# Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users

Rick Wash
School of Journalism
Michigan State University
wash@msu.edu

Emilee Rader
Dept. of Media and Information
Michigan State University
emilee@msu.edu

## ABSTRACT

Home computers are frequently the target of malicious attackers because they are usually administered by non-experts. Prior work has found that users who make security decisions about their home computers often possess different mental models of information security threats, and use those mental models to make decisions about security. Using a survey, we asked a large representative sample of United States Internet users about different causal beliefs related to computer security, and about the actions they regularly undertake to protect their computers. We found demographic differences in both beliefs about security and security behaviors that pose challenges for helping users become more informed about security. Many participants reported weakly held beliefs about viruses and hackers, and these were the least likely to say they take protective actions. These results suggest that all security knowledge is not the same, educating users about security is not simply a more-is-better issue, and not all users should receive the same messages.

## 1. INTRODUCTION

For most people, protecting their home computers from hackers and viruses is rather difficult. They see celebrities having their information stolen by hackers [39]; they don't understand how anti-virus software works [36]; they can't see the benefits of patching, but frequently see downsides [34]; and even when they decide they want to protect themselves, they can't always correctly configure their computers [38].

Despite this difficulty, home computer users have to make many security-relevant decisions every day. They receive links to invalid or suspect websites in their email, and have to decide whether to click on them. They hear about problems with computer viruses, and need to decide whether to purchase and use anti-virus software. The anti-virus scan slows down their computer, and they need to decide whether to postpone it to regain their use of the computer or wait until it finishes.

Wash [36] found that individuals can possess multiple different "folk models" of the threats they are worried about. Each different folk model leads users to make different choices when faced with these everyday computer security decisions. Rather than characterizing users along a continuum of less knowledge to more knowledge as is traditionally done [25], Wash's work suggests that there are a number of different beliefs that each can lead to different behaviors.

To better understand how different types of beliefs about computer security threats can affect the way people make choices to protect their computers, we asked a large nationally representative sample of U.S. Internet users about both their computer security beliefs and the security actions they take. The most common beliefs involve more direct and visible threats, and these beliefs are associated with more positive security decisions. More sophisticated security beliefs that involve more technological knowledge often are associated with fewer precautions. We also find that more educated users and older adults (50+) tend to have these more sophisticated beliefs, where younger people and people with lower levels of education tend to focus more on direct and visible threats.

## 2. RELATED WORK

### 2.1 Home Computer Security

A recent Pew survey shows that over 76% of the US population accesses the Internet from their home [29]. Computers in people's homes have changed our society, but they have also imposed new risks; home computers are under constant threat. Additionally, we are now seeing increased use of mobile phones, tablets, and other Internet of Things devices that are connected to the Internet and all of these are potential targets [32].

Despite not being security experts, home users are tasked with administering and making security decisions for their computers and devices. This makes protecting these computers difficult. Home computer users who feel psychological ownership for the computer are more likely to engage in protective measures [3]. One of the major strategies they use is to find ways to delegate the responsibility for security to some external entity, which could be technological (like a firewall), social (another person or IT staff), or institutional (like a bank) [13].

Gross and Rosson [18] studied what security knowledge end users, who were not directly responsible for security but had access to sensitive information, possessed in the context of large organizations. Users' security knowledge was "neither comprehensive nor sufficient" to maintain proper secu-

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

rity, but common security actions such as locking the screen when away were better understood and practiced. Users in both organizational [18] and home settings [34] also frequently conflate security and functionality failures or problems. For example, users refuse to install future security updates because past updates changed critical user interface elements [34].

A wide variety of security advice has been provided to computer users, particularly in large organizations. Researchers in large organizations have investigated the effects of different kinds of training programs and security policies on security outcomes [2, 12]. Wash [36] lists 12 pieces of security advice found from Microsoft, CERT, and US-CERT that are specifically targeted at home computer users. Egelman and Peer [14] identified 30 security behaviors that represent common advice given to home computer users. Hoban et al. [20] examined much of this advice, and found that online educational materials often focus on virus and phishing threats, but rarely mention hackers as a threat. Larose, Rifon, and Enbody [25] recommend emphasizing personal responsibility to encourage users to engage in protection behaviors. However, users often do not follow the advice found in security education materials and persuasive appeals. Herley [19] argues that when non-expert users reject security advice, it is often rational for them to do so. Advice to end users often ignores the costs of their time and effort, and therefore overestimates the net value of security.

All of this research uses small non-representative samples. This makes it impossible to understand how prevalent different folk models or different security behaviors are in society [11]. We seek to measure both mental models of security and security behaviors in a large, representative sample. By understanding what behaviors are common and among whom, we can examine society-wide vulnerabilities due to mental models. We can also better understand what types of beliefs make people particularly vulnerable to security problems.

## 2.2 Mental Models of Security

An important aspect of the security decisions of home computer users is their existing knowledge about computers and computer security issues. More knowledge about common security issues is frequently found to be correlated with intention to behave securely [25]. However, most studies find that knowledge is not enough, and that additional motivations must be in place for people to make secure decisions [3, 26].

Most research in this area has measured security knowledge on a continuum from *less knowledge* to *more knowledge*. For example, Kumar et al. [24] counts the number of common security measures that the user is aware of, and finds that people familiar with more of these are more likely to engage in security behaviors. Shillair et al. [31] measure knowledge by asking about two forms of malware (spyware and Trojans), and combining answers into a single measure of low versus high knowledge. They did not find that more knowledge increased behavior, but prior knowledge had an important interaction effect for what type of communication was best for increasing security intentions.

However, knowledge about security does not easily fall into a more-is-better continuum. Wash found that home computer users have a variety of different "mental models" of security threats [36], and that these models are often used to make security decisions. Mental models describe how a user thinks about a problem; it is the model in the person's mind of how things work. People use these models to make decisions about the effects of various actions [21, 17] by cognitively simulating the actions and running the model forward in time to examine potential outcomes. Mental models are not the same as knowledge; rather, they usually represent a set of causal*beliefs* that a person possesses that he or she uses to guide decisions and behavior [9]. Security experts differ from non-experts in the mental models that they use. Asghapour et al. [4] conducted a card sorting experiment; participants were instructed to match words with a set of computer security related concepts. They found that experts and non-experts show differences in which analogy (medical, crime, etc.) they felt the concepts were closest to.

Economists often talk about products as being horizontally differentiated or vertically differentiated [35]. Products are differentiated vertically when everyone agrees which product is better than the other (e.g. $100 is better than $10). Products are horizontally differentiated when some people like one product, and other people prefer the second product (e.g. baseball vs. football).

Using this to draw an analogy, the traditional way of measuring security knowledge is vertically differentiated: more knowledge is better for making good security decisions than less knowledge. However, we follow the lead of Wash [36] and treat security knowledge as horizontally differentiated: there are a variety of causal beliefs about computer security, and even simplified or incorrect beliefs can lead to good decisions. We seek to measure a number of different types of beliefs about computer security, and then identify which types are associated with positive security behaviors.

### 2.2.1 Measuring Mental Models of Security

Many scholars have examined mental models in the anthropology tradition. For example, Kempton [22] used both semi-structured interviews and analysis of log data to study folk models of thermostat technology in an attempt to understand the wasted energy that stems from poor choices in home heating. Wash [36] used similar interviews to study mental models of home computer users. D'Andrade [9] summarizes interview-based approaches to discovering mental models. Interview-based approaches allow examination of the details of an individual's mental model; unfortunately, interview-based approaches are too time-intensive to be used on large samples, and therefore cannot be used to measure prevalence of different beliefs in a population.

A number of psychologists have studied mental models using human-subjects lab experiments. Johnson-Laird has done a long series of studies examining how mental models form, how they represent time, and how they are used to make decisions [21]. He never measures the models directly, but rather provides information to participants and then measures behaviors and decisions that emerge from the models. This technique finds patterns in what mental models look like and how people use them, but does not reveal the details of the specific models.

A promising approach for measuring mental models of security is the card-sorting method used by Asghapour et al. [4]. They made a list of a number of words related to security and risk, and asked participants to sort these words into piles. They then analyzed these piles using dimension reduction techniques to find patterns among the words. They pre-specified which words were associated with each mental model, which limited the ability of the study to discover new models. It is also difficult to measure reliability and prevalence of the discovered models using this method.

## 2.3 Security Intentions and Behaviors

Measuring actual security decisions and behaviors with a survey is very difficult. Most security decisions depend on the *context* of the decisions [13]. While a person might generally prefer to run regular anti-virus scans, they might forgo a scan if they are in the middle of an important phone call or up against an important deadline. It is difficult to replicate real-world contexts in a survey.

Additionally, most security decisions are *repeated* [36]. Installing anti-virus software may be a one-time decision, but decisions like "should I click on this shady link?" or "what password should I choose for this site?" happen frequently. Surveys cannot usually ask about every instance of a security decisions, both because that would make the survey too long and tedious, and because it is difficult to know about the security decisions ahead of time.

To address these issues, most surveys follow one of two approaches. The first approach is to ask questions about *general behavioral intentions* [11]. Asking about intentions focuses on the future and what the participant wants to do. Intentions are a natural focus when conducting research that involves a manipulation, and you want to measure whether the person's future behavior is likely to have changed. There are also theoretical reasons why intentions are likely to map to behavior [1].

A second approach is to ask questions asking participants to recall how frequently they have undertaken a behavior in the past. This approach can be subject to recall bias, since it depends on the memory and honesty of the participant. However, it focuses directly on actual behavior rather than relying on the theoretical link between intention and behavior [11]. In security in particular, participants often find it difficult to enact their intentions due to lack of skill [13] or confusing interfaces [38], and therefore past frequency might have a stronger connection to actual behavior than intention.

A number of recent surveys have used different questions to ask about general intentions toward computer security decisions. Larose et al. [25] ask their participants about their intentions to do eight specific security behaviors on a 7-point Likert scale, and then average the results as an overall measure of preventative intentions. Anderson and Agarwal [3] use two sets of questions about general protective intentions. One set asks general intentions (e.g. "I am likely to take security measures") about the participant's home computer, and the other set asks similar questions about protecting "the Internet." Egelman and Peer recently developed a new security behavior intentions scale that focuses on four common security behaviors: device securement, password generation, proactive awareness, and software updating [14].

With the exception of the scale created by Egelman and Peer (which was not yet published at the time we conducted this study), all of the measures we found used a unidimensional more-is-better measure of intention to make secure decisions. We wanted a measure that could capture more horizontal differentiation between security decisions. We seek to understand how different beliefs – different mental models – can lead people to make different types of security-related decisions. Our approach does not assume from the outset that more knowledge, or more sophisticated beliefs, are better for security. Rather, our survey is a measure of the association between patterns of causal beliefs and protective behaviors. Our survey allows us to better understand how different types of security knowledge can lead to *different* security decisions, rather than simply *more or less secure* decisions.

# 3. SURVEY INSTRUMENT

We sought to develop a survey instrument to help us understand how different types of security knowledge are associated with different types of security behaviors. By administering such a survey to a representative sample of United States Internet users, we can characterize which mental models are most common in the population, and which behaviors are commonly associated with these models.

Most existing security survey instruments measure vertically differentiated security knowledge; they assume that more-is-better and attempt to measure how much knowledge a person has. Egelman and Peer [14] is one exception, though they focus on horizontally-differentiated *behavioral intentions* rather than horizontally differentiated knowledge. However, Wash [36] found that many people differ horizontally in their beliefs, and therefore their knowledge, about computer security and that these differences lead to different behaviors. Because few of the existing methods for measuring security knowledge allow for horizontal differentiation, we developed new measures for this study.

## 3.1 Questions About Security Beliefs

Mental models are not traditionally measured with a survey [9]. They often include multiple types of knowledge: factual knowledge [36], counter-factual knowledge [22], knowledge about process [21], and understandings of what context is relevant [9, 36]. This presented a challenge for designing effective survey questions. We began by brainstorming a variety of survey question types based on suggestions in Dillman [11] and the findings in Wash [36]. We then conducted eight rounds of think-aloud trials with 3-5 participants each, to gain a better understanding of how our participants understood the questions we had generated. These think-alouds involved a member of the research team sitting together with a potential participant while they were taking the survey, and asking them to "think aloud" while reading and answering the questions. After these initial trials, we decided to focus on straightforward statements of *beliefs*. Beliefs form the fundamental components of a mental model, and appropriate sets of beliefs can indicate which mental model a person possesses. Statements of belief can easily be presented in a survey. We structured the questions to ask the participant to what extent they agree with each statement of belief.

Once we decided on the type of questions, we worked to develop a set of reliable questions. We identified a set of 132 beliefs that can indicate different mental models. Each belief focuses on exactly one salient aspect of one of the mental models that Wash [36] identified, and were drawn from that paper. For example, we asked participants to what extent they agree with the statement "Hackers target rich and important people." Participants who agree with this statement are likely to possess the *Big Fish* mental model, where participants who disagree with this statement are more likely to possess the *Burglar* or *Digital Graffiti Artist* models of hackers. Since Wash's [36] folk models were divided into two large categories — models about viruses and models about hackers — we divided our statements of belief into similar categories. We began with 67 belief questions about viruses and 65 belief questions about hackers. We also focused on beliefs that are independent of specific technologies (e.g. "Using anti-virus is important" rather than "Use McAfee Anti-virus") to help ensure our resulting instrument would remain valid for some time into the future.

During this analysis, we also found that many questions naturally grouped together into sets that represent high-level beliefs about hackers and viruses. For example, most people answered "Hackers target rich and important people" and "Hackers target large businesses" in very similar ways; both of these beliefs suggest a belief that hackers mostly target other people who aren't me. Rather than measuring mental models by pre-specifying an association between belief and model, we decided to directly measure beliefs and identify which beliefs were frequently held in common.

Over 100 questions is too long for a survey instrument. Following normal scale development practices [10], we identified a smaller number of questions accurately captured patterns in the larger dataset. We conducted an initial trial with 149 participants on Mechanical Turk (paying \$2 per survey) to identify which questions are internally reliable and which questions cluster into common beliefs. An initial factor analysis is available in Appendix A.

We identified a total of 34 beliefs that we believed were reliable and had high construct validity. Eighteen questions concerned viruses, and these questions clustered into three sets of 6 questions in the factor analysis. Sixteen questions concerned hackers, which clustered into three sets of five questions along with an additional, standalone question. We verified the reliability of these questions with a pair of validation surveys (one for viruses, and one for hackers), each of which involved 200 participants from Mechanical Turk. Each set of questions had high reliability (Cronbach's $alpha > 0.75$), which suggests that participants answered each question in the set similarly to the way they answered the other questions in the set. Each set of questions also loaded onto a single factor in the factor analysis (loading $> 0.400$).

## 3.2 Questions About Security Behaviors

We are interested in the decisions that home computer users make and the behaviors they undertake to protect their computers. We decided to measure the frequency of past behavior rather than the intentions for future behavior. We did this because security actions often don't match up with intentions due to contextual or skill reasons. We are more concerned with what actually happened on the computer than what a person wants to happen.

Since Wash [36] found that most people separate threats from viruses and threats from hackers, we divided potential behaviors into two categories: behaviors that can protect against viruses and other malware, and behaviors that can protect computers from active attacks from hackers.

To develop a set of questions about behavior, we followed the same procedure as above: we identified 20 questions and tested them in the same initial survey on Mechanical Turk (N=149). In our trials we found extremely similar results: behaviors to protect against viruses always clustered into the same two factors (software and "be careful"), and behaviors that protect against hackers fell into the same three factors (software, "be careful", and expert behaviors). As such, we don't include our initial factor analyses here; the final factor analysis is in Appendix B.

# 4. MAIN SURVEY

## 4.1 Sample

We conducted a large scale survey with a representative sample of the US Internet-using population. Unfortunately, there is no "phone book" for the Internet from which a truly random sample can be drawn [5]. Following the guidance from the American Association for Public Opinion Research, we decided to use sample quotas to get an appropriate variety of participants for our survey that matched as closely as possible the demographics of the target population [5].

Before we did this, though, we needed to understand the demographics of Internet users in the United States. We focused on getting a representative distribution of age and education level; we believe that those two demographic factors are likely to have the largest variation in information security mental models and behaviors. We began by finding the distribution of age and education in the US population by looking at the 2010 US Census [33]. We then found the percentage of the US population with Internet access grouped by those same age and education groupings from the Pew Internet and American Life project [29]. The Pew data provides a percentage *within* a category that uses the Internet. Multiplying those together and rescaling back to 100% then provides an estimate for the percentage of Internet-using US persons that fit into each demographic category. These estimates are in Table 1.

To recruit participants, we contracted with Qualtrics to provide access to panels of Internet-using adults (>18 years old) meeting certain demographic constraints. We provided our demographic quotas to Qualtrics, and asked them to recruit a total of 2000 Internet users in the United States who met our demographic quotas. Our study was approved as minimal risk by our institutions's IRB; we never collected identifiable information about our participants. We paid Qualtrics \$5 per participant, of which \$1.50-\$2.50 was paid to the participants in accordance with the conditions of their panel membership.

Qualtrics recruited participants to take the survey according to the quotas we specified. Two attention check questions were included in the survey, and participants who answered these questions incorrectly were removed before finishing the survey and did not count as part of the quota. Qualtrics ended up recruiting a total of 2006 participants who passed these attention checks. We removed from this sample anyone who finished the survey in less than 2 minutes, anyone who took more than 4 hours to complete the survey, and anyone who answered exactly the same answer for all Likert scale questions. After this cleaning, the final number of valid responses was 1993.

## 4.2 Demographics

The demographics of our sample largely reflect the demographics of the US Internet-using population, along the lines of age, education, and race. Table 1 contains the detailed demographics of our sample.

The only major deviation from the US Internet-using population in terms of demographics is gender. Internet survey

| | | |
|---|---|---|
| Men | 749 | 37.6% |
| Women | 1157 | 58.0% |
| No Gender Reported | 87 | 4.3% |
| Democrat | 769 | 38.6% |
| Independent | 618 | 31.0% |
| Republican | 595 | 29.9% |
| Some High School | 245 | 12.3% |
| High School Grad | 984 | 49.4% |
| College | 562 | 28.2% |
| Grad School | 202 | 10.1% |
| Age 18–29 | 420 | 21.1% |
| Age 30–49 | 777 | 39.0% |
| Age 50–64 | 552 | 27.7% |
| Age 65+ | 244 | 12.2% |
| No Children | 796 | 40.0% |
| Has Children | 1193 | 59.9% |
| White | 1629 | 81.7% |
| American Indian or Alaskan Native | 25 | 1.2% |
| Asian or Pacific Islander | 51 | 2.5% |
| Black or African American | 161 | 8.0% |
| Hispanic or Latino | 94 | 4.7% |
| Other or Not Specified | 33 | 1.7% |

**Table 1: Demographics of our sample.**

panels are known to be skewed toward women [5], and our sample has a similar skew. It is about 58% women and only 37.6% men, which differs significantly from the population, where men and women use the Internet in approximately equal numbers [29].

To measure political party affiliation, we used the method and wording of the question in Gallup polls. We first ask whether the participant considers herself a Democrat, Independent, or Republican. The vast majority of Americans answer "Independent". Then, only for the people who answer "Independent", we also ask "As of today, do you lean more to the Democratic Party or the Republican Party?" By combining the two answers, we are able to effectively measure political leaning. Gallup polls performed around the time when this survey was run (summer 2014) showed approximately 40% of the US leans Republican and approximately 40% of the US leans Democrat [15]. Our sample slightly under-represents Republicans and over-represents independents.

We set demographic quotas on Age range and Education level for people responding to our survey. As such, those two demographics exactly match our best estimates of the US Internet-using population. We decided to enforce these quotas because we suspected that age and education would be the largest influences on security beliefs. And indeed, as shown below in Table 4, we found the most variation along those two demographics.

The question measuring race was designed to be comparable with data from the Pew Internet and American Life data [29], and the sample racial demographics of our survey approximately match the US Internet-using population, with the exception that we slightly underrepresent Hispanics.

# 5. SCALES

We first present the scales that we created, along with data that indicate how common each belief is in our sample of US Internet users. In addition to the scales we developed for security beliefs, we also included a number of questions asking how often participants undertake behaviors that can protect them from these threats.

## 5.1 Factors About Beliefs

Using data from the large representative sample, we conducted an exploratory factor analysis to confirm the presence of the same factors we found during scale development. EFA can also be used for confirmation of factors [6], and is often preferred when there is not a strong theoretical motivation for associating a set of items as factors.

One of the factors about hacker beliefs proved unreliable; it had a low Cronbach's $\alpha$ ($< 0.70$) and the factor analysis did not identify it as a factor. We dropped that factor, leaving us with three factors about virus beliefs and two factors about beliefs about hackers. Table 2 contains details of the factors. The final factor analyses is available in Appendix B.

### *Virus Beliefs.*

Most people express a strong concern about malicious software. They usually group all malicious software under the term "virus." [36]. Beliefs about how viruses operate are likely to have an impact on the way that people make decisions to protect their computers. For example, Wash [36] found that people who believe that viruses are simply buggy software don't feel like they need to install anti-virus systems because they can simply not choose to download risky software. We identified three sets of beliefs about how viruses operate.

The first major belief about viruses is that **Viruses create visible problems**. This factor includes questions that indicate a belief that viruses infect home computers and cause a variety of problems that are mostly visible to home computer users. Most US Internet users agree with these questions; 91.4% of our participants averaged above a 3.0 on these questions.

A second major belief about viruses is that **You can protect yourself from viruses**. Viruses come from intentional choices like downloads and viewing ads, and either not downloading, or using an anti-virus software to scan downloads can prevent them. Only about 22.4% of the US Internet using population would agree with this belief, though 4.5% strongly agree with this (mean $> 4.0$).

The third major belief about viruses is that **Viruses are caught on the Internet**; often, there is little that can be done (other than possibly avoiding the shady parts of the Internet) to prevent them. Clicking on advertisements, downloading files, watching pornography, or simply visiting the wrong webpages can all cause you to catch a virus. Approximately 63.0% of the US Internet-using population share this belief.

### *Hacker Beliefs.*

Many people also express a strong concern about hackers [36, 20]. "Hacker" is often a catch-all term for bad people who operate via computers and all of the associated concerns. Different beliefs about hackers have been found to influence protective behaviors. For example, Wash [36] found that people who believe hackers only steal information en

|  | *Alpha* | *Mean* | *SD* |
|---|---|---|---|
| **Belief: Viruses create visible problems** | **0.76** | **4.00** | **0.62** |
| A virus causes computers to crash |  | 4.13 | 0.87 |
| A virus causes annoying problems |  | 4.45 | 0.77 |
| A virus erases important files on the computer |  | 3.86 | 0.89 |
| A virus steals personal and/or financial information |  | 3.92 | 0.93 |
| Being aware of what websites I go to will help me avoid getting a virus |  | 3.62 | 0.91 |
| **Belief: You can protect yourself from viruses** | **0.80** | **2.59** | **0.82** |
| You can't get a virus if you keep your anti-virus software up to date |  | 2.65 | 1.13 |
| Anti-virus software always detects viruses |  | 2.72 | 1.12 |
| The only way to get a virus is by downloading something |  | 2.38 | 1.11 |
| You can't get a virus if you never download things from the Internet |  | 2.35 | 1.11 |
| Blocking pop-ups makes it very difficult to get a virus |  | 2.87 | 1.00 |
| **Belief: Viruses are caught on the Internet** | **0.75** | **3.40** | **0.75** |
| Clicking on advertisements will give you a virus |  | 3.22 | 0.93 |
| Downloads from the Internet will give you a virus |  | 3.47 | 0.96 |
| Merely visiting the wrong webpages will give you a virus |  | 3.52 | 1.01 |
| Watching pornography on the Internet will give you a virus |  | 3.39 | 1.06 |
| **Behavior: Use Security Software** | **0.90** | **4.32** | **0.86** |
| Check anti-virus software to make sure it is up to date |  | 4.21 | 0.99 |
| Regularly scan the computer with anti-virus software |  | 4.23 | 1.02 |
| Use anti-virus software |  | 4.48 | 0.93 |
| Use security software such as firewall |  | 4.34 | 0.99 |
| **Behavior: Be Careful on the Internet** | **0.82** | **4.36** | **0.70** |
| Avoid downloading anything without knowing what exactly is being downloaded |  | 4.39 | 0.87 |
| Be aware of what websites you visit |  | 4.37 | 0.83 |
| Avoid clicking on email attachments from people you do not know |  | 4.53 | 0.84 |
| Block pop-ups |  | 4.17 | 0.95 |

|  | *Alpha* | *Mean* | *SD* |
|---|---|---|---|
| **Belief: Hackers target home users** | **0.83** | **3.84** | **0.69** |
| A hacker watches what you are doing on your computer |  | 3.86 | 0.88 |
| A hacker intentionally puts viruses on the computer |  | 4.00 | 0.91 |
| A hacker makes a record of everything on the computer |  | 3.64 | 0.95 |
| Hackers target home computer users |  | 3.90 | 0.85 |
| A hacker installs monitoring software on the computer |  | 3.82 | 0.87 |
| **Belief: Hackers target others** | **0.85** | **3.58** | **0.78** |
| Hackers target rich and important people |  | 3.31 | 1.08 |
| Hackers target large businesses |  | 3.81 | 0.97 |
| Hackers target the upper class |  | 3.18 | 1.06 |
| Hackers target banks |  | 3.76 | 0.94 |
| Hackers target large databases |  | 3.85 | .92 |
| **Behavior: Expert Security Settings** | **0.82** | **3.32** | **1.08** |
| Disable scripting on emails |  | 3.14 | 1.38 |
| Disable scripting on websites |  | 3.13 | 1.33 |
| Back up your information on an external hard-drive, network, or server |  | 3.41 | 1.32 |
| Update patches regularly |  | 3.58 | 1.31 |
| **Behavior: Be Careful on the Internet** | **0.76** | **4.20** | **0.73** |
| Be careful downloading software from the Internet |  | 4.34 | 0.88 |
| Use good passwords (good passwords include uppercase and lowercase letters, numbers, and symbols) |  | 4.31 | 0.87 |
| Avoid clicking on attachments |  | 4.03 | 0.98 |
| Always sign out of accounts when you are done with that website |  | 4.11 | 1.09 |
| **Behavior: Use Security Software** | **0.71** | **4.23** | **0.90** |
| Use some pre-existing security software such as firewall |  | 4.26 | 1.06 |
| Scan your computer regularly with anti-virus software |  | 4.21 | 0.98 |

**Table 2: Questions and Scales. All belief items use a 'Strongly Disagree' to 'Strongly Agree' 5-point Likert scale, converted for analysis to numbers 1–5. All behavior items were phrased "How often do you do the following security precautions to avoid getting a virus? / avoid being hacked?" and use a 'Never', 'Rarely', 'Sometimes', 'Often', 'Always' scale, also converted to numbers 1-5. Virus-related questions are in the left column; hacker-related questions are on the right. The names of the factors that we assigned to each group of questions are bolded.**

mass from large websites often don't take actions to protect their personal computers. From our survey, we identified two distinct sets of beliefs about hackers.

The first belief about hackers is that **Hackers target home computer users**; they break into home computers, monitor everything you do on your computer, and install viruses on your computer. 84.5% of the US Internet-using population would agree or strongly agree with this belief.

The second belief is that **Hackers target others**, mostly rich and important individuals and banks. This is a belief that hackers intentionally choose targets, and those targets are often other people with more money or power. About 71.3% of the US Internet-using population would agree with this belief.

## 5.2 Factors about Behavior

We were also interested in understanding what kinds of security behaviors people undertake to protect themselves. We asked how often the participant would do specific security-related behaviors that he or she could take *for the specific purpose of avoiding a virus/hacker.* We found that these actions almost always clustered into two major categories, with most people answering questions in each category very similarly.

The first cluster was behaviors that place trust in **using security software** (trust-in-software): anti-virus, firewall, and security products. Most users claimed to do this both to protect against viruses and to protect against hackers. 67.4% of the US Internet-using population would state that they use security software to protect against viruses at least "Often" (mean > 4.0 on a 1-5 scale). 58.1% would claim at least "Often" to protect against hackers.

The second cluster of behaviors place trust in oneself; they involved things that are frequently described as **be careful on the Internet** (trust-in-self): use good passwords, don't click on unknown things, block popups, and sign out of accounts when done. 69.1% of the US Internet-using population stated that they do these actions at least "Often" to protect against viruses, and 59.0% claim at least "Often" to protect against hackers.

Both of these clusters are likely influenced by *social desirability bias*: participants believe that it is socially desirable to be seen as doing these actions they are told are important, so they report doing it more often than they actually do [28]. Still, the *variation* in responses – exactly who reported doing these rarely – provides useful correlations with beliefs.

We found a third cluster of behaviors that some people used to protect against hackers. These behaviors are more advanced, **expert security settings**: disabling scripting on web pages, updating software patches, and backing up information. Many fewer US Internet-users do these behaviors, with only 24.2% of people reporting doing these "Often" or "Always". And due to the previously mentioned social desirability bias, the true number is likely to be lower.

# 6. RESULTS

## 6.1 The Relationship Between Belief and Behavior

We ran a series of regressions to better understand the relationship between the beliefs that a person has and the self-reported behaviors that they undertake to protect themselves. Table 3 contains the detailed results.

### *Protecting Against Viruses.*

There is a relationship between the beliefs that people possess and the actions that they state they take to protect themselves. People who believe that viruses cause visible problems report taking both trust-in-software and trust-in-self actions more often. This makes sense and is good; people who see viruses as causing problems for personal computers report trying to protect themselves (reading across the second row in Table 3).

People who believe that you can protect yourself from viruses by avoiding downloads and running anti-virus software actually report lower levels of use of both trust-in-software and trust-in-self actions (row 3). The effect size of this negative relationship is smaller but still statistically significant. This is interesting, and suggests that believing that you can protect yourself actually leads to more risky behavior.

Finally, people who believe that viruses are caught simply by browsing the Internet did not show any correlation, positive or negative, with actions to protect themselves (row 4).

### *Protecting Against Hackers.*

We found two clearly distinct sets of beliefs about hackers: that hackers target home computers, and that hackers target others. These two beliefs are not mutually exclusive – they have a 0.60 Pearson correlation – but they represent different worries about what hackers might do. This can be seen by looking at how they correlate with behaviors.

If a person believes that hackers target home computers, then they take positive actions to protect their computers. There is a positive and statistically significant relationship between this belief and all three type of actions – trust-in-software actions, trust-in-self actions, and expert actions (row 5). This makes sense; more concern about being attacked leads to more effort to protect themselves.

On the other hand, we found no relationship between a belief that hackers target others and any actions to protect computers (row 6). Our estimates are both very small and not statistically significant. The fact that this isn't negative suggests that participants don't necessarily feel safer on their computers. But rather, this belief is largely unrelated to the security precautions that people undertake when using their computers.

## 6.2 How Beliefs and Behaviors Vary: Demographics

Since we have a representative sample of US Internet users, we can compare beliefs about viruses and hackers across demographic groups. To do this, we calculated $g_\psi$, an effect size measure for each comparison of demographic groups. $g_\psi$ is a generalization of Hedge's $g$ designed to be used in situations where there are more than two groups to be compared. $g_\psi$ uses as the standardizer the estimated standard deviation of the whole population; this way, all $g_\psi$ estimates are in the same units and can be compared with each other [23].

A potentially more traditional approach to this is to directly compare means of groups, and then conduct a statistical hypothesis test for each comparison. The hypothesis tests normally accomplish two goals: they account for vari-

|  |  | *V. Software* | *V. Careful* | *H. Expert* | *H. Careful* | *H. Software* |
|---|---|---|---|---|---|---|
| 1 | (Intercept) | 2.35*** | 2.62*** | 1.13*** | 2.24*** | 2.05*** |
| 2 | Virus: Visible Problems | 0.27*** | 0.32*** | 0.08 | 0.23*** | 0.27*** |
| 3 | Virus: Can Protect Yourself | -0.09*** | -0.10*** | 0.14*** | -0.05* | -0.11*** |
| 4 | Virus: Caught on Internet | 0.06. | 0.03 | 0.07 | 0.01 | 0.03 |
| 5 | Hacker: Target Home Users | 0.12** | 0.06. | 0.22*** | 0.20*** | 0.17*** |
| 6 | Hacker: Target Others | -0.01 | 0.01 | 0.01 | -0.00 | 0.03 |
| 7 | Woman | -0.07. | 0.02 | -0.14** | 0.01 | -0.10* |
| 8 | Independent | 0.05 | 0.01 | 0.14* | 0.08* | 0.11* |
| 9 | Republican | 0.01 | -0.03 | 0.03 | 0.04 | 0.07 |
| 10 | Age 30-49 | 0.29*** | 0.16*** | 0.15* | 0.20*** | 0.35*** |
| 11 | Age 50-64 | 0.49*** | 0.24*** | 0.14. | 0.32*** | 0.54*** |
| 12 | Age 65+ | 0.55*** | 0.37*** | 0.15 | 0.28*** | 0.61*** |
| 13 | HS Grad | 0.26*** | 0.28*** | 0.34*** | 0.20*** | 0.30*** |
| 14 | College | 0.28*** | 0.22*** | 0.35*** | 0.20*** | 0.35*** |
| 15 | Grad School | 0.19* | 0.25*** | 0.40*** | 0.15* | 0.27** |
| 16 | Has Children | -0.04 | -0.03 | -0.03 | -0.05 | -0.11** |

**Table 3: Regression Results. Each column is a regression, with the dependent variable being the title of the column. The intercept represents the baseline category: man, republican, age 18-29 who didn't complete high school and doesn't have children.**

ance in the underlying measurements, and they provide an indirect measure of effect size (larger effects lead to lower p-values). However, in this situation, hypothesis tests are problematic. Uncorrected, they have a problem with false positives (too many tests are determined to be statistically significant). However, using a multiple comparisons correction like Bonferroni dramatically reduces the power of the tests, which also leads to improper interpretation of results and a bias towards overestimates [16, 7]. Instead of relying on an indirect estimate of effect size, we chose to report $g_\psi$, which directly estimates the size of the difference. Since $g_\psi$ normalizes by standard deviation, it also properly takes into account variance in the underlying measurements, and is more directly interpretable as a measure of the size of a difference in a population. No correction is necessary for $g_\psi$ since there is no acceptance/rejection decision.

Tables 4 and 5 contain these results for beliefs and behaviors. For the purposes of this paper, we take an effect size larger than 0.10 to be small but worth comment, and an effect larger than 0.30 to be moderate to large [8, 23].

There appear to be almost no differences in beliefs about either viruses or hackers between Men and Women. Also, there are relatively few, and mostly small differences between people across the political spectrum. The most interesting comparison here is between Republicans and Democrats; Republicans tend to be 0.16 standard deviations lower on the belief that you can protect yourself from Viruses by using antivirus software and not downloading files.

There are some noticeable differences between people with different amounts of education. People with only high school educations generally report higher agreement with all beliefs we found, and also report that they engage in more behaviors described as being careful on the Internet (use good passwords, don't click on unknown things, etc.). But this difference is particularly large for two beliefs. People without a college education are much more likely to agree with statements that indicate that you can catch viruses by casually browsing the Internet than people with college educations. Also, people who have attended grad school are much less likely to believe that hackers target home computer users. This suggests that greater education is associated with beliefs that they are less vulnerable online.

There are some differences across age cohorts. The largest difference is for the belief that casually browsing the Internet can cause you to catch a virus. Adults 50 years old and older are much less likely to agree with this belief than younger adults. Also, adults age 30–49 are the most likely to agree that viruses cause visible problems on home computers; adults 65 and older are very unlikely to agree with that belief. Regarding behaviors, older age cohorts are more likely to report that they engage in careful behaviors to protect themselves from viruses and hackers.

There are small differences in beliefs that emerge between people who have children and people who don't. People with children are more likely to believe in threats to their own computers: that viruses cause visible problems on home computers, and that hackers target home computers.

Finally, though white Americans are the most populous racial group in the US, they have very different beliefs than other races. Whites are less likely to believe viruses can be caught on the Internet, and that viruses can be protected against. However, they are most likely to report that they do behaviors that place trust in security software (anti-virus, firewall, and security products).

## 6.3 Grouping Participants

To better understand what kinds of beliefs happened together, we clustered participants using K-Means clustering. This clustering technique partitions participants into $K$ groups where each cluster has a mean, or prototype, and

| | *Virus: Visible Problems* | *Virus: Protect Yourself* | *Virus: Caught on Internet* | *Hacker: Target Home* | *Hacker: Target Others* |
|---|---:|---:|---:|---:|---:|
| Woman - Man | 0.02 | -0.03 | -0.01 | 0.01 | -0.07 |
| Independent - Democrat | -0.14 | -0.11 | -0.06 | -0.09 | -0.12 |
| Republican - Democrat | -0.08 | -0.16 | -0.04 | -0.04 | -0.09 |
| Republican - Independent | 0.05 | -0.06 | 0.01 | 0.05 | 0.02 |
| High School Grad - Some High School | 0.07 | -0.16 | 0.06 | 0.05 | 0.06 |
| College - High School Grad | -0.15 | -0.18 | -0.21 | -0.10 | -0.05 |
| Grad School - College | -0.05 | -0.16 | -0.24 | **-0.33** | -0.10 |
| Age 30-49 - 18-29 | 0.19 | -0.10 | 0.06 | 0.16 | 0.03 |
| Age 50-64 - 30-49 | -0.04 | -0.17 | -0.24 | 0.03 | 0.11 |
| Age 65 or over - 50-64 | -0.17 | -0.01 | -0.25 | -0.14 | -0.14 |
| Has Children - No Children | 0.10 | -0.04 | 0.03 | 0.11 | -0.01 |
| American Indian or Alaska Native - White | **0.41** | **0.56** | **0.57** | 0.23 | 0.26 |
| Asian or Pacific Islander - White | 0.08 | **0.81** | **0.37** | -0.02 | 0.26 |
| Black or African American - White | 0.14 | **0.71** | 0.28 | 0.18 | 0.12 |
| Hispanic or Latino - White | -0.00 | **0.54** | 0.15 | 0.11 | -0.05 |

**Table 4: Comparing Beliefs Across Demographic Groups. Each value is $g_\psi$, an estimate of the effect size of the difference, in units of standard deviation of the whole variable. $g_\psi$ is a generalization of Hedge's $g$. Positive values indicate that the group on the left agrees with the belief more than the group on the right. For example, in the row "College - High School Grad", the effect size for "Virus: Caught on Internet" is $-0.21$. This means that High School Grads agree more that you can catch a virus simply by browsing the Internet than College graduates do. An effect size larger than $0.10$ is small but worth comment, and an effect larger than $0.30$ is moderate to large.**

| | *Virus: Software* | *Virus: Be Careful* | *Hacker: Expert* | *Hacker: Be Careful* | *Hacker: Software* |
|---|---:|---:|---:|---:|---:|
| Woman - Man | -0.12 | 0.01 | -0.15 | -0.02 | -0.16 |
| Independent - Democrat | -0.06 | -0.09 | 0.08 | 0.02 | -0.01 |
| Republican - Democrat | 0.07 | -0.02 | 0.04 | 0.06 | 0.14 |
| Republican - Independent | 0.13 | 0.07 | -0.03 | 0.04 | 0.15 |
| High School Grad - Some High School | 0.28 | **0.39** | 0.25 | 0.23 | **0.30** |
| College - High School Grad | 0.10 | 0.01 | -0.01 | 0.05 | 0.16 |
| Grad School - College | -0.06 | 0.05 | -0.05 | -0.10 | -0.07 |
| Age 30-49 - 18-29 | **0.38** | 0.29 | 0.16 | **0.33** | **0.44** |
| Age 50-64 - 30-49 | 0.22 | 0.09 | -0.05 | 0.14 | 0.22 |
| Age 65 or over - 50-64 | 0.05 | 0.13 | 0.04 | -0.04 | 0.09 |
| Has Children - No Children | 0.09 | 0.05 | 0.01 | 0.03 | 0.02 |
| American Indian or Alaska Native - White | -0.21 | 0.01 | 0.29 | 0.08 | -0.20 |
| Asian or Pacific Islander - White | -0.02 | -0.09 | -0.07 | **-0.40** | -0.26 |
| Black or African American - White | -0.22 | -0.20 | 0.04 | -0.08 | **-0.37** |
| Hispanic or Latino - White | -0.11 | -0.16 | 0.07 | -0.07 | -0.16 |

**Table 5: Comparing Security Behaviors Across Demographic Groups. Each value is $g_\psi$, an estimate of the effect size of the difference, in units of standard deviation of the whole variable. $g_\psi$ is a generalization of Hedge's $g$. Positive values indicate that the group on the left engages in the behavior more frequently than the group on the right. An effect size larger than $0.10$ is small but worth comment, and an effect larger than $0.30$ is moderate to large.**

| *Cluster* | *Virus:<br>Visible<br>Problems* | *Virus:<br>Protect<br>Yourself* | *Virus:<br>Caught on<br>Internet* | *Group<br>Size* | *Virus<br>Behavior:<br>Software* | *Virus<br>Behavior:<br>Careful* |
|---|---|---|---|---|---|---|
| 1 | 4.32 | 2.07 | 3.77 | 680 | 4.50 | 4.53 |
| 2 | 4.34 | 3.71 | 3.99 | 406 | 4.39 | 4.40 |
| 3 | 3.56 | 2.45 | 2.80 | 790 | 4.14 | 4.23 |

**Table 6: K-Means Clustering of Participants based on their answers to questions about virus beliefs, with $K = 3$. The three beliefs (the left three columns) were clustered, and then means calculated for behaviors for each cluster (the right two columns). $K = 3$ was determined by examining a plot of variance explained and choosing the elbow.**

| *Cluster* | *Hacker:<br>Target<br>Home* | *Hacker:<br>Target<br>Others* | *Group<br>Size* | *Hacker<br>Behavior:<br>Expert* | *Hacker<br>Behavior:<br>Careful* | *Hacker<br>Behavior:<br>Software* |
|---|---|---|---|---|---|---|
| 1 | 3.95 | 3.73 | 792 | 3.26 | 4.22 | 4.30 |
| 2 | 4.71 | 4.66 | 368 | 3.72 | 4.51 | 4.53 |
| 3 | 2.23 | 1.73 | 43 | 3.48 | 4.10 | 4.22 |
| 4 | 3.35 | 2.94 | 673 | 3.15 | 4.03 | 4.01 |

**Table 7: K-Means Clustering of Participants based on their answers about hacker beliefs, with $K = 4$. The two beliefs (the left two columns) were clustered, and then means calculated for behaviors for each cluster (the right three columns). $K = 4$ was determined by examining a plot of variance explained and choosing the elbow.**

each participant is grouped into the cluster with the most similar mean across all the variables [27]. This method allows us to find common patterns of beliefs that might be non-linear in nature.

### *Clustering By Virus Beliefs.*

To begin, we clustered all participant according to their answers to the questions about virus beliefs. This clustering technique allows us to characterize "prototype" individuals for each cluster by examining the mean value for each measure. We first need to decide how many clusters to find. By examining a plot of variance explained by number of clusters (not shown) [27], we decided that we should look for $K = 3$ clusters of participants; this would explain over 50% of the variance in virus beliefs. Table 6 contains these results.

Clusters 1 and 2 are similar in many respects; individuals in both clusters strongly agree that viruses cause visible problems on home computers, and strongly agree that viruses can be caught by browsing the Internet. However, individuals in these two clusters disagree about whether you can protect yourself by using anti-virus and not downloading files. Cluster 1 disagrees with the belief that you can protect yourself from viruses; while Cluster 2 strongly agrees with it.

Cluster 1 has the highest self-reported compliance with both using virus software and being careful about viruses on the Internet. Cluster 2 still complies with all virus protection behaviors, though less so than Cluster 1.

Cluster 3 is different; individuals in this cluster weakly agree that viruses can cause visible problems for home computers, and weakly disagree that viruses can be caught simply from browsing the Internet. Individuals in this cluster also report the lowest use of anti-virus software and the least often behavior of being careful on the Internet. This is also the largest cluster. This suggests that this mental model – believing that you can't randomly catch viruses on the Internet and only slightly believing that viruses cause visible problems on home computers – is associated with the fewest security actions.

### *Clustering by Hacker Beliefs.*

We also clustered participants by their beliefs about hackers. Examining the plot of variance explained by number of clusters, we determined that the optimal number of clusters here would be $K = 4$. Table 7 shows the results of this clustering. Both beliefs tended to vary together, with Cluster 2 having the highest belief for both targeting home computers and targeting others, and Cluster 3 having the lowest belief in both.

Individuals in Clusters 1 and 2 agree with both beliefs about hackers. These people tended to do the most behaviors to protect themselves against hackers. Individuals in Cluster 3 disagree with both beliefs about hackers. And individuals in Cluster 4 are on the fence, neither agreeing nor disagreeing with the beliefs. The people in Cluster 4 actually report the fewest behaviors to protect themselves, and might represent a group that doesn't really think much about hackers.

## 7. LIMITATIONS

A limitation of this study is that we didn't measure actual actions taken by our participants; rather, we measured what participants were willing to say about their actions. These answers might not match actual actions for at least two reasons: 1) *Social Desirability Bias* [28], or 2) *Imperfect Recall.* Social desirability bias means that participants might intentionally answer incorrectly because they believe they should be taking that action, even if they aren't. They believe it is socially desirable to be seen taking that action. Imperfect recall means that participants might unintentionally answer incorrectly because they do not accurately remember their actions. This is often due to the fact that survey questions ask about general trends (e.g. “Don’t click on shady links on the Internet”) which are often aggregates of multiple individual events, and some of those individual events might be more salient (not clicking on a link believed to be shady) than others (clicking on a link you didn’t realize was shady).

We do not see social desirability bias as a problem in our dataset. There was interesting variation in people’s answers to the behavior questions, and those answers varied by belief. We suspect that which actions people see as socially desirable also depend on which folk models people believe. If you don’t think hackers attack home computers, then it isn’t socially desirable to protect yourself against them. Social desirability bias then works in our favor by emphasizing actions associated with a folk model and de-emphasizing actions that contrast with a folk model. This bias should strengthen our correlational results, but means our exact estimates of how many people undertake a given action might be off.

Imperfect recall is a problem in our data (and any survey). It definitely can add noise to the data; however, our large sample should allow us to distinguish signal from noise. But imperfect recall might not just be random noise; it might be biased in one direction or the other. This means that the absolute level of how much an action is taken might be incorrect, but relative measures — for example, comparing across demographic groups or clustering folk models — should still yield accurate comparisons.

Additionally, many of our questions were framed positively, which could increase the social desirability bias. We did this intentionally; we were trying to capture the horizontal differentiation of mental models expressed in Wash [36]. Mental models are often incomplete, and are not transitive; and they are sometimes self-contradictory. Even if a mental model includes a belief about a positive statement, it does not necessarily follow that the model also includes a disbelief in the equivalent negative statement. We generated questions based on statements of beliefs from Wash’s [36] findings, and inverting these statements to make them negative would change their meaning. Still, this framing could possibly be why the means of some of the scales are higher than expected. However, since we are drawing comparisons between groups rather than examining absolute responses, any bias due to positivity should be approximately equivalent across groups.

In order to get more accurate measures, we intend to directly measure security actions in future work to avoid these problems with self-reported measures.

## 8. DISCUSSION AND CONCLUSIONS

Accurate knowledge about computer security is very hard for everyday computer users to attain, because their decisions can be hard to execute correctly [38], may not lead to correct behaviors [34], and the outcomes of their behaviors are often not visible [37]. But, using the Internet means these users still have to act. In a sense, all Internet users have experience with making computer security decisions, because they make them so often. They just rarely know for sure if they are making the “correct” ones.

Most people struggle to learn from past experiences how to protect themselves from computer security threats [37]. For example, if a user delegates security protection to antivirus, how can he or she be sure the software is doing its job correctly? Users must trust that their actions, such as clicking on one link while not clicking on another, produce positive security outcomes that are difficult to see or verify. These beliefs about actions and outcomes form a mental model about causal relationships [17]: “If I do X (use antivirus), Y outcome will result (my computer will be protected)”. People incorporate lessons and analogies like “don’t go to the shady parts of the Internet or you’ll catch a virus” into beliefs about what causes problems [36, 30]. Many beliefs can exist side-by-side [21], are called upon when mental models relevant to the decision to be made are activated, and are associated with different behaviors. By focusing on variation in beliefs instead of amount of knowledge, we identify relationships between what everyday computer users are thinking and doing. These relationships suggest new ways to help users make better decisions beyond simply providing more knowledge.

Consistent with literature in other domains, we found a number of causal beliefs that are associated with self-reported security behaviors. Additionally, in a representative sample of the United States Internet-using population, we found that there are demographic differences in both beliefs about security, and security behaviors.

Less educated people are more likely to believe computers can catch viruses by casually browsing the Internet, but at the same time least likely to believe it is possible protect their computers from viruses and hackers. People with less than a high school degree are also least likely to report taking any kind of protective actions related to viruses or hackers. People with lower levels of general education are vulnerable because they they feel helpless, like there is nothing they can do to protect themselves.

Older people are much less likely to agree that casually browsing the Internet can give you a virus, and people with more years of education are less likely to believe hackers target home computers. Older people, and people with a high school education or greater report taking more protective actions. These people believe they can protect themselves, but often don’t think that they are a target.

These results suggest an interesting relationship between demographics, beliefs and behavior: younger and less educated Internet users are vulnerable in different ways than older and more educated users. This vulnerability likely arises because of differences in their beliefs.

Differences in beliefs make communicating with and educating users about security challenging. Emphasizing vulnerability and using scare tactics is unlikely to help younger or less education users, since they often don’t believe there is anything they can do about it. On the other hand, that may work for older adults, where teaching protective measures won’t work because they don’t believe they are a target.

An important characteristic of mental models is that many different, but related, causal beliefs can be held by a single person at the same time [36]. Our survey showed that people who strongly believe that viruses cause visible computer problems also strongly agree that viruses can be caught by browsing the Internet. But, some of these people believe they can take actions to protect themselves (n=406), while others do not (n=680). In addition, people who agree that hackers target other people instead of themselves also believe that hackers target home computers; these people do the most behaviors to protect themselves. Seeing yourself as a target isn’t necessary to undertake protective actions.

The relationship between mental model and behavior is not as straightforward as Wash [36] suggests.

Interestingly, people with weak beliefs about viruses ($n = 790$) and hackers ($n = 673$) also had weak beliefs about how they should protect themselves. They also reported the lowest amount of protection behaviors. It seems like having a strong belief about cause and effect — any cause and effect — may be related to taking protective actions. Interventions intended to influence behavioral outcomes should focus on users whose causal beliefs are weakest.

More-is-better measures of security knowledge do not capture the range of beliefs that real users possess. By characterizing beliefs, we identify groups of users that have different challenges in understanding computer security. This work suggests that different demographic segments of the population are likely to respond differently to persuasive and educational messages, and a one-size-fits-all education approach is inappropriate for computer security.

# 9. ACKNOWLEDGMENTS

Special thanks to Alex Hinck, Michelle Rizor, and Kami Vaniea for their work on the early versions of the survey. Thanks to everyone in the BITLab for providing feedback. This material is based upon work supported by the National Science Foundation under Grant Nos. CNS-1116544 and CNS-1115926.

# 10. REFERENCES

[1] I. Ajzen. From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl and J. Beckmann, editors, *Action Control: From Cognition to Behavior*, pages 11–39. Springer Berlin Heidelberg, Berlin, Heidelberg, Sept. 1985.

[2] E. Albrechtsen and J. Hovden. Improving information security awareness and behaviour through dialogue, participation and collective reflection. An intervention study. *Computers & Security*, 29(4):432–445, June 2010.

[3] C. Anderson and R. Agarwal. Practicing Safe Computing: A Multimethod Empirical Examination of Home Computer User Security Behavioral Intentions. *MIS Quarterly*, 34(3), September 2010.

[4] F. Asgharpour, D. Liu, and L. Camp. Mental models of computer security risks. In *Workshop on the Economics of Information Security (WEIS)*, 2007.

[5] R. Baker, S. Blumberg, J. Brick, M. Couper, M. Courtright, J. Dennis, D. Dillman, M. Frankel, P. Garland, R. Grovers, C. Kennedy, J. Krosnick, and P. Lavrakas. Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly*, 74(4):711–781, Winter 2010.

[6] D. Bandalos and S. Finney. Factor analysis: Exploratory and confirmatory. In G. R. Hancock and R. O. Mueller, editors, *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, chapter 8. Routledge, 2010.

[7] K. S. Button, J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, May 2013.

[8] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, second edition, 1998.

[9] R. D'Andrade. *The Development of Cognitive Anthropology*. Cambridge University Press, 2005.

[10] R. F. DeVellis. *Scale Development: Theory and Applications*. SAGE Publications, Inc, third edition, June 2011.

[11] D. Dillman. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Wiley, third edition, 2008.

[12] N. F. Doherty, L. Anastasakis, and H. Fulford. The information security policy unpacked: A critical study of the content of university policies. *International Journal of Information Management*, 29(6):449–457, Dec. 2009.

[13] P. Dourish, R. E. Grinter, J. Delgado De La Flor, and M. Joseph. Security in the wild: User strategies for managing security as an everyday, practical problem. *Personal and Ubiquitous Computing*, 8(6):391–401, 2004.

[14] S. Egelman and E. Peer. Scaling the Security Wall. In *ACM Conference on Human Factors in Computing (CHI)*, pages 1–10, Jan. 2015.

[15] Gallup. Party affiliation. `http://www.gallup.com/poll/15370/party-affiliation.aspx`, September 2014.

[16] A. Gelman and D. Weakliem. Of beauty, sex and power. *American Scientistq*, 97, 2009.

[17] E. Goldvarg and P. N. Johnson-Laird. Naive causality : a mental model theory of causal meaning and reasoning. *Cognitive Science*, 25(4):565–610, 2001.

[18] J. Gross and M. Rosson. Looking for Trouble: Understanding End-User Security Management. In *Proceedings of the 2007 Symposium on Computer Human interaction For the Management of information Technology*, pages 30–31, 2007.

[19] C. Herley. So Long , And No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *NSPW '09 Proceedings of the 2009 New Security Paradigms Workshop*, 2009.

[20] K. Hoban, E. Rader, R. Wash, and K. Vaniea. Computer security information in stories, news articles, and education documents. In *Poster at SOUPS*, Palo Alto, CA, July 2014.

[21] P. N. Johnson-Laird. Inaugural Article: Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 2010, Oct. 2010.

[22] W. Kempton. Two Theories of Home Heat Control. *Cognitive Science*, 10(1):75–90, 1986.

[23] R. Kline. *Beyond Significance Testing*. American Psychological Association, Washington DC, 2005.

[24] N. Kumar, K. Mohan, and R. Holowczak. Locking the door but leaving the computer vulnerable: Factors inhibiting home users' adoption of software firewalls. *Decision Support Systems*, 46(1):254–264, Dec. 2008.

[25] R. LaRose, N. J. Rifon, and R. Enbody. Promoting personal responsibility for internet safety. *Communications of the ACM*, 51(3):71–76, Mar. 2008.

[26] D. Lee, R. LaRose, and N. Rifon. Keeping our network safe: a model of online protection behaviour.

*Behaviour & Information Technology*, 27(5):445–454, Sept. 2008.

[27] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[28] D. L. Paulhus. Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, and L. S. Wrightsman, editors, *Measures of personality and social psychological attitudes*, chapter 2, pages 17–59. Academic Press, San Diego, CA, 1991.

[29] Pew Research Center's Internet and American Life Project. How americans go online. `http://www.pewinternet.org/2013/09/25/how-americans-go-online/`, September 25 2013.

[30] E. Rader, R. Wash, and B. Brooks. Stories as informal lessons about security. In *SOUPS '12: Proceedings of the Eighth Symposium on Usable Privacy and Security*. ACM, July 2012.

[31] R. Shillair, S. R. Cotten, H.-Y. S. Tsai, S. Alhabash, R. LaRose, and N. J. Rifon. Online safety begins with you and me: Convincing Internet users to protect themselves. *Computers in Human Behavior*, 48:199–207, July 2015.

[32] Symantec Corporation. Internet Security Threat Report, Volume 18, 2013.

[33] U.S. Census Bureau. Summary of population and housing characteristics, 2010 census of population and housing. `http://www.census.gov/prod/cen2010/cph-1-1.pdf`, January 2013.

[34] K. Vaniea, E. Rader, and R. Wash. Betrayed by updates: How negative experiences affect future security. In *Proceedings of the ACM Conference on Human Factors in Computing (CHI)*, Toronto, Canada, 2014.

[35] H. Varian. *Microeconomic Analysis*. W. W. Norton and Company, 1992.

[36] R. Wash. Folk models of home computer security. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, Redmond, WA, July 2010.

[37] R. Wash and E. Rader. Influencing mental models of security: a research agenda. In *Proceedings of the New Security Paradigms Workshop (NSPW)*, 2011.

[38] R. Wash, E. Rader, and K. Vaniea. Out of the loop: How automated software updates cause unintended security consequences. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, Palo Alto, CA, July 2014.

[39] Wikipedia. August 2014 celebrity photo leaks. `http://en.wikipedia.org/wiki/August_2014_celebrity_photo_leaks`, September 21 2014.

# APPENDIX

## A. INITIAL TRIAL

We conducted an initial survey to evaluate questions about security beliefs. We recruited participants using Amazon's Mechanical Turk. We paid $2 per survey, and required that participants had completed at least 500 HITs and had a 90% approval rate.

Table 8 shows the results of an initial factor analysis of the questions related to beliefs about viruses. From this, we extracted 18 questions representing 3 common factors.

Table 9 shows the results of an initial factor analysis of the questions related to beliefs about hackers. From this, we extracted 16 questions. Questions *hacker.1.3*, *hacker.1.4* and *hacker.1.5* are all very similar, and we decided to keep question *hacker.1.4*. In addition to the top 5 questions for each factor, we also kept question X because we thought it might be interesting.

## B. MAIN SURVEY

Tables 10, 11, 12, and 13 contain the final Exploratory Factor Analysis for the full, nationally representative sample. Participants were paid approximately $1.50 via Qualtrics for their participation in the survey.

| *ID* | *Question* | 1 | 2 | 3 |
|---|---|---|---|---|
| virus.1.1 | I closely monitor what I download from the Internet | | | |
| virus.1.2 | I can tell if a website isn't safe | | 0.511 | |
| virus.1.3 | Downloading things from popular websites is safe | | | |
| virus.1.4 | I know if I have a virus | | | |
| virus.1.5 | It is extremely difficult for Macintosh computers to get viruses | | | |
| virus.1.6 | I can get a virus just from going to a website | | | |
| virus.1.7 | Anti-virus software always detects viruses | | | 0.761 |
| virus.1.8 | The only way to get a virus is by downloading something | | | 0.593 |
| virus.1.9 | A website is shady when there are a lot of pop-ups | | | |
| virus.1.10 | Being aware of what websites I go to will help me avoid getting a virus | | 0.555 | |
| virus.1.11 | Viruses are undetected if no anti-virus software is installed | | | |
| virus.1.12 | When I download something from the Internet, I probably won't get a virus | | | |
| virus.1.13 | Limiting my Internet use will help me avoid getting a virus | | | |
| virus.2.1 | Playing games on the Internet makes it easy to get a virus | | | |
| virus.2.2 | Using an Apple computer means you can't get a virus | | | 0.540 |
| virus.2.3 | Purchased anti-virus software is better than free anti-virus software | | | |
| virus.2.4 | Not paying attention to cookies can result in getting a virus | 0.524 | | |
| virus.2.5 | Blocking pop-ups makes it very difficult to get a virus | | | 0.529 |
| virus.2.6 | You can't get a virus if you never download things from the Internet | | | |
| virus.2.7 | Being careful with what you click on while browsing the Internet makes it much more difficult to catch a virus | | | |
| virus.2.8 | You probably won't catch a virus if you do not use the Internet frequently | | | |
| virus.2.9 | Turning off your computer when you are not using it helps protect against viruses | | | |
| virus.2.10 | You cannot get a virus if you keep your anti-virus software up to date | | | 0.522 |
| virus.3.1 | Strange emails will give you a virus | 0.520 | | |
| virus.3.2 | Downloads from the Internet will give you a virus | 0.584 | | |
| virus.3.3 | Merely visiting the wrong webpages will give you a virus | 0.617 | | |
| virus.3.4 | A virus can be caught and spread automatically without you doing anything | | | |
| virus.3.5 | Watching pornography on the Internet will give you a virus | 0.695 | | |
| virus.3.6 | Clicking on advertisements will give you a virus | 0.602 | | |
| virus.3.7 | You will get a virus if someone hacks into the your computer and installs a virus | | | |
| virus.3.8 | You can get a virus if you actively click on a link on the Internet | | | |
| virus.3.9 | You will catch a virus from randomly searching for things on the Internet | | | |
| virus.4.1 | Causes computers to crash | | 0.595 | |
| virus.4.2 | Displays images such as skulls and crossbones every time the computer turns on | | | |
| virus.4.3 | Causes annoying problems | | 0.654 | |
| virus.4.4 | Downloads pornography | | | |
| virus.4.5 | Erases important files on the computer | | 0.609 | |
| virus.4.6 | Steals personal and/or financial information | | 0.606 | |
| virus.4.7 | Kicks me out of applications that are running | | 0.525 | |
| | Cronbach's $\alpha$ | 0.807 | 0.796 | 0.759 |
| | Variance Explained | 0.105 | 0.100 | 0.087 |
| | Cumulative Variance Explained | 0.105 | 0.205 | 0.291 |

**Table 8: Exploratory Factor Analysis (using maximum likelihood factor analysis) of virus questions from initial trial of N=149 participants from Mechanical Turk. Loadings are the result of varimax rotation. Loadings $< 0.5$ were removed. Three factors were chosen based on the elbow of the Scree plot. We focused on the top 6 questions from each of the 3 factors for future analysis. Cronbach's $\alpha$ is for the top 6 questions in each factor.**

| *ID* | *Question* | 1 | 2 | 3 |
|---|---|---|---|---|
| hacker.1.1 | A hacker makes a record of everything on the computer | 0.663 | | |
| hacker.1.2 | The hacker intentionally puts viruses on the computer | 0.713 | | |
| hacker.1.3 | A hacker steals personal and financial information | 0.686 | | |
| hacker.1.4 | A hacker sells personal information to other hackers | 0.634 | 0.540 | |
| hacker.1.5 | A hacker works with other hackers to steal personal and financial information | 0.643 | 0.545 | |
| hacker.1.6 | A hacker breaks stuff on the computer | 0.548 | | |
| hacker.1.7 | A hacker installs monitoring software on the computer | 0.707 | | |
| hacker.1.8 | A hacker watches what you are doing on your computer | 0.697 | | |
| hacker.1.9 | A hacker sells personal and financial information to criminals | 0.642 | 0.528 | |
| hacker.2.1 | Hackers choose targets randomly | | | |
| hacker.2.2 | Hackers target home computer users | | | |
| hacker.2.3 | Hackers target people with weak computer security | | | |
| hacker.2.4 | Hackers target large businesses | | | 0.703 |
| hacker.2.5 | Hackers target rich and important people | | | 0.820 |
| hacker.2.6 | Hackers are choose their victims based on exploiting immediate circumstances | | | |
| hacker.2.7 | Hackers target banks | | | 0.529 |
| hacker.2.8 | Hackers target the upper class | | | 0.740 |
| hacker.2.9 | Hackers do not target anyone specifically | | | |
| hacker.2.10 | Hackers target large databases | | | 0.553 |
| hacker.3.1 | Hackers only target really important people; therefore, I do not need to protect myself | | | |
| hacker.3.2 | Staying away from unfamiliar websites will protect me from hackers | | | |
| hacker.3.3 | It is important to shut off the computer when it is not in use to avoid being hacked | | | |
| hacker.3.4 | Only provide personal information to websites you trust to avoid being hacked | | | |
| hacker.3.5 | Install anti-virus software to keep hackers from breaking in to the computer | | | |
| hacker.3.6 | Always sign out of accounts and websites when you are done using them to avoid being hacked | | | |
| hacker.3.7 | Use strong passwords (includes numbers, symbols, and upper and lowercase letters) to prevent hackers from breaking in | | | |
| hacker.3.8 | Preinstalled firewall prevents hackers from breaking in | | | |
| hacker.3.9 | There is no way to protect myself from being hacked | | | |
| hacker.3.10 | Don't check your bank account online to prevent being hacked | | | |
| hacker.3.11 | Shop in stores instead of online to avoid being hacked | | | |
| hacker.4.1 | Hackers are college-age technology-savvy students | | | |
| hacker.4.2 | Anyone can be a hacker | | | |
| hacker.4.3 | Hackers are lonely college students | | | |
| hacker.4.4 | Hackers are professional criminals | | 0.630 | |
| hacker.4.5 | Hackers are members of organized crime | | 0.657 | |
| hacker.4.6 | Hackers are a type of criminal | | 0.726 | |
| hacker.4.7 | Hackers work with other criminals | | 0.866 | |
| hacker.4.8 | Hackers have no morals | | | |
| hacker.4.9 | Hackers are mischevious | | | |
| | Cronbach's $\alpha$ | 0.863 | 0.851 | 0.847 |
| | Variance Explained | 0.128 | 0.118 | 0.077 |
| | Cumulative Variance Explained | 0.128 | 0.246 | 0.323 |

**Table 9: Exploratory Factor Analysis (using maximum likelihood factor analysis) of hacker questions from initial trial of N=149 participants from Mechanical Turk. Loadings are the result of varimax rotation. Loadings $< 0.5$ were removed. Three factors were chosen based on the elbow of the Scree plot. We focused on the top 5 questions from each of the 3 factors for future analysis. Cronbach's $\alpha$ is for the top 5 questions in each factor.**

| *ID* | *Question* | 1 | 2 | 3 |
|---|---|---|---|---|
| virus.1 | Watching pornography on the Internet will give you a virus | | | 0.504 |
| virus.2 | Merely visiting the wrong webpages will give you a virus | | | 0.539 |
| virus.3 | Clicking on advertisements will give you a virus | | | 0.649 |
| virus.4 | Downloads from the Internet will give you a virus | | | 0.562 |
| virus.5 | Not paying attention to cookies can result in getting a virus | | | |
| virus.6 | Strange emails will give you a virus | 0.476 | | 0.454 |
| virus.7 | Anti-virus software always detects viruses | | 0.722 | |
| virus.8 | The only way to get a virus is by downloading something | | 0.663 | |
| virus.9 | Using an Apple computer means you can't get a virus | | 0.530 | |
| virus.10 | Blocking pop-ups makes it very difficult to get a virus | | 0.556 | |
| virus.11 | You can't get a virus if you keep your anti-virus software up to date | | 0.756 | |
| virus.12 | You can't get a virus if you never download things from the Internet | | 0.626 | |
| virus.13 | A virus causes annoying problems | 0.618 | | |
| virus.14 | A virus causes computers to crash | 0.687 | | |
| virus.15 | A virus erases important files on the computer | 0.577 | | |
| virus.16 | A virus steals personal and/or financial information | 0.544 | | |
| virus.17 | Being aware of what websites I go to will help me avoid getting a virus | 0.482 | | |
| virus.18 | A virus kicks me out of applications that are running | 0.484 | | |
| | Variance Explained | 0.155 | 0.151 | 0.109 |
| | Cumulative Variance Explained | 0.155 | 0.307 | 0.416 |

**Table 10: Exploratory Factor Analysis (using maximum likelihood factor analysis) of virus questions from the full survey of N=1993 participants sampled via Qualtrics. Loadings are the result of varimax rotation. Loadings $<0.4$ were removed. EFA was used in a confirmatory manner, and three factors were chosen based the previous trials. Question 6 loaded on multiple factors and was removed. Question 5 did not load on any factors and was removed. We constructed scales out of up to 5 questions (highest loaded) for each factor.**

| *ID* | *Question* | 1 | 2 |
|---|---|---|---|
| hacker.1 | Hackers work with other criminals | 0.504 | |
| hacker.2 | A hacker intentionally puts viruses on the computer | 0.656 | |
| hacker.3 | Hackers are members of organized crime | | |
| hacker.4 | Hackers are professional criminals | 0.547 | |
| hacker.5 | Hackers have no morals | 0.599 | |
| hacker.6 | Hackers are mischievous | 0.545 | |
| hacker.7 | A hacker watches what you are doing on your computer | 0.680 | |
| hacker.8 | A hacker makes a record of everything on the computer | 0.628 | |
| hacker.9 | A hacker installs monitoring software on the computer | 0.605 | |
| hacker.10 | A hacker breaks stuff on the computer | | |
| hacker.11 | Hackers target home computer users | 0.612 | |
| hacker.12 | Hackers target rich and important people | | 0.717 |
| hacker.13 | Hackers target large businesses | | 0.706 |
| hacker.14 | Hackers target the upper class | | 0.687 |
| hacker.15 | Hackers target banks | | 0.617 |
| hacker.16 | Hackers target large databases | | 0.606 |
| | Variance Explained | 0.255 | 0.197 |
| | Cumulative Variance Explained | 0.255 | 0.452 |

**Table 11: Exploratory Factor Analysis (using maximum likelihood factor analysis) of hacker questions from the full survey of N=1993 participants sampled via Qualtrics. Loadings are the result of varimax rotation. Loadings $<0.5$ were removed. EFA was used in a confirmatory manner. Three factors were originally extracted, but many indicators suggested a poor fit. Instead, we extracted two factors. We constructed scales out of up to 5 questions (highest loaded) for each factor.**

| *ID* | *Question* | 1 | 2 |
|---|---|---|---|
| virus.prevent.1 | Check anti-virus software to make sure it is up to date | 0.816 | |
| virus.prevent.2 | Regularly scan the computer with anti-virus software | 0.797 | |
| virus.prevent.3 | Use anti-virus software | 0.751 | |
| virus.prevent.4 | Use security software such as firewall | 0.637 | |
| virus.prevent.5 | Avoid downloading anything without knowing what exactly is being downloaded | | 0.728 |
| virus.prevent.6 | Be aware of what websites you visit | | 0.726 |
| virus.prevent.7 | Avoid clicking on email attachments from people you do not know | | 0.684 |
| virus.prevent.8 | Block pop-ups | | 0.500 |
| | Variance Explained | 0.337 | 0.288 |
| | Cumulative Variance Explained | 0.337 | 0.625 |

**Table 12: Exploratory Factor Analysis (using maximum likelihood factor analysis) of questions about protection behaviors from viruses from the full survey of N=1993 participants sampled via Qualtrics. Loadings are the result of varimax rotation. Loadings $< 0.4$ were removed. EFA was used in a confirmatory manner, and two factors were chosen based the previous trials.**

| *ID* | *Question* | 1 | 2 | 3 |
|---|---|---|---|---|
| hacker.prevent.1 | Use some pre-existing security software such as firewall | | | 0.613 |
| hacker.prevent.2 | Disable scripting on emails | 0.865 | | |
| hacker.prevent.3 | Back up your information on an external hard-drive, network, or server | 0.465 | | |
| hacker.prevent.4 | Scan your computer regularly with anti-virus software | | | 0.631 |
| hacker.prevent.5 | Avoid clicking on attachments | | 0.555 | |
| hacker.prevent.6 | Be careful downloading software from the Internet | | 0.745 | |
| hacker.prevent.7 | Disable scripting on websites | 0.837 | | |
| hacker.prevent.8 | Update patches regularly | 0.454 | | |
| hacker.prevent.9 | Always sign out of accounts when you are done with that website | | 0.454 | |
| hacker.prevent.10 | Use good passwords (good passwords include uppercase and lowercase letters, numbers, and symbols) | | 0.579 | |
| | Variance Explained | 0.213 | 0.187 | 0.140 |
| | Cumulative Variance Explained | 0.213 | 0.400 | 0.539 |

**Table 13: Exploratory Factor Analysis (using maximum likelihood factor analysis) of questions about protection behaviors from hackers from the full survey of N=1993 participants sampled via Qualtrics. Loadings are the result of varimax rotation. Loadings $< 0.4$ were removed. EFA was used in a confirmatory manner, and three factors were chosen based the previous trials.**