They will be instructed to read and summarize a news article in the first task session. An example is shown below. The other half of the participants will be asked to summarize a news article irrelevant to Internet or privacy.



Employee Monitoring: It's Not Paranoia—You Really Are Being Watched!

BY BRITTANY PETERSEN MAY 26, 2008 10:40AM EST 14 COMMENTS

Your employer could be watching you. We tell you how to know whether you're being monitored and what you can do about it.

100
SHARES

It's possible that someone has been reading your e-mails, listening to your phone calls, and tracking your Internet use. No, it's not a foreign spy. It's not even your ex—it's your employer. And she doesn't even need to tell you she's doing it.

Employee Monitoring: It's Not Paranoia—You Really Are Being Watched!

Other Places You're Being Watched

Employers can legally monitor their workers however they want. They can log and review all computer activity as long as they own the machines. The most popular method of keeping tabs on employees is to track Internet use: A whopping 66 percent of companies monitor employee Internet activity, according to a survey released in February by the American Management Association and the ePolicy Institute. What are they looking for? Frequent visits to sexually explicit sites, game sites, and social-networking sites like Facebook on company time. Almost a third of those who said they monitor their employees have fired someone for inappropriate Web surfing.

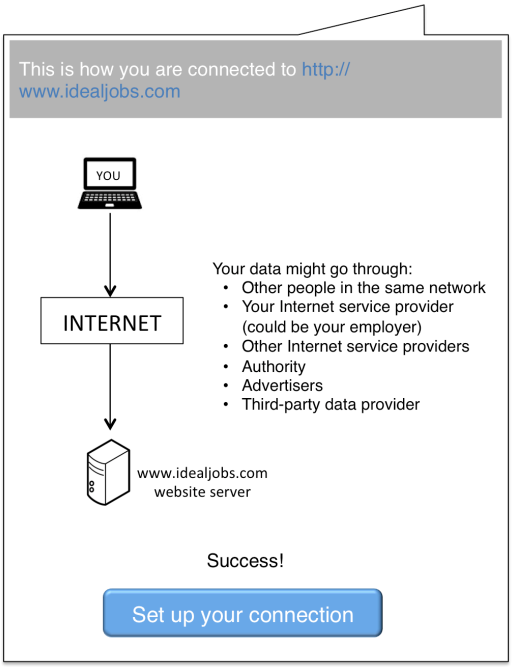
Figure 7. An example of a news article

To manipulate the Internet model, I will show each participant one of the models randomly selected from Figure 8. Participants will be instructed to help evaluate a new browser plug-in. They will be given a scenario to imagine themselves using the Internet for. An example scenario is:

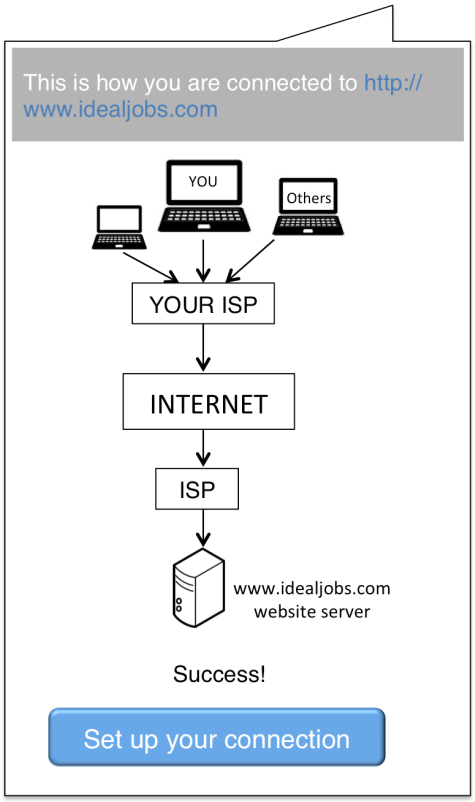
You are not satisfied with your current job and want to change jobs. You need to search for other jobs online. Your home Internet broke down so you have to use company Internet to do the searches. You found a new website for job search: www.idealjobs.com During registration, it asks for some personal information such as your age, gender, current occupation, current financial status. You have to fill out the registration form first to see available jobs.

The browser plug-in tells users how they are connected to the webpage through the Internet. Because the previous interview study revealed that a more detailed mental model have two main features: complex structure and awareness of entities involved, I plan to test the effect of these two features by manipulating four different models as shown in Figure 8.

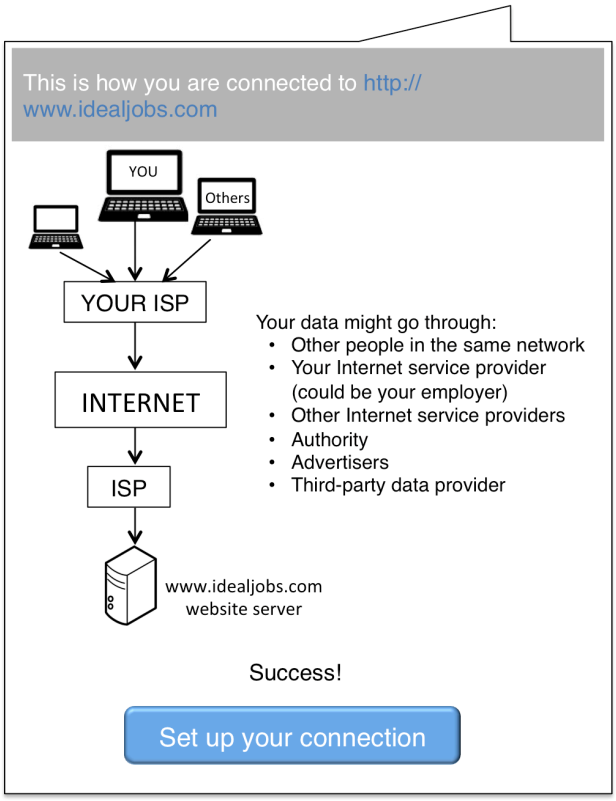
a. Simple model



b. Simple model with a list of entities



c. complex model



d. complex model with a list of entities

Figure 8. Simple and complex models of the Internet

Dependent variables

In 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 the proposed study, I plan to measure their behavior by measuring how participants set up their connection in this evaluation. Participants will be told that we are testing some new function to be imbedded in the browser. The tradeoffs of most privacy protection tools are the cost of time and effort. I plan to simulate the time cost in this experiment. Participants will be told to imagine using this setup as how they will actually use it in their browser. They can choose to use a proxy server to hide their identity or keep the original connection. The time tradeoff is the extra three steps to set up the proxy server.

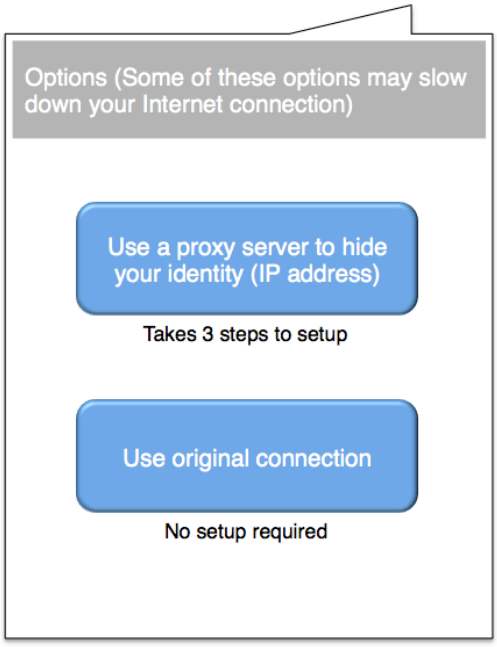


Figure 9. Measure of decision to hide

In addition, we will measure people's tendency to disclose personal information as an additional dependent variable. Participants will be asked to fill out the registration form for www.idealjobs.com using the previous setup. If they do not want to disclose, they can select "decline to answer".

The disclosure items include both nonsensitive and sensitive information:

- Age, gender, address, phone number, current occupation, current yearly income, marital status, personal debt, Facebook account, twitter account.

The other two dependent variables will be measured by survey questions:

Perception of threat (Strongly disagree/Disgree/Neutral/Agree/Strongly agree)

- If my Internet activity is monitored by my employer, it would be a serious threat.
- It is likely that my Internet activity will be monitored by my employer.

Perception of efficacy (Strongly disagree/Disgree/Neutral/Agree/Strongly agree/not sure)

- I understand how my employer could monitor my Internet activity.
- I know how to hide my Internet activity from my employer.
- Using a proxy server can hide my Internet activity from my employer.

Other measures

I will also ask these questions in the post-test survey.

- Match with users' own mental model:
  - Do you find any of the information provided by the browser plug-in surprising?

- Manipulation check:
  - Which of the following article have you reviewed yesterday?

- General privacy concern
  - How concerned are you about each of the following people or organizations seeing your information on the Internet?
    - Your Internet service provider
    - Your employer or supervisor
    - Your family, friends, or people who know you in real life
    - Government or authorities
    - Companies who run the website that you are accessing or the app that you are using
    - Other people who use the same network
    - Hackers and criminals
    - Advertisers
  - How likely do you think each of the following people or organization can access your information on the Internet?
  - How much do you feel in control of your information on the Internet?

- Previous negative experience
- Knowledge test
- Demographic information

Procedure

We will split the threat and model manipulation into two tasks to minimize the experimenter demand. Participants who complete the first task will be given task two on the second day.

Task 1: review and summarize a news article.

- Pre-test survey (Internet experience, technical knowledge, self-efficacy)
- Review news article
- Summary of the article
- Multiple choice question (reading comprehension)

Task 2: evaluate a browser plug-in

- Be presented with the scenario
- Be presented with the browser plug-in screenshot
- Select the set-up
- Disclosure form
- Post-test survey (include perception of threat, perception of efficacy, and other measures)

6.3 Timeline

Jan – Feb, 2015: Revise experimental design and pilot test

March, 2015: Run experiment

April – May, 2015: Analyze data

June – July, 2015: Write up thesis

August, 2015: Defend thesis

References

Acquisti, A. and Gross, R. (2006). Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. *Privacy Enhancing Technologies*, 2006. 36-58.

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is privacy worth?. *The Journal of Legal Studies*, *42*(2), 249-274.

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

Altman, I. The environment and social behavior. Monterey, CA: Brooks/Cole (1975).

Angulo, J., Wästlund, E., & Högberg, J. (2014). What Would It Take for You to Tell Your Secrets to a Cloud?. In *Secure IT Systems* (pp. 129-145). Springer International Publishing.

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, *54*(15), 2787-2805.

Baumeister, R. F. (1982). A self-presentation view of social phenomena. *Psychological Bulletin*, 91, 3-26.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, *5*(4), 323.

Berendt, B., Günther, O., & Spiekermann, S. (2005). Privacy in e-commerce: stated preferences vs. actual behavior. *Communications of the ACM*, *48*(4), 101-106.

Bargh, J.A., McKenna, K.Y.A., and Fitzsimons, G.M. Can You See the Real Me ? Activation and Expression of the " True Self " on the Internet. *Journal of social issues* 58, 1 (2002), 33–48.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. Political Analysis, 20(3), 351-368.

Bernstein, M. S., Bakshy, E., Burke, M., & Karrer, B. Quantifying the invisible audience in social networks. In *Proc. of CHI 2013*, ACM (2013), 21-30.

Bravo-Lillo, C. C., Downs, L., & J Komanduri, S. (2011). Bridging the gap in computer security warnings: A mental model approach. *Security and privacy*, IEEE, 18-26.

Brewer, M., and Chen, Y. Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological review* 114, 1 (2007), 133-151.

Chen, K., & Rea, A. I. (2004). Protecting personal information online: A survey of user privacy concerns and control techniques. *Journal of Computer Information Systems*, *44*(4), 85-92.

Chow, C. W., Harrison, G. L., McKinnon, J. L., & Wu, A. Cultural influences on informal information sharing in Chinese and Anglo-American organizations: An exploratory study. *Accounting, organizations, and society.* 24(1999). 561-582.

Cranor, L. F. (2008). A Framework for Reasoning About the Human in the Loop.*UPSEC*, 8, 1-15.

Culnan, M. J. (1993). " How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS quarterly*, 341-363.

DiMicco, J. Millen, D. (2007). Identity management: multiple presentations of self in Facebook

Debatin, B., Lovejoy, J. P., Horn, A. K., & Hughes, B. N. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication,* 15, 1(2009), 83-108.

Das, S., and Kramer, A. Self-censorship on Facebook. In *Proc. of ICWSM 2013*, AAAI (2013), 120-127.

Das, S., Kim, T. H-J., Dabbish, L., and Hong, J. The Effect of Social Influence on Security Sensitivity. In *SOUPS 2014,* USENIX (2014), 143-157.

Edman, M., & Yener, B. On anonymity in an electronic society: A survey of anonymous communication systems. *ACM CSUR,* 42, 1 (2009), 5.

Farnham, S. & Churchill, E. Faceted identity, faceted lives: social and technical issues with being yourself online. In *Proc. of CSCW 2011*, ACM (2011), 359-368.

Hofstede, G. *Culture's consequences: International differences in work-related values*. Sage (1984).

Iachello, G., & Hong, J. End-user privacy in human-computer interaction. *Foundations and Trends in Human-Computer Interaction*, 1, 1 (2007), 1-137.

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

Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: an empirical study. *MIS quarterly*, *34*(3), 549-566.

Joinson, A., Reips, U., Buchanan, T., & Schofield, C. Privacy, trust, and self-disclosure online. *Human–Computer Interaction,* 25, 1 (2010), 1-24.

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., and Kiesler, S. Why do people seek anonymity on the Internet?: informing policy and design. In *Proc. of CHI 2013,* ACM (2013), 2657-2666.

Kang, R., Brown, S., Dabbish, L., & Kiesler, S. Privacy Attitudes of Mechanical Turk Workers and the US Public. In *Proc. of SOUPS 2014*,USENIX (2014),38-49.

Katikalapudi, R., Chellappan, S., Montgomery, F., Wunsch, D., & Lutzen, K. (2012). Associating Internet usage with depressive behavior among college students. *Technology and Society Magazine, IEEE*, *31*(4), 73-80.

King, J., Lampinen, A. and Amolen, A. Privacy: Is There an App for That? In *Proc. of SOUPS 2011*, ACM(2011).

Kim, T. H-J., Stuart, H. C., Hsiao, H-C., Lin, Y-H, Zhang, L., Dabbish, L., Kiesler, S. YourPassword: Applying feedback loops to improve security behavior of managing multiple passwords. *ACM ASIA CCS '14*. NY: ACM Press(2014).

Klasnja, P., Consolvo, S., Jung, J., Greenstein, B. M., LeGrand, L., Powledge, P., & Wetherall, D. When I am on Wi-Fi, I am fearless: privacy concerns & practices in everyday Wi-Fi use. In *Proc. of CHI 2009*, ACM (2009), 1993-2002.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proc. of CHI 2007*, ACM (2007), 905-914.

Lampinen, A., Lehtinen, V., A., L. and Tamminen, S. We're in It Together: Interpersonal Management of Disclosure in Social Network Services. In *Proc. of CHI 2011*, ACM (2011), 3217-3226.

LaRose, R., Rifon, N. J., & Enbody, R. (2008). Promoting personal responsibility for internet safety. *Communications of the ACM*, *51*(3), 71-76.

Leary, M. R., and Kowalski, R. M. Impression management: A literature review and two-component model. *Psychological bulletin*, 107, 1 (1990), 34-47.

Lin, J., Amini, Sh., Hong, J., Sadeh, N., Lindqvist, J., and Zhang, J. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proc. UbiComp 2012*. ACM,501-510.

Litt, E., & Hargittai, E. A bumpy ride on the information superhighway: Exploring turbulence online. *Computers in Human Behavior*, 36 (2014), 520-529.

Litt, E., Spottswood, E., Birnholtz, J., Hancock, J., Smith, M. E., & Reynolds, L. Awkward Encounters of an "Other" Kind: Collective Self-Presentation and Face Threat on Facebook. In *Proc. of CSCW* 2014, ACM (2014), 449-460.

Markus, H.R., and Kitayama, S. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review* 98, 2 (1991), 224-253.

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

Milne, G. R., Rohm, A. J., & Bahl, S. (2004). Consumers' protection of online privacy and identity. *Journal of Consumer Affairs*, *38*(2), 217-232.

Naylor, D, et, al. Experience with XIA: Architecting a more Trustworthy and Evolvable Internet, SIGCOMM CCR, ACM (2014), To appear.

Newman, M. W., Lauterbach, D., Munson, S. A., Resnick, P., & Morris, M. E. It's not that i don't have problems, i'm just not putting them on facebook: challenges and opportunities in using online social networks for health. In *Proc. of CSCW 2011*, ACM (2011), 341-350.

Nguyen, D. H., Kobsa, A., & Hayes, G. R. An empirical investigation of concerns of everyday tracking and recording technologies. In *Proc. of the 10th Ubicomp,* ACM (2008), 182-191.

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

Omarzu, J. (2000). A disclosure decision model: Determining how and when individuals will self-disclose. *Personality and Social Psychology Review*, *4*(2), 174-185.

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.

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

Petronio, S. *Boundaries of Privacy: Dialectics of Disclosure*. State University of New York Press, Albany, NY, 2002.

Poole, E. S., Chetty, M., Morgan, T., Grinter, R. E., & Edwards, W. K. Computer help at home: methods and motivations for informal technical support. In *Proc. of CHI 2009*, ACM (2009), 739-748.

Rader, E. Awareness of Behavioral Tracking and Information Privacy Concern in Facebook and Google. In *Proc. of SOUPS 2014*, USENIX (2014), 51-67.

Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., & Dabbish, L. Anonymity, Privacy, and Security Online. Pew Research Center (2013). http://www.pewinternet.org/2013/09/05/anonymity-privacy-and-security-online/

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

Schecter, 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 *Proc. IEEE Symposium on Security and Privacy (S&P)* (pp. 51-65).

Shay, R., Ion, I., Reeder, R. W., and Consolvo, S. "My religious aunt asked why I was trying to sell her viagra": Experiences with account hijacking. *Proc. of CHI 2014*, ACM (2014), 2657-2666.

Shklovski, I. A., Mainwaring, S. D., Skúladóttir, H. H., Borgthorsson, H. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proc. of CHI 2014*, ACM (2014), 2437-2356.

Smith, H. J., Dinev, T., & Xu, H. Information privacy research: an interdisciplinary review. *MIS quarterly*, 35, 4 (2011), 989-1016.

Smith, H. J., Milberg, S. J., & Burke, S. J. Information privacy: measuring individuals' concerns about organizational practices. *MIS quarterly*, (1996). 167-196.

Snyder, M., & Gangestad, S. On the nature of self-monitoring: matters of assessment, matters of validity. *Journal of personality and social psychology,* 51, 1(1986). 125-139.

Solove, D. J. (2007). "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego law review*, *44*, 745-772.

Stuart, H.C., Dabbish, L., Kiesler, S., Kinnaird, P., and Kang, R. Social transparency in networked information exchange: a theoretical framework. *Proc. of CSCW 2012*, ACM (2012), 451–460.

Stutzman, F. and Hartzog, W. Boundary regulation in social media. In *Proc. of CSCW 2012*, ACM (2012).

Suler, J.R. (2002). Identity Management in Cyberspace. Journal of Applied Psychoanalytic Studies, 4, 455-460

Triandis, H.C. The self and social behavior in differing cultural contexts. *Psychological review* 96, 3 (1989), 506-520.

Turner, E. C., & Dasgupta, S. (2003). Privacy on the Web: An examination of user concerns, technology, and implications for business organizations and individuals. *Information systems management*, *20*(1), 8-18.

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

Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proc. of SOUPS 2012*, ACM.

Vance, A., Eargle, D., Ouimet, K., & Straub, D. (2013, January). Enhancing Password Security through Interactive Fear Appeals: A Web-Based Field Experiment. In *System*

Sciences (HICSS), 2013 46th Hawaii International Conference on (pp. 2988-2997). IEEE.

Vaniea, K. E., Rader, E., & Wash, R. Betrayed by updates: how negative experiences affect future security. In Proc. of CHI 2014, ACM (2014), 2671-2674.

Vandello, J. A., & Cohen, D. Patterns of individualism and collectivism across the United States. *Journal of personality and social psychology* 77, 2(1999), 279-291.

Vitak, J., & Kim, J. You can't block people offline: examining how Facebook's affordances shape the disclosure process. In *Proc. of CSCW 2014*, ACM (2014), 461-474.

Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., & Cranor, L. F. (2011, July). I regretted the minute I pressed share: A qualitative study of regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security* (p. 10). ACM.

Weinstein, N. D. (1989). Optimistic biases about personal risks. Science, 246,1232-1233.

Witte, K. (1994). Fear control and danger control: A test of the extended parallel process model (EPPM). *Communications Monographs*, *61*(2), 113-134.

Woodruff, A. Necessary, unpleasant, and disempowering: reputation management in the internet age. In *Proc. of CHI 2014,* ACM (2014), 149-158.

Woodruff, A., Pihur, V., Consolvo, S., Schmidt, L., Brandimarte, L., & Acquisti, A. (2014, July). Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. In SOUPS 2014.

Xu, H., Dinev, T., Smith, H. & Hart, P. Information Privacy Concerns: Linking Individual Perceptions with Institutional Privacy Assurances. *JAIS*, 12, 12(2011), 798-824.

Zhang-Kennedy, L., Chiasson, S., & Biddle, R. (2013, September). Password advice shouldn't be boring: Visualizing password guessing attacks. In eCrime Researchers Summit (eCRS), 2013 (pp. 1-11). IEEE.

Appendix: Survey questions

Note: We only show the questions analyzed in this paper. 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

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 <u>people from your past, or people who might criticize, harass, or target you</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 11. Have you ever tried to use the Internet in such a way that <u>law enforcement, the government, or companies or people that might want payment for the files you download such as songs, movies, or games</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 12. Have you ever tried to use the Internet in such a way that <u>hackers, criminals, or advertisers</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.

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 in MTurk Survey]

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. □ True □ False □ I'm not sure

Incognito mode / private browsing mode in browsers prevents websites from collecting information about you.
Website cookies can store users' logins and passwords in your web browser.

No one, except for the sender and intended receiver, can reveal the content of an encrypted email.
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.

[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 (alpha = 0.74)
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 (alpha = 0.53)
I often do "my own thing".
I enjoy being unique and different from others in many ways.
Segmented identity (alpha = 0.76)
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)_____