"My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security

Ruogu Kang[1], Laura Dabbish[1,2], Nathaniel Fruchter[1], Sara Kiesler[1]

Human-Computer Interaction Institute[1], Heinz College[2]

Carnegie Mellon University

Pittsburgh, PA

{ruoguk, dabbish, nhf, kiesler}@andrew.cmu.edu

ABSTRACT

Many people use the Internet every day yet know little about how it really works. Prior literature diverges on how people's Internet knowledge affects their privacy and security decisions. We undertook a qualitative study to understand what people do and do not know about the Internet and how that knowledge affects their responses to privacy and security risks. Lay people, as compared to those with computer science or related backgrounds, had simpler mental models that omitted Internet levels, organizations, and entities. People with more articulated technical models perceived more privacy threats, possibly driven by their more accurate understanding of where specific risks could occur in the network. Despite these differences, we did not find a direct relationship between people's technical background and the actions they took to control their privacy or increase their security online. Consistent with other work on user knowledge and experience, our study suggests a greater emphasis on policies and systems that protect privacy and security without relying too much on users' security practices.

1. INTRODUCTION

Today, the Internet is a ubiquitous vehicle for information, communication, and data transportation, and central to many lives. Most who use the Internet, however, have a limited understanding of its technical underpinnings (e.g., [24,30]) and how their personal information is used and accessed online ([28], [39]). Here, we argue that we need to understand what people know and do not know about how the Internet works for two reasons. First, understanding how users think will enable us to design more effective privacy and security controls that match user perceptions. Second, understanding how users think the Internet works will help us develop more effective educational programs so that users, as citizens, can be better informed about privacy policies and other aspects of Internet governance. Towards these goals, we examined users' mental models of how the Internet works.

Part of the challenge in understanding the Internet is its rapid evolution. The Internet is now massive and embedded into many contexts. It connects billions of individuals around the world through many different types of devices [46]. Many entities are involved in transmitting data and tracking user behavior including third party caching services, first and second level ISPs, cellular network providers, web services, search engines, and ad networks. More personal data than ever is transmitted via the Internet as mobile access proliferates [9] and service providers expand their tracking, creating privacy and security challenges far beyond the ability of end users to manage [38]. Network security tools are not widely used and do not help users understand why or how well they work.

The Internet is not an automated device that works in a simple way to accomplish simple goals. Users have to make decisions that affect their privacy and security, ranging from whether to access public Wi-Fi at an airport to how to share a file with a colleague to constructing a new password for a shopping site. We don't know the influence of users' understanding of the Internet on their daily privacy and security practices on the Internet. Does technical knowledge about the Internet help people make good privacy-protecting decisions?

Some previous work has explored user mental models of networking, but has mainly focused on specific domains such as home networking [30,42] and wireless Internet access [24], or specific privacy mechanisms such as firewalls [32]. This previous work does not describe users' overall mental model of the Internet across domains and its implications for how they think about and take action to protect their privacy on the Internet.

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

2. RELATED WORK

Why do end users need to understand the Internet? Early literature [12] suggests that Internet literacy is associated with inequality and political participation, as well as economic, legal, and policy decisions. During everyday Internet use, most people may not need to understand technical details such as how a webpage got delivered to their desktop, how caching works, and where their data is being sent; however, when problems occur (e.g., the SSL Heartbleed bug [1] or Target data breach[2]), a better

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22-24, 2015, Ottawa, Canada.

understanding of how the system works can help users understand problems, picture the potential consequences to their personal privacy decisions, and protect themselves from future invasions, such as by obeying recommendations to increase password strength. A clear picture of how users think about the Internet can also help system designers develop technologies that meet users' expectations and help policy makers communicate in ways that are easily understood by lay people [34].

A commonly used method in psychology to elicit users' understanding about a problem is mental models, which are "psychological representations of real, hypothetical, or imaginary situations." [22] Mental models describe how a user thinks about a problem or system; it is the model in the person's mind of how things work. These models are used to make decisions by supporting mental simulation of the likely effect of an action [16]. Mental models of a system can be useful in informing interface design or educational materials because they suggest natural ways to visualize complex system components or user interactions with them.

2.1 Users' mental models of the Internet

A number of researchers have adopted the mental models approach to understand users' perceptions of the Internet [24,30] [30] and Internet-related systems or technologies, such as home computer security [42], firewalls [44], and web security [16].

Diagramming exercises are considered a good way of capturing mental models in addition to traditional verbal reports [22], and this method is frequently used in user-centered Internet research. Poole et al. [30] used a sketching task in order to understand laypersons' knowledge of home networks. Their results suggest that most users, even those who are technically sophisticated, have a poor understanding of home networking structures. Klasnja et al. [24] also used a diagraming task when studying how users understand and use Wi-Fi. Their study revealed that users had an incomplete understanding of important privacy risks when they were connected to Wi-Fi, such as malicious access points, and did not protect themselves against threats, such as seeking SSL encryption. Four out of the eleven participants they observed were aware that other people could possibly access their information being transmitted over Wi-Fi, but this understanding did not raise concerns.

Having a deficient mental model may indicate a lack of awareness of the security risks surrounding Internet activities. Some prior work specifically examined users' perceptions of security systems. Wash [42] interviewed people about how they understood security threats to their home computer and summarized different folk models about home computer security including models centered on viruses and models centered on hackers. Friedman et al. [16] also addressed security risks, interviewing 72 participants and asking them to do a drawing task to illustrate their understanding of web security. They found that the majority of participants relied on simple visual cues like the presence of HTTPS and a lock icon to identify secure connections. Raja et al. [44] studied users' mental models of personal firewalls on Windows Vista using a structured diagramming task. They gave participants images of a computer, firewall, and the Internet depicted as a cloud, and asked participants to connect those pictures with arrows. They then improved understanding of firewalls by showing participants an interface prototype with contextual information.

Many studies show that more technically advanced users have a different understanding of the Internet and computer systems compared to more novice users. Bravo-Lillo and colleagues [8] compared advanced and novice users' differences in their mental models about computer security warnings, finding that advanced users had much more complex models than novice users. Vaniea et al. [41] 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. [45]'s found that a 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. Hmelo-Silver and Pfeffer [19] compared experts' and novices' conceptualization of a complex system and found that novices' understanding focuses more on "perceptually available" (concrete) components, whereas experts mention more "functional and behavioral" (conceptually abstract) components. A few other studies [20,33] found that people often assume centralized control and single causality, especially domain novices, whereas experts think about decentralized control and multiple causes when asked to describe a complex system.

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

2.2 Users' knowledge of the Internet

Various attempts have been made to measure users' knowledge of the Internet. Page & Uncles [27] categorized Internet knowledge into two categories: the knowledge of facts, terms or attributes about the Internet (declarative knowledge), and the knowledge of how to take actions or complete tasks on the Internet (procedural knowledge). Following this argument, Potosky [31] developed an Internet knowledge measure (iKnow) that asks people to rate their agreement as to whether or not they understand terms related to the Internet (e.g., "I know what a browser is"), and whether or not they are able to perform Internet-related tasks (such as "I know how to create a website"). An important question researchers have asked is what impact these two kinds of knowledge have on user security and privacy behavior.

Park [28] measured user knowledge in three dimensions: technical familiarity, awareness of institutional practices, and policy understandings. He found higher user knowledge correlated with online privacy control behavior. Other studies emphasize the role of user skills. Das et al. [11] proposed three factors influence the adoption of security and privacy tools: awareness of security threats and tools, motivation to use security tools, and the knowledge of how to use security tools. Litt [25] found that higher

Internet skills were positively associated with more content generation online and managing one's online presence. boyd and Hargarttai [6] found that users with more Internet skills were more likely to modify their privacy settings on Facebook. Hargittai and Litt [18] developed a scale to specifically measure privacy-related skills. They asked people to evaluate their level of understanding of privacy-related Internet terms such as "privacy settings," "tagging," and email "bcc." Their survey showed that higher privacy-related knowledge was positively associated with better privacy management of social media profiles.

Having more declarative knowledge or skill has not always been shown to predict more secure online behaviors. Dommeyer and Gross [13] found that consumers are aware of privacy-protective strategies, but do not use them. In a study by Nguyen and colleagues [26], some participants expressed uncertainty about how store loyalty cards would be used, but they did not take any protective actions to protect their personal information. Furnell et al. [15] studied how people manage security threats to home PC systems and found advanced technical users did not use more effective security practices than novice users.

The Internet today is much different than what it was 10 years ago, so people may perceive or use it very differently today, especially in managing their privacy. In 2003, the majority of American Internet users expressed strong concern about information used by governments and corporations, but they had little knowledge of how their data flows among companies [39]. A more recent 2011 review of the literature suggests that people's awareness of organizations collecting their personal information increases their privacy concerns [35], but there remains little understanding of how people think the Internet works. In late 2014, Pew Research Center conducted a national U.S. sample survey to test Internet users' knowledge of the Web by asking 17 questions about Internet terms (e.g., "URL"), famous technology celebrities (e.g., identifying Bill Gates' photo), and the underlying structure of the Internet (e.g., explanation of Moore's law) [29]. Their survey indicated that the majority of Internet users recognize everyday Internet usage terms, but very few are familiar with the technical details of the Internet and most do not understand Internet-related policies.

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

3. METHOD

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

3.1 Participants

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

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

Table 1. Study participants (Total = 28; N = non-technical participants; C = community participants; T = technical participants; *T11 was recruited with the community sample).

Identifier	Gender	Age	Education background
Lay participants (N = 17)			
N01	M	19	Finance
N02	M	22	Finance
N03	M	22	Biomedical Engineering
N04	F	18	Geology
N05	F	22	English
N06	M	22	Law
N07	F	21	Cognitive science
N08	F	19	Statistics; psychology
N09	F	22	Legal studies
N10	M	30	Music; foreign languages
N11	F	18	Neuroscience
C01	M	64	Engineering; public health
C02	M	32	Culinary arts
C03	M	62	Communication arts; religion
C04	M	49	Psychology
C05	F	58	MBA
C06	F	30	Foreign policy
Technical participants (N = 11)			
T01	F	19	Computer science
T02	F	21	Computer science
T03	F	27	Computer science & HCI
T04	M	25	Information technology
T05	F	24	Electrical/CS engineering
T06	M	26	Computer science
T07	M	25	Information technology
T08	M	23	Computer science
T09	M	27	Software engineering
T10	M	24	Software engineering
T11*	M	32	Computer science

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

3.2 Procedure

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

After completing the survey, participants were guided through the main drawing tasks. Every participant was first prompted to explain how the Internet works, and asked to draw a general diagram of it in whatever form they chose on a large sheet of paper in front of them. Participants were instructed to verbalize their thought process as they drew, consistent with traditional think aloud protocols [14]. A video camera captured participants' drawings and voices. All recordings were labeled using anonymous identifiers. No personally identifiable information was collected or recorded.

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

After the drawing tasks, participants filled out a post-task survey with demographic questions, as well as a series of Internet knowledge questions (attached in Appendix C). The knowledge questions included self-rated familiarity with nine technical terms on a 5-point scale (IP address, cookie, encryption, proxy servers, SSL, Tor, VPNs, privacy settings, and private browsing), and seven true/false questions about security and privacy knowledge (e.g., "Tor can be used to hide the source of a network request from the destination.") We developed the knowledge questions by consulting domain experts in computer security and tested their reliability with two independent samples (see Appendix C for details).

3.3 Technical knowledge level

All 28 participants filled out the same post-task survey. Besides differences in academic background, technical participants performed significantly better than lay participants in both the self-rated familiarity questions (mean: technical = 3.59, lay = 2.47, t [26] = 4.32, $p < .001$) and correctness on the true/false questions (mean number correct: technical = 4.27, lay = 1.53, t [26] = 5.83, $p < .001$).

3.4 Data Analysis

We qualitatively analyzed participants' think aloud responses to identify key differences across mental models. We conducted our analysis iteratively, carrying out three rounds of data collection and subsequent analysis, allowing the first analysis process to guide our second round of data collection, and then the third. Our initial analysis occurred after the first 12 sessions with participants (predominantly non-technical participants). We focused on the diagrams they generated during our sessions as well as the video and audio recorded during our sessions. By comparing and contrasting across user models, we generated a set of codes that indicated dimensions on which the models varied. To verify and extend codes and themes identified in our first round of data analysis, we conducted a second round of analysis, extending codes identified in our first round based on new features of the second set of models. In the last round of data collection, we added a few questions to the interview based on results from the previous two rounds. The third round of data collection expanded the age range of our sample and let us examine the influence of users' past experience and concerns on their perception and behavior. Six interview recordings were lost due to equipment problems but field notes on paper were available. The remaining 22 of the 28 interviews were recorded and transcribed (9 technical, 7 nontechnical, and 6 community participants). Aside from analyzing the drawings, we performed qualitative data analysis of the verbal transcripts and field notes using a grounded theory approach [10]. The data were coded in Dedoose (http://www.dedoose.com/). A second researcher independently coded 15% of all the interviews. Our analysis showed a good inter-coder agreement between the two researchers (kappa = 0.79).

4. RESULTS

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

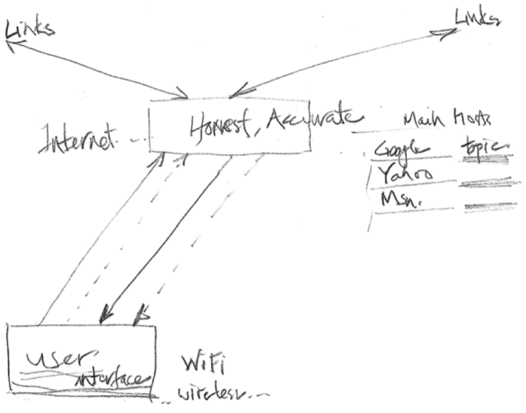

Figure 1. Internet as service (C01)

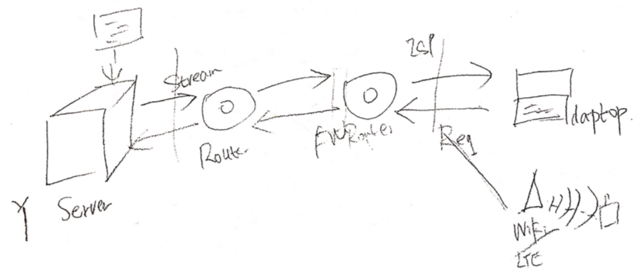

Figure 2. Articulated model with hardware components (T10)

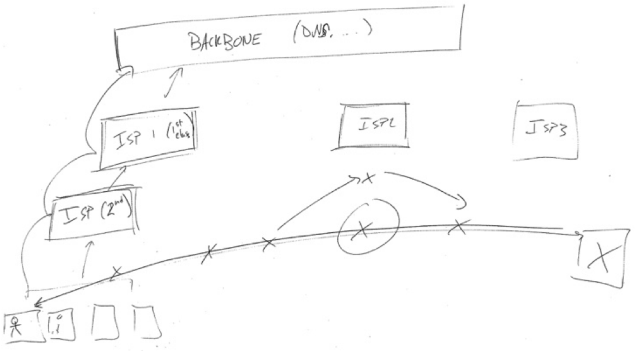

Figure 3. Articulated model with multiple layers of the network (T06)

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

4.1 Users' knowledge of the Internet

Participant models varied in their representation of the Internet as a simple system or service (the "Internet" in Figure 1) or as an articulated, technically complex system (Figure 2 and 3).

4.1.1 Simple vs. articulated system mental models

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

"So everything that I do on the Internet or that other people do on the Internet is basically asking the Internet for information, and the Internet is sending us to various places where the information is and then bringing it back." (C01, Figure 1)

Most lay participants only expressed surface-level awareness of organizations and services that they interacted with directly such as Google and Facebook, but did not mention any of the underlying infrastructure. When talking about making online payments, for example, they mentioned a number of different organizations involved in the process such as "the bank," "Amazon," and "PayPal." Some were aware of physical objects that helped them connect to the Internet (see N05's drawing of a router in Figure 4). Three lay participants also drew mobile towers when describing a cellular network. Three thought satellites played a role in connecting them to the Internet, but none of the technical participants mentioned this.

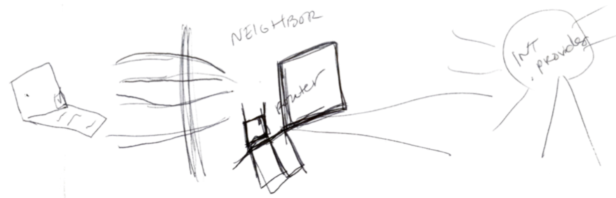

Figure 4. Drawing of how she uses neighbor's Wi-Fi (N05)

In most technical participants' drawings, we seldom saw a simple system or service representation of the Internet. Instead, users had more articulated models of the Internet as a complex system with varied hardware components and a more involved set of connections among components (Figure 2 and Figure 3). Ten technical and four lay participants belonged in this category. The number and presence of entities and organizations within participants' sketches mirrored to some extent their Internet literacy levels. The presence of other computers, servers, ISPs, DNS, routers, servers/clients, and infrastructure hardware spoke to a participant's knowledge and understanding of the Internet as a complex system.

Some technical participants articulated their view of multiple layers of the network (Figure 3), whereas most lay participants described one layer of the network. A few technical participants mentioned physical layers ("fiber cable", T05), or concepts potentially associated with a physical layer such as physical location (such as a "U.S. server," or a university as a physical entity). Most technical participants (9 out of 11) expressed broader awareness of entities and organizations involved in the Internet. For example, 6 technical participants noted there were many different ISPs. Furthermore, technically advanced users had specialized knowledge. Five technical participants mentioned network protocols such as "TCP/IP", "SMTP", or "IMAP", but none of the lay participants mentioned these concepts. Some

Table 2. Differences between simple and articulated models

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

technical participants also mentioned logical elements such as "routing" or "peering." The differences between these two types of mental models are explained in Table 2.

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

4.1.2 Awareness of security and privacy[1]

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

4.1.2.1 Public vs. private communication

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

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

4.1.2.2 Protection mechanisms

We coded users' expressed awareness of protection mechanisms such as encryption, passwords, certification of websites, and verification steps implemented by websites. One lay participant and seven technical participants said that their email, online payments, or connections could be encrypted. T04 said, *"If I'm going to use Gmail then I assume that, by default, the connection is going to be encrypted between my PC and the Gmail server."* Another technical participant drew a little lock sign in his model to indicate that the connections are encrypted (Figure 5).

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

4.1.2.3 Trust

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

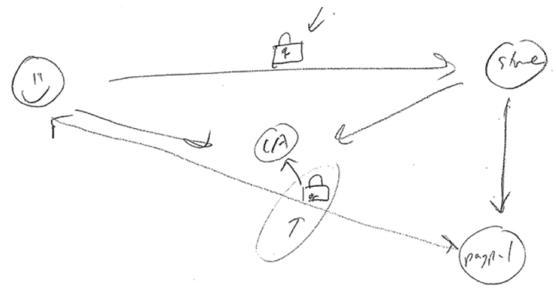

Figure 5. Model of making an online payment to a shoe store (T09)

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

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

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

4.1.2.4 Mobile phones vs. computers

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

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

4.2 Users' perceptions of their data

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

4.2.1 Where does my data go?

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

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

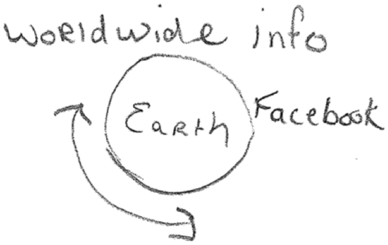

Figure 6. A depiction of where his information goes online (C04)

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

Regarding where their data is stored, participants mentioned "Google's large storage banks," cloud storage, ISPs, and advertising companies. One participant said, *"Once something is online, it's there forever."* (T11) A few others were not sure if information would be stored permanently, using the evidence of having seen webpages removed.

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

4.2.2 Who can see my data?

After each participant completed their drawing of the Internet, the interviewer asked, "Are there any other people, organizations, or companies that can see your connections and activities?" Privacy threats participants identified in frequency order include: companies that host the website (e.g., YouTube, Amazon) (mentioned by 18 out of 22 participants), third parties (e.g., advertisers or trackers) (mentioned by 14 participants), the government (mentioned by 12 participants), hackers or 'man in the middle' (mentioned by 12 participants), other people (e.g., other users online, other people using the same Wi-Fi) (mentioned by 11 participants), internet service providers (mentioned by 8 participants), employer (mentioned by 2 participants), and browser owners (mentioned by 1 participant). Figure 7 shows a fairly complete representation of all the people and organizations that the participant thought had access to his information, including the government, hackers, company, ISP, and third parties. This participant (T11) studied computer science in school, but stated that his current job was not related to technology.

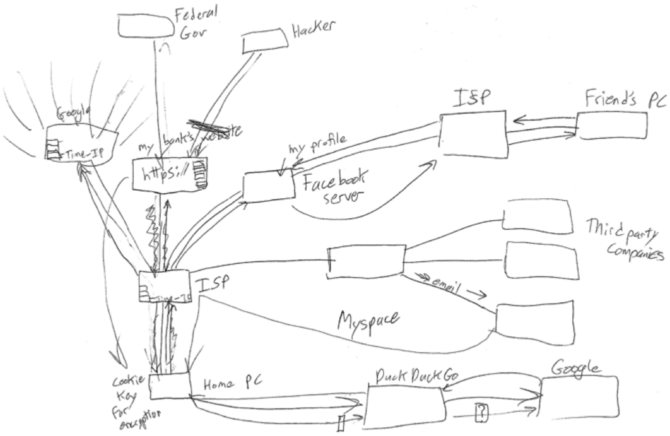

Figure 7. Model of the Internet including who can access his information (T11)

We compared how much lay and technical participants' mentioned the six most frequently mentioned threats. These two groups did not differ significantly in their general awareness of who has access their data. Lay participants mentioned on average 3.23 threats (out of 6), whereas technical participants mentioned on average 3.67 threats, a small non-significant difference overall. As shown in Figure 8, technical participants were significantly more likely, however, to mention hackers having access to their data than lay participants did. Across the categories of threat, they were more specific in identifying threat such as ISPs, whereas lay participants mentioned more vague threat such as third parties: "*whoever tries to make money off of you.*" (C02) This generality was probably due to the more simplistic mental models lay participants had about the Internet.

Although technical education did not seem to influence participants' overall perception of privacy threat, the mental models (simple vs. articulated) were somewhat predictive of the number of threats people perceived. We found that, on average, participants with articulated models mentioned more sources that might have access to their data than those with simple models (mean number of threats mentioned by people with articulated models = 4 and the number mentioned by those with simple models = 2.56, *t* [20] = 2.80, *p* = .01). Those with articulated mental models expressed higher awareness of privacy threats from government, hackers, and ISPs. This higher level of awareness may be caused by these people's better understanding of where risks could occur in the network. For example, with a mental model like Figure 1, there is no way the user would know what privacy risk his ISP could bring to his data on the Internet.

Besides these specific threats, some participants thought that "everyone" could access their information, either in the general

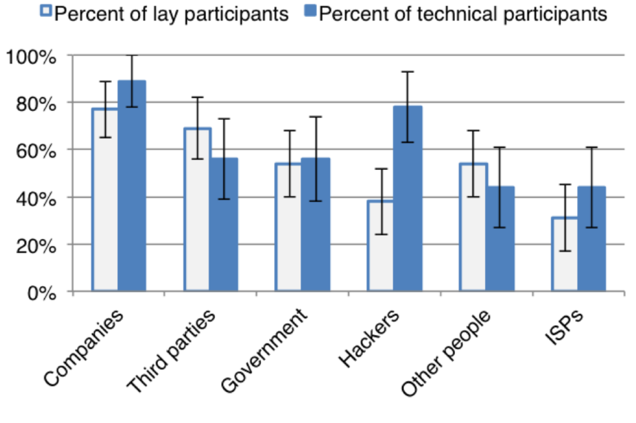

Figure 8. Percent of lay or technical participants who mentioned each group that might have access to their data.

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

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

4.2.3 Different types of information

Previous research has shown that users consider some personal information more sensitive than others [3]. From our interviews, we saw different user privacy expectations for different types of information, including not only personal information, but also technical identifiers. For instance, three lay participants thought companies could access their purchase history but not credit card information (N06, C02, C06). N11 suspected companies would be more interested in what she watched on YouTube than her emails, so she expected more protection on emails. Some participants were aware of the differences between identifiable information (such as names) and non-identifiable information like an ID number or IP address (C05), but she also said *"they could find it [my name] from this ID."* T09 pointed out that even for encrypted messages, his ISP could see all the packets and they could still tell *"where the origin, which is me, and what it's going."*

4.3 How do people protect their information?

4.3.1 Protective actions

We asked people: "Did you do anything to prevent any others from seeing your connections or activities?" Participants mentioned a wide range of protective actions they had tried, such as not logging in to websites, watching for HTTPS, and using cookie blockers or tracker blockers. We categorized the actions participants used into four categories as shown in Table 3. *Proactive risk management* includes general precautionary steps people take in daily use of the Internet. *Event-based risk management* includes people's actions towards specific requests or intrusions. *Controlling digital traces* includes actions that mask or remove people's digital or physical footprints. *Securing connections* indicates methods people take to make sure their connection to a certain site or their general Internet connection is secure or anonymized.

Table 3. Protective actions used by lay participants and technical participants

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

Although our technical participants were more knowledgeable of how the Internet works in the backend, they did not in general take more steps to protect their information online, in comparison with lay participants (Mean types of actions used by technical participants = 2.33, lay participants = 2.23, *n.s.*). As shown in Table 3, the only difference was that technical participants were somewhat more likely to mention securing their connections than lay participants and the comparison shows a trend approaching significance (t [20] = 1.99, p = .07). This finding contrasts with some of the prior work that has shown a correlation between technical knowledge and privacy practices [28], but this ostensible contradiction may stem from how we and other authors explored the influence of technical knowledge. Our study was focused on how people understand the Internet and its infrastructure whereas other studies [18,28] have mainly focused on users' Internet literacy and their familiarity with privacy practices.

We counted the diversity of privacy threats that participants mentioned among six frequently-mentioned sources of threat in Figure 8: companies, third parties, government, hackers, other people, and ISPs. We then compared how perceptions of threat were related to protective action, by conducting a nonparametric correlation analysis on the number of threats they mentioned and the number of protective action types they took. The analysis yielded a moderate correlation (r_s = .40, p = .06). This result indicates that the awareness of privacy threats is probably a stronger indicator of people's protective actions than their general technical background. This comparison points to a difference in the impact of general technical knowledge, which does not seem to predict actions, and the awareness of Internet privacy risks.

Many participants had some knowledge of protective actions but had not used them. This may be one consequence of the privacy paradox [4] whereby people have general desire for privacy but do not act on this desire. We wanted to know what our participants would say about why they did not take steps to protect their information.

4.3.2 *What prevents people from taking action?*

Four categories emerged when participants talked about why they did not take actions to protect their information from being seen. The most common explanation was similar to the statement, "I've nothing to hide" [37].

Eleven participants (8 lay and 3 technical participants) said they were not worried about their information being accessed or monitored or did not have the need to use tools. Many participants were not concerned because they did not do anything very subversive, illegal, or had little to protect. T10 said, *"I don't care who sees and reads my email"* although he was aware that *"hackers can act as mail servers."* Two participants were not worried also because *"I don't put that much information out there."* (C03 and C04) Three participants said they had too little money to protect, *"I don't have much money to worry about."* (C03) C01 said he was not worried because his data is among *"an awful lot of data."* T11 said he knew a lot of methods that other people had used to mask their IP address, such as proxy servers, but he never pirated so much music that he felt the need to do so. T04 mentioned Tor as a protective method during the interview, but also said, "*Till now I haven't had the need to use Tor."*

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

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

For a minority, a feeling of helplessness and lack of procedural knowledge prevented them from taking any action [36]. C05 said that hackers would probably hack into the website servers instead of individual users, and there was nothing he could do about it. Four lay participants said they lacked enough information to discuss actions they could do to prevent others' access to their information. C03 said he deleted cookies and then said, *"I don't know how to do anything else."*

The relationship of risk perception and action is also shown in participants' remarks. A technical background could influence awareness of threats and risks to some extent, but risk perception could also be shaped by personal experience. T11 started using DuckDuckGo after hearing about news related to Target's data breach and NSA monitoring. He became worried about how many people could see his information online. T11 had also been harassed by a Craigslist job poster because he gave out his phone number and email address. The Target data breach was also mentioned by C02, C07 and T11. After C07 was notified of the breach, she was not sure whether she was a victim or not, so she kept checking her statements carefully for a few months. Consistent with previous research [23], these instances suggest that past negative experience triggers more secure online behavior and a heightened level of privacy concern. In contrast, people who had not experienced a negative event seemed to be habituated to the convenience brought by the Internet and were less motivated to take protective actions online. A community participant (C04) had a friend who experienced identity theft, but hearing about this story did not make him worry about his information, and he stated, *"unless it happens to you it's hard to walk in somebody else's shoes."*

5. DISCUSSION AND IMPLICATIONS

Our study suggested technical education determined whether people viewed the Internet as a simple, service-like system or as an articulated technical system. Those with a more articulated model of the Internet expressed higher awareness of the different people or organizations that could access their data. However, technical participants did not take more steps to protect their online information than those with lower technical knowledge. After the second round of data collection, we invited five networking and computer security experts to review several lay and technical participants' models and discuss implications for security and privacy.

5.1 The role of knowledge in privacy decisions

Previous research is unclear as to whether or not Internet knowledge is associated with better management of one's privacy and security. We found little difference in the actions that people with more technical Internet knowledge took versus the actions lay participants used except that technical participants were slightly more likely to secure their connections (Table 3). Many technical participants expressed that they did not need to take action, and that the tools were inconvenient. These observations echo the finding in [15] that technical users complained about practical factors that prevented them from taking secure actions (e.g., "security is too expensive"). Also, expert reviewers pointed out that technical participants might be overconfident about their knowledge, which may cause a *"skewed view of security"*.

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

"*When it's just a magic black box, you tend to say well, I trust the magic black box, and so I would worry a little bit more that someone with this level of abstraction would not think as much about who could be sniffing on their communications or changing it or how they interpret security warnings and things like that.*"

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

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

5.2 Uncertainty in knowledge and concerns

Across all three rounds of data collection, participants expressed a great deal of uncertainty or lack of knowledge about how the Internet works, how their data is collected, shared or stored, what protective actions they could use, and whether the protection is effective or not. This finding echoes Acquisti et al.'s [3] work demonstrating the privacy uncertainty. For example, N11 used a Google app to block trackers but she was not sure how effective it was and still concerned: *"I don't think it blocks everything."* Several nontechnical and community participants were confused about how attacks or problems happened. Finally, three technical participants expressed doubts about who had access to their data. These different uncertainties may prevent people from accurately estimating their privacy and security risks.

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

Much previous research about the privacy paradox discusses people who claim they are concerned but do not take steps to protect their information. Our study reveals another possibility: participants who showed less concern about privacy in some situations actually took protective actions in other situations. For instance, two participants (C02 and T11) mentioned reading websites' terms of service to figure out how the companies handled their data, which indicates they are pretty cautious about data privacy. But when the interviewer asked about their privacy concern levels, C02 said he was only moderately concerned. Although T11 said he was worried about privacy, the reason he switched back to Google from DuckDuckGo was *"I don't really have anything to hide."* A number of researchers have shown that general privacy segmentations like Westin's do not sufficiently capture people's complex privacy needs, and concerns do not align with their behavior [43]. Our finding suggests that we need more detailed measures of privacy concerns instead of a general privacy concern scale.

5.3 Implications for design and policy

People rely on their specific experiences and on observable cues to understand how their information is accessed, used, or protected online. Experiences include actions they've taken (e.g., password or monthly payment) and received (e.g., spam). Cues include interface cues (e.g., lock sign, dots replacing password), dynamic information (e.g., tailored advertisements), and social information (e.g., comments on a post). Most of our participants were aware of personalized services or advertisements, which spoke to their high awareness of companies and third parties having access to their data. A technical participant explicitly noted this transparency: *"They are totally telling you that they know what you're viewing, because they recommend videos for you"* (T06). Social cues on sites like YouTube and Facebook (e.g., user profiles, number of views, and uploader's profile) indicated the presence of other users, which rendered participants' activity on those sites more public. Regardless of their technical knowledge, participants seem to have made most of their privacy-related decisions based on these experiences and cues.

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

The dilemma of multiple sources of risk implies that even if we raise awareness about some sources of risk (e.g., tracking and third party advertisements), there are others, and if we try for more comprehensive warnings, we may cause overload, annoyance, and security tool abandonment or lack of adoption [41]. Moreover, warnings can raise user confidence, which can in turn increase their risk behaviors.

In half of the interviews we coded, participants said they trusted institutions or companies to take care of their security. Some expert reviewers and a few technical participants suggested that users are putting too much trust in the system or the software, and taking too little responsibility: *"If you want to* [*achieve*] *privacy, you have to take that into your own hands"* (T06). This attitude reflects a laissez-faire policy perspective that our data and prior work challenges. If users cannot understand or control their own privacy, or if they have a limited role, where should responsibility rest? Policy makers could enforce more strict laws and regulations to mandate organizational practices, but users may still feel uncertain and helpless if we fail to provide good education programs about the influence of policies on their personal data. In the last round of interviews, we asked the community participants whether or not they thought current laws provided enough protection for their privacy. All but one participant thought laws did not provide enough protection; however, when the interviewer asked participants about their knowledge of Internet-related policy, most participants could not articulate anything beyond what they heard in news reports, suggesting a strong need for investigative journalism.

5.4 Limitations and future work

Because we used a think-aloud style qualitative study, our observations were influenced by the questions we posed and the knowledge people recalled. Participants may have had more knowledge of the Internet or security mechanisms than they expressed. Another limitation of conducting a qualitative study is that we have a comparatively small sample size. The small sample size may prevent us from detecting small but real effects of declarative and procedural knowledge on motivations and behavior. We are conducting larger sample surveys that will provide more statistical power to detect correlations among users' knowledge, expressed awareness of threats, and use of protection tools.

Individual demographic differences such as age, profession, and area of the country may also influence people's Internet perceptions and behaviors but are more appropriately compared in a quantitative study. Our sample is especially sparse in some age ranges. In future work, the interplay between demographic factors and users' technical background should be examined.

Another limitation of this work is its scope. Our study specifically examined participants' knowledge of the underpinnings of the Internet, how they tried to control data access by others, and, in the post-test survey, their understanding of some tools and concepts for controlling privacy on the Internet. We did not measure participants' knowledge of how attacks occur or how they understood different privacy and security threats in detail. We used their awareness of who had access to their data as a proxy for their awareness of risks and threats. Our data showed gaps in people's understanding of how attacks occur, but we do not know how these gaps influenced their risk perceptions or behaviors. There is much more to learn about these and other dimensions of Internet knowledge. More extensive measures are needed to explore the relationships among technical knowledge, understanding of threats, and people's privacy behaviors. For instance, it will be important to understand if knowledge of the complexity, layers, entities, and operation of the Internet help people to understand how and where threats can occur, or whether they simply need to have in mind a mental list of threats and methods to lower risk. Future work should examine whether education or design interventions can improve specific aspects of users' mental models of the Internet such as entity awareness. We also need to understand better how people understand security versus privacy—or whether they even need such a distinction.

6. CONCLUSION

As the Internet becomes more technically complex and, at the same time, more intertwined with everyday life and the well being of organizations, we face the question of how to educate users to help them protect their privacy. We conducted a qualitative study to investigate users' mental models of the Internet and their knowledge of data flow on the Internet. We examined how they conceptualize the process of connecting to the Internet and how they think others can access their data online. Our analysis revealed strong differences among users with different educational backgrounds. The majority of those without computer science education had simple, service-oriented mental models whereas those with a background in computer science had an articulated many-layer model of the Internet that included key entities and organizations. People with a more articulated model expressed higher awareness of specifically who might have access to their personal data and communications. Yet technical background was not directly associated with more secure behavior online. Almost universally, participants' privacy protective actions or lack of action were informed by personal context and experiences, such as a feeling they had nothing to hide, and in some cases by immediate cues in the online environment such as a security emblem or famous company name. Our work suggests a need for more research into privacy protections that reduce the responsibility on users to understand how the Internet works and to make myriads of privacy protection decisions based on their technical knowledge.

7. ACKNOWLEDGEMENTS

We thank Katherine Mashek and Zachery Mullins for their help conducting interviews, and we thank Sauvik Das, members of the CoEX lab, David Naylor, Peter Steenkiste and other members of the XIA project for their feedback and suggestions. We also thank all the participants for sharing their experiences and thoughts, and the anonymous reviewers for their insightful comments. Support for this work was provided by NSF grants CNS1345305 and CNS1221006.

8. REFERENCES

[1] OpenSSL 'HeartBleed' vulnerability: https://www.us-cert.gov/ncas/alerts/TA14-098A Retrieved March 12, 2015.

[2] Target confirms massive credit-card data breach: http://www.usatoday.com/story/news/nation/2013/12/18/secret-service-target-data-breach/4119337/ Retrieved March 12, 2015.

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

[4] Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science,* 347(6221), 509-514.

[5] 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.

[6] boyd d and Hargittai E (2010) Facebook privacy settings: Who cares? First Monday 15(8). Available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086/2589

[7] Bonné, B., Quax, P, & Lamotte, W. Your mobile phone is a traitor! – Raising awareness on ubiquitous privacy issues with SASQUATCH. International Journal on Information Technologies & Security, 3 (2014), 39-53.

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

[9] Brown, A., Mortier, R., & Rodden, T. MultiNet: reducing interaction overhead in domestic wireless networks. In Proc. of CHI '13. ACM (2013), New York, NY, USA, 1569-1578.

[10] Corbin, J.M. and Strauss, A.L. Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage Publications, Inc, 2008.

[11] Das, S., Kim, T. H. J., Dabbish, L. A., & Hong, J. I. The effect of social influence on security sensitivity. In Proc. SOUPS (2014).

[12] DiMaggio, P., Hargittai, E., Neuman, W. R., & Robinson, J. P. (2001). Social implications of the Internet. Annual review of sociology, 307-336.

[13] Dommeyer, C. J., & Gross, B. L. What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy protection strategies. *Journal of Interactive Marketing* (2003), 17(2), 34-51.

[14] Ericsson, K. A., & Simon, H. A. Verbal reports as data. Psychological review, 87, 3 (1980), 215-251.

[15] Furnell, S. M., Bryant, P., & Phippen, A. D. Assessing the security perceptions of personal Internet users. *Computers & Security* (2007), *26*(5), 410-417.

[16] Friedman, B., Hurley, D., Howe, D. C., Nissenbaum, H., & Felten, E. Users' conceptions of risks and harms on the web: a comparative study. In *CHI'02 Extended Abstracts*, ACM (2002), 614-615.

[17] Hargittai, E. Survey measures of web-oriented digital literacy. Social Science Computer Review (2005), 23(3), 371-379.

[18] Hargittai, E., & Litt, E. New strategies for employment? internet skills and online privacy practices during people's job search. IEEE security & privacy (2013), 11(3), 38-45.

[19] Hmelo-Silver, C. E., & Pfeffer, M. G. Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. Cognitive Science (2004), 28(1), 127-138.

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

[21] Jensen, C. and Potts, C. Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. In *Proc. of CHI 04,* ACM Press (2004), 471–478.

[22] Jonassen, D. & Cho, Y. H. Understanding Models for Learning and Instruction, chapter Externalizing Mental Models with Mindtools, pages 145–159. Springer US, 2008.

[23] Kang, R., Brown, S., and Kiesler, S. Why do people seek anonymity on the internet?: informing policy and design. In *Proc. of CHI 13*. ACM (2013), New York, NY, USA, 2657-2666.

[24] 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.

[25] Litt, E. Measuring users' internet skills: A review of past assessments and a look toward the future. New Media & Society (2013), 15(4), 612-630.

[26] 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.

[27] Page, K., & Uncles, M. Consumer knowledge of the World Wide Web: Conceptualization and measurement. Psychology & Marketing (2004), 21(8), 573-591.

[28] Park, Y. J. Digital literacy and privacy behavior online. Communication Research (2011), 40 (2), 215-236.

[29] Pew Research Center. What Internet Users Know about Technology and the Web. http://www.pewinternet.org/2014/11/25/web-iq/

[30] Poole, E. S., Chetty, M., Grinter, R. E., & Edwards, W. K. More than meets the eye: transforming the user experience of home network management. In Proc. of DIS 2008, ACM (2008), 455-464.

[31] Potosky, D. The Internet knowledge (iKnow) measure. Computers in Human behavior (2007), 23(6), 2760-2777.

[32] Raja, F., Hawkey, K., & Beznosov, K. Revealing hidden context: improving mental models of personal firewall users. In SOUPS 2009, ACM (2009).

[33] Resnick, M., & Wilensky, U. (1998). Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. The Journal of the Learning Sciences, 7(2), 153-172.

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

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

[36] 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.

[37] Solove, D. J. 'I've got nothing to hide'and other misunderstandings of privacy. *San Diego law review*, *44* (2007), 745-772.

[38] Tbahriti, S., Ghedira, C., Medjahed, B., & Mrissa, M. Privacy-Enhanced Web Service Composition, IEEE Transactions on Services Computing, 7, 2 (2013), 210-222.

[39] Turow, J. Americans & online privacy: The system is broken (2003). Annenberg Public Policy Center, University of Pennsylvania, 3-35.

[40] 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, ACM Press (2012).

[41] 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.

[42] Wash, R. Folk models of home computer security. In Proc. of SOUPS 2010, ACM (2010).

[43] 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 Proc. of SOUPS 2014.

[44] Raja, F., Hawkey, K., & Beznosov, K. Revealing hidden context: improving mental models of personal firewall users. In SOUPS 2009, ACM (2009).

[45] 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.

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

APPENDIX A. Prescreen Survey

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

Technical Network Knowledge Scale

1. What is a three-way handshake in TCP/IP?

□ Three or more computers connected and communicating together
■ A method to establish a connection between two computers
□ Three computers on the same LAN or WLAN
□ A deal made between an ISP and a customer regarding Internet service
□ I'm not sure

2. Which of the following protocols work on the Data-Link layer of the OSI Model?

□ SMTP
□ HTTP
□ UDP
■ ARP
□ I'm not sure

3. Which of the following is the correct order for the OSI model layers?

□ Physical, Data Link, Transport, Network, Presentation, Session, Application
□ Physical, Data Link, Network, Transport, Presentation, Session, Application
■ Physical, Data Link, Network, Transport, Session, Presentation, Application
□ Physical, Data Link, Transport, Network, Session, Presentation, Application
□ I'm not sure

4. Which numbers below represent an IP address?

□ 2042.1.6.227
□ 125.120.255
□ 72.1380.12.86
■ 138.5.221.113
□ I'm not sure

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

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

6. Which of these statements about SSL/CAs is NOT correct?

□ CAs can be compromised by attackers
□ A CA is a third party organization
□ A CA issues digital certificates
■ Using trusted certificates from a CA always guarantees the owner's identity
□ I'm not sure

*7. What does the wireless network encryption tool WEP stand for?

■ Wired Equivalent Privacy
□ Wireless Equivalent Privacy
□ Wireless Equivalent Protocol
□ None of the above
□ I'm not sure

8. Of the following choices, what is the best choice for a device to filter and cache content from web pages?

□ Web security gateway
□ VPN concentrator
■ Proxy server
□ MAC filtering
□ I'm not sure

APPENDIX B. Interview Script

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

making a payment online; receiving an online advertisement; browsing a website.

Interviewer:

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

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

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

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

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

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

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

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

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

APPENDIX C. Post-test Knowledge Survey

This survey was given to participants in our interview study to assess their technical knowledge of the Internet, privacy and security. It was also given to students in a graduate level computer networking class and MTurk participants in another research study. We computed the scale reliability by combining these three datasets together (the participants in this study, the students in the networking class, and the MTurk participants in another research study). Total N = 432. Cronbach's alpha for *Internet Know-How Self Report Scale* is 0.88. Cronbach's alpha for the *Technical Knowledge of Privacy Tools Scale* is 0.66. Note: The correct answers are marked in black boxes.

Internet Know-how Self Report Scale
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.
IP address	□	□	□	□	□
Cookie	□	□	□	□	□
Incognito mode / private browsing mode in browsers	□	□	□	□	□
Encryption	□	□	□	□	□
Proxy server	□	□	□	□	□
Secure Sockets Layer (SSL)	□	□	□	□	□
Tor	□	□	□	□	□
Virtual Private Network (VPN)	□	□	□	□	□
Privacy settings	□	□	□	□	□

Technical Knowledge of Privacy Tools Scale
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.	□	■	□
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 request coming from a proxy server cannot be tracked to the original source.	□	■	□