Supporting Security and Privacy Decisions With Data

Blase Ur
Carnegie Mellon University

July 29, 2015

Abstract

Despite decades of research into developing abstract security advice and improving interfaces, users still struggle to make security and privacy decisions. In particular, users often make security and privacy decisions that they are unsure about, that are based on misunderstandings of reality, or that do not reflect their preferences. Prior work suggests that a major cause of these problems is that users do not have the necessary contextual information about themselves and about greater ecosystems to make decisions.

To better support users' decisions, I propose building just-in-time data about a user's own behaviors and situations into security and privacy interfaces. When considered relative to data about greater ecosystems, a single user's own data can help the user make decisions that are objectively more secure or private, that he or she feels more confident about, that reflect a greater awareness of risks, and that better match the user's preferences.

I will examine this premise through one security case study and two privacy case studies. First, I will test the effectiveness of giving users feedback on precisely what they are doing wrong in creating a password. This approach leverages data I have collected through detailed analyses of both the semantic structure and guessability of large data sets of passwords, in addition to my studies of password-strength meters. Second, to counteract users' privacy misunderstandings related to online tracking and help them make privacy decisions that better match their preferences, I will examine the impact of visualizing different abstractions of how a user's own web browsing has been tracked. These visualizations rely on "tracking the trackers," but also build on my qualitative understanding of how users perceive privacy tradeoffs in the context of online behavioral advertising. Third, I will provide average consumers an interactive database that merges data I have collected about over 6,000 U.S. financial institutions' privacy practices with data about those institutions' branch locations. Using this interactive database, I will test whether surfacing this large-scale, comparative privacy information impacts users' willingness to consider switching banks, as well as how a willingness to find a more privacy-protective bank interacts with the logistical barriers of actually switching.

Thesis statement: **The objective of this thesis is to demonstrate how integrating just-in-time data about a user's own behaviors and the state of the greater ecosystem into security and privacy interfaces can better equip users to make security and privacy decisions. I demonstrate this idea through case studies in helping users create more secure passwords, make privacy decisions about third-party online tracking, and compare the privacy practices of financial institutions.**

Chapter 1

Committee Members

Lorrie Faith Cranor (Chair), Carnegie Mellon University

Alessandro Acquisti, Carnegie Mellon University

Lujo Bauer, Carnegie Mellon University

Jason Hong, Carnegie Mellon University

Michael K. Reiter (External), UNC Chapel Hill

Chapter 2

Introduction

When average users need to make a security or privacy decision, they are often not equipped to do so. When asked to make a password, the average user has only a vague, and sometimes incorrect, idea of what characteristics make a password hard to guess [145]. Often, the user does not know how password-guessing attacks work [161], nor how to navigate the greater ecosystem around passwords [14, 49]. Similarly, most average users are confused [144] about the complex mechanics behind online behavioral advertising (OBA), in which advertisers track a user's web browsing for the purpose of targeting advertisements based on the individual user's interests. Average users generally prefer receiving relevant advertising, yet are unsure how to weigh that benefit with their privacy concerns [144]. Similarly, if an average consumer wishes to do business with a privacy-protective financial institution, it has been unclear where he or she should begin.

In this thesis, I will work towards better supporting users as they make these decisions. This support will take the form of just-in-time information distilled automatically from data collected about the user's own behaviors and situations, as well as greater security and privacy ecosystems. I hypothesize that when considered relative to data about greater ecosystems, a single user's own data can help the user make decisions that are objectively more secure or private, that he or she feels more confident about, that reflect a greater awareness of risks, and that better match the user's preferences.

I investigate my hypotheses through case studies in providing data-driven decision support to help users make more secure passwords, make privacy decisions about third-party tracking for online behavioral advertising, and find a privacy-protective financial institution. In Chapter 3, I survey related work on data-driven decision support in security and privacy, as well as in the application domains of my case studies. My proposed thesis encompasses eight related studies, four of which have already been completed. I outline both the completed and proposed studies in Chapter 4. Finally, in Chapter 5, I provide an outline of the chapters in my thesis and a timeline for its completion.

Chapter 3

Related work

I outline related work in four steps. First, in Section 3.1, I give an overview of data-driven feedback for helping users make security and privacy decisions, which is most closely related to my thesis. I then provide domain-specific overviews of work related to passwords (Section 3.2) and online behavioral advertising (Section 3.3). Finally, I survey related work on privacy notices, with a focus on the U.S. financial industry, in Section 3.4.

3.1 Data-driven Feedback

A handful of other researchers have also evaluated the effectiveness of using personalized examples derived from users' own data or the overall ecosystem in helping users make privacy and security decisions. These studies rely on different techniques and focus on different application areas than my work. Furthermore, a number of these techniques do not present this information to users "just in time" as they are about to make a security or privacy decision, in contrast to the focus of my thesis.

Overall, providing users information about privacy or security can impact their decisions. For example, Tsai et al. found that consumers will pay a premium price to make purchases from more privacy-protective businesses when information about privacy is made accessible to consumers [141].

Much of the research on data-driven support for users in making security and privacy decisions has centered on smartphones. For example, Almuhimedi et al. found that showing users how frequently different smartphone apps access sensitive data can nudge users to restrict apps' access to this information [3]. That project builds on work by Harbach et al., who demonstrated that personal examples of the data accessible to smartphone apps help users understand otherwise abstract smartphone permission requests [60]. Similarly, Balebako et al. demonstrated that summary visualizations and just-in-time notices of smartphone privacy leakages help correct users' misconceptions about data sharing [7]. Their tool collects data about these leaks using the Taintdroid platform [42]. In a slightly different focus area, Consolvo et al. proposed bringing transparency to the information that is leaked unencrypted over wi-fi networks by showing this unencrypted data to users [27].

Some researchers have investigated the use of data-driven methods and feedback during the password-creation process, but using different types of data than I do, and often using this data in a less directed manner. For decades, researchers have suggested providing users proactive feedback based on the password they are typing [9, 11]. Two studies have investigated the underlying premise of whether meters impact user behavior in creating passwords. Complementary to my own online study of password-strength meters [143] (Section 4.1 of this thesis), Egelman et al. conducted a laboratory study of password-strength meters, also finding that the presence of a meter resulted in stronger passwords, yet only for high-value accounts [41].

The amount of data that drives this feedback on passwords has varied, however. Most websites currently rate passwords using length and character-class heuristics [33]. In the academic community, researchers have suggested using peer pressure by ranking passwords comparatively to other users [131] or estimating password strength using Markov models [22]. In this same vein, the Telepathwords project [75] uses large corpora of leaked passwords and dictionaries to guess what the user might type next. In contrast to my work, neither the scheme based on Markov models nor the Telepathwords scheme provides directed feedback based on the semantic characteristics or explicit guessability of the passwords. Furthermore, neither scheme directly takes into account users' misperceptions of password security.

Only a few projects have applied data-driven methods to privacy decisions regarding online tracking or online behavioral advertising. Closely related to my work, Wills and Zeljkovic examined data-driven decision support in privacy decisions related to online behavioral advertising. They prototyped a tool that examines browser history to help a user understand when their data has been tracked [158]. Their prototype presents a table of websites a user has visited, along with a list of the third-parties tracking on each site. While I start from a similar initial premise, I instead show what each advertising company knows about the user, visualizing a detailed list of the individual webpages the company has seen the user visit, as well as estimating what inferences could be made about the user's interests based on that set of visited pages and companies' own privacy dashboards. Angulo et al. have proposed using a plugin to visualize similar privacy leaks, such as the user's name and email address, in the browser [4]. As I discuss further in Section 4.7, software tools like Ghostery and Disconnect present data on what tracking is occurring at the current moment, but they do not collect longitudinal information. In contrast, Mozilla Lightbeam collects longitudinal information, but uses this information to display a graph of how different third-party companies and first-party websites are connected.

Other projects have used semi-automated methods or crowdsourcing to gather data about privacy and then present this data to users. For instance, Liu et al.'s Privacy Grade project crowdsources analysis of the permissions requested by different smartphone apps [91, 92]. They assign each app a "privacy grade" based on the appropriateness of the permissions the app requests, as judged by crowdworkers. A group of CMU researchers aims to provide users with similarly succinct summaries of privacy information, but for full-length privacy policies rather than smartphone app permissions [123]. They aim to do so by relying on both natural language processing and crowdsourcing. Projects like "Terms of Service; Didn't Read" (TOS;DR) have already used crowdsourcing on a small scale to put information about companies' privacy policies and terms of service into a standardized, usable format [138]. In the rare cases when machine-readable privacy information is available, as was the case for some websites using the P3P standard, crowdsourcing is not necessary [40].

In a number of other application domains within privacy and security, collected data drives decision making by software programs, rather than informing a user's decision. For instance, data-driven techniques are popular in the detection of phising messages [1,10,104], intrustion detection systems, and the identification of malware. Because such schemes do not directly visualize this information for users, they are out of scope of my thesis.

3.2 Passwords

In this section, I cover additional related work on passwords. I first focus on security metrics and approaches to cracking passwords. Afterwards, I focus on studies of user behavior in creating passwords. My data-driven approach to passwords examines both security metrics and user behaviors.

3.2.1 Password Security Metrics

While estimated entropy was once a leading password strength metric [19], it does not reflect what portion of a set can be cracked easily [12, 71, 154]. Two main classes of metrics have emerged in its place: statistical metrics and parameterized metrics. Both classes focus on *guessability*, the number of guesses needed by an adversary to guess a given password or a fraction of a set.

Statistical metrics are particularly valuable for examining password sets as a whole. For example, Bonneau introduced partial guessing metrics [12] for estimating the number of guesses required for an idealized attacker, who can perfectly order guesses, to guess a given fraction of a set. Since password distributions are heavy-tailed, very large samples are required to determine a set's guessability accurately.

Parameterized metrics instead investigate guessability under a cracking algorithm and training data [13, 71, 154]. These metrics thus model an adversary using existing tools, rather than an idealized attack, though the metric is only as good as the chosen algorithm and training data. Parameterized metrics can also be used to compare password sets without fully running the algorithm [93].

In contrast to statistical metrics, parameterized metrics have two important properties. First, they estimate the guessability of each password individually. Estimating guessability per-password is important for security audits (e.g., identifying weak passwords) and to provide feedback to a user about a password she has created. This latter promises to become more widespread as proactive feedback tools move from length-and-character-class heuristics [33] to data-driven feedback [22, 75]. Second, parameterized metrics aim to estimate security against real-world, rather than idealized, attacks. Researchers previously assumed automated techniques approximate real-world attackers [71, 154]; the "Biases" study that is part of my thesis (Section 4.2) is the first to test this assumption against attacks by professionals.

Parameterized metrics have been used to measure password strength in a number of previous studies [22, 32, 34, 43, 50, 71, 76, 93, 98, 119, 127, 143, 150, 154, 160]. While there are many different methods for cracking passwords, time and resource constraints lead many researchers to run only a single algorithm per study. However, prior to the "Biases" study [146]

(Section 4.2), it was an open question whether this strategy accurately models real-world attackers, or whether choosing a different algorithm would change a study's results.

Throughout this thesis proposal, we refer to the *guess number* of a password, or how many guesses a particular parameterized algorithm took to arrive at that password. Because the algorithm must be run or simulated, there is necessarily a *guess cutoff*, or maximum guess after which remaining passwords are denoted "not guessed."

3.2.2 Types of Guessing Attacks

The selection of a guessing attack to model is crucial to analyzing password guessability. Researchers have long investigated how to guess passwords. A handful of studies [25, 34, 119] have compared the aggregate results of running different cracking approaches. Other studies have compared results of running different cracking approaches based on guess numbers [24, 37, 93]. In this section, I highlight four major types of attacks.

Brute-force and mask attacks Brute-force attacks are conceptually the simplest. They are also inefficient and therefore used in practice only when targeting very short or randomly generated, system-assigned passwords.

Mask attacks are directed brute-force attacks in which password character-class structures, such as "seven lowercase letters followed by one digit" are exhausted in an attacker-defined order [133]. While this strategy may make many guesses without success, mask attacks can be effective for short passwords as many users craft passwords matching popular structures [79, 140]. Real-world attackers also turn to mask attacks after more efficient methods exhaust their guesses.

Probabilistic context-free grammar In 2009, Weir et al. proposed using a probabilistic context-free grammar (PCFG) with a large training set of passwords from major password breaches [148] to model passwords and generate guesses [155]. They use training data to create a context-free grammar in which non-terminals represent contiguous strings of a single character class. From the passwords observed in its training data, PCFG assigns probabilities to both the structure of a password (e.g., *monkey99* has the structure {*six letters*}{*two digits*}) and the component strings (e.g., "99" will be added to the list of two-digit strings it has seen). A number of research studies [24, 34, 38, 71, 93, 98, 127, 143, 154, 160] have used PCFG or a close variant to compute guessability.

Kelley et al. proposed other improvements to Weir et al.'s PCFG algorithm, like treating uppercase and lowercase letters separately and training with structures and component strings from separate sources [71]. Because they found these modifications improved guessing effectiveness, I incorporate their improvements in my tests. In addition, multiple groups of researchers have proposed using grammatical structures and semantic tokens as PCFG non-terminals [119, 150]. More recently, Komanduri proposed a series of PCFG improvements, including supporting hybrid structures and assigning probabilities to unseen terminals [74].

Markov models Narayanan and Shmatikov first proposed using a Markov model of letters in natural language with finite automata representing password structures [109]. Castelluccia et al. used a similar algorithm for password meters [22]. John the Ripper and Hashcat offer simple Markov modes in their cracking toolkits as well.

Recently, Duermuth et al. [37] and Ma et al. [93] independently evaluated many variations of Markov models and types of smoothing in cracking passwords, using large sets of leaked passwords for training. Both groups compared their model with other probabilistic attacks, including Weir et al.'s original PCFG code, finding particular configurations of a Markov model to be more efficient at guessing passwords for some datasets.

Mangled wordlist attacks Perhaps the most popular strategy in real-world password cracking is the dictionary attack. First proposed by Morris and Thompson in 1979 [105], modern-day dictionary attacks often combine *wordlists* with *mangling rules*, string transformations that modify wordlist entries to create additional guesses. Wordlists usually contain both natural language dictionaries and stolen password sets. Typical mangling rules perform transformations like appending digits and substituting characters [114, 134].

Many modern cracking tools, including John the Ripper [115], Hashcat [132], and PasswordsPro [65], support these attacks, which I term *mangled wordlist attacks*. The popularity of this category of attack is evident from these tools' wide use and success in password-cracking competitions [78, 116]. Furthermore, a number of research papers have used John the Ripper, often with the default mangling rules [24, 32, 33, 43, 50, 59, 76, 159] or additional mangling rules [34, 38, 160].

Expert password crackers, such as those offering forensic password-recovery services, frequently perform a variant of the mangled wordlist attack in which humans manually write, prioritize, and dynamically update rules [55]. I term these manual updates to mangling rules *freestyle rules*. My co-authors and I evaluate guessability using off-the-shelf tools relying on publicly available wordlists and mangling rules. We also contract a password recovery industry leader to do the same using their proprietary wordlists and freestyle rules.

3.2.3 Understanding Password Composition

A number of researchers have surveyed users about password creation. Unsurprisingly, most passwords contain meaningful elements [162], such as the name of the user or a relative, geographic locations, and names of sports teams [84, 103].

Over time users' passwords appear to be becoming longer and more complex. In 1997, Zviran and Haga found an average password length of six characters, with 14% including digits and less than 1% including symbols [162]. By the mid-2000s, the average length had increased to eight characters, at least 40% used digits, and 3-16% used symbols [18]. A survey by von Zezschwitz et al. asking users to compare passwords they used at various times similarly finds that users are now choosing stronger passwords than they used to [152].

A number of researchers have analyzed passwords obtained from public leaks of password data, including sets from RockYou (2009), Sony and Gawker (2011), LinkedIn (2012), and Adobe (2013). Common passwords in these sets include *password*, *password1*, and *123456* [35, 96, 120]. For Sony and Gawker, 14% of passwords included people's names, 25% included dictionary words, and 8% included place names, with the most common modification being the addition of digits [63]. Names, dictionary words, keyboard patterns, and numbers were similarly prevalent in RockYou [35, 151].

Several studies have considered factors affecting password selection. Users choose passwords more carefully (and reuse them less frequently) for accounts they perceive to have higher value [48, 111]. Others find correlations with demographics [12, 98] and with users' annoyance with the password creation process [98]. Zhang et al. found that up to 17% of passwords created to replace an expiring prior password can be broken in under five guesses if the old password is known. Common transformations included incrementing a number and replacing one symbol with another [160].

3.2.4 Linguistic and Semantic Properties of Passwords

Some recent work has examined passwords explicitly from a linguistic perspective. Bonneau and Shutova searched for predictable Amazon payphrases, which must contain two or more words but no digits or symbols [16]. They found that many expected phrases from the arts, sports, and geography had been chosen, and that noun phrases were common. Rao et al. investigated grammatical structures within long passwords as an aid to cracking [119]. However, their mostly manual analysis examines a sample of only 144 passwords and does not consider substitutions or interstitial digits and symbols.

Jakobsson and Dhiman built an automated parser to study how dictionary words are modified as part of passwords, with particular focus on concatenated words, "leet" substitutions, and misspellings [66]. In my "Art" study (Section 4.3), I instead crowdsourced this process to reduce false positives and false negatives.

Veras et al. automatically separated passwords into linguistic chunks (words) based on large corpora of text, tagged each word with its part of speech, and further classified nouns and verbs semantically using a lexical database of English-language concepts [150]. They found nouns to be overwhelmingly popular among parts of speech, while top semantic categories included names, cities, and words related to love.

In contrast to this past work, in the "Art" study (Section 4.3), I instead use a combination of crowdsourcing and automated techniques to reverse engineer passwords, providing greater accuracy in creating the abstract semantic representation of a password. I then perform analyses on a number of different levels to provide a holistic understanding of user behavior in constructing passwords.

3.2.5 The Password Ecosystem

The overall password ecosystem has also been a major area of investigation. A number of studies have investigated password-management practices [2, 54, 58, 64, 130] and how users respond to password-creation requirements [117,153]. Researchers have studied how users recall multiple passwords, including text passwords and graphical passwords [23]. Researchers have also explored automatically increasing password strength by adding random characters, which study participants could shuffle until arriving at a configuration they liked. The authors found that inserting two random characters increased security, yet adding more characters hurt usability [50].

More recently, Stobert and Biddle interviewed 27 participants about their strategies for password management and usage. Participants had an average of 27 accounts and five passwords. They often made tradeoffs between following password advice and expending too much effort [135].

Other studies have focused on password reuse [32], economic analyses of the passwor ecosystem [14], internationalization issues in the administration of password systems [17] and designing systems that make offline attacks more easily detectable [69]. Finally, a handful of studies have examined the memorability of system-assigned passwords, finding that humans can learn long, random secrets through spaced repetition [15] and that system-assigned passphrases are not significantly more memorable than system-assigned passwords [126].

3.3 Privacy and Online Behavioral Advertising

The second case study in my thesis covers online behavioral advertising (OBA), which is the targeting of advertisements based on a user's web browsing. OBA remains a major source of privacy invasion [62, 97].

Online advertisers track users as they traverse the Internet, constructing profiles of individuals to enable targeted advertising based on each user's interests. Targeting advertisements can provide benefit to advertisers, helping advertisers find users who are more likely to be interested in the advertised product [45]. Furthermore, since advertising networks can charge higher prices to serve targeted advertising rather than general ads, OBA is "a way to support the websites and products you care about" and may reduce the number of ads consumers see that are not relevant to their interests [107].

I begin by explaining how OBA works from a technical perspective. I then discuss notice and choice mechanisms for OBA in the United States. Finally, I discuss prior studies of consumer sentiment towards advertising and OBA. These studies, conducted by both academics and the advertising industry, have found a range of positive and negative consumer attitudes about OBA. The interview results I report in the "Smart, Useful" study (Section 4.6) provide deeper insight into the genesis and interrelationship between the attitudes reported in prior surveys.

3.3.1 The Mechanics of OBA

Since the details of data collection and usage by specific advertising networks for the purpose of OBA can be considered trade secrets, the exact mechanics of how OBA works are generally not public. Nevertheless, basic mechanisms for enabling tracking, as well as methods currently used in the wild, have been examined in the literature.

In general, the goal of online behavioral advertising is to create a profile of a user's Internet activities, such as the websites he or she visits. This profile can later be used to target advertisements. When a user visits a web page, that page's content can come from both a first party (the page that the user is explicitly visiting) and third parties (companies that have a relationship with the first party allowing them to place content, visible or not, on that page). Third parties include advertising networks, analytics companies, and social networks that contract with first-party websites. These third parties can set a unique identifier on a user's computer. Then, as the user visits different websites that include content from the same third party, that third party can associate these visits with the same computer. In recent years, a small number of third parties have increasingly served content on a larger number of pages, enabling these companies to track a user's browsing across the Internet [83].

On a technical level, this tracking can be accomplished in many ways. In one of the simplest cases, an advertiser can set a cookie with a unique identifier on a user's computer, correlating browsing activity with that unique identifier [83]. In a study proposing a method for measuring behavioral advertising, Balebako et al. found that blocking third-party cookies in a web browser achieved a reduction in behavioral targeting similar to opting out of behavioral advertising using industry opt-out websites, albeit only testing behavioral targeting in text ads from Google [8]. However, there exist myriad means of uniquely profiling a particular computer, ranging from browser fingerprinting [39] to using Flash Local Shared Objects (LSOs), HTML5 local storage, or other methods of maintaining a unique identifier on a user's computer over time [97]. Many of these techniques aim to uniquely associate browsing activity with a particular computer, rather than with an individual's real-life identity. The Network Advertising Initiative, a U.S. trade group for advertisers, notes, "It is possible to merge PII [personally identifiable information] and Non-PII for OBA and other uses. However, no NAI member ad networks currently engage in this practice" [106].

3.3.2 Notice and Choice

Behavioral advertising has led a number of parties to voice privacy concerns. For instance, the U.S. Federal Trade Commission has noted that data collection can be invisible, privacy notices may be difficult to understand, consumer profiles are sometimes very detailed, and that there is a "risk that data collected for behavioral advertising – including sensitive data regarding health, finances, or children – could fall into the wrong hands or be used for unanticipated purposes" [47].

The most visible attempts in the United States to provide consumers notice and choice about OBA have come about as a result of advertising industry self-regulation by groups such as the Digital Advertising Alliance (DAA) and Network Advertising Initiative (NAI). Policies from both groups have far-reaching implications; for instance, the DAA notes that its members "comprise 85% of the OBA marketplace" [80].

Both groups' principles include the idea of providing consumers notice and choice about behavioral advertising [36,108]. For instance, the DAA's *consumer control* principle requires that users be able to opt out of receiving targeted advertisements, although it does not require that users be able to opt out of being tracked online [36]. These opt-outs can be enabled via the DAA's opt-out website,[1] on which consumers can opt out of OBA on a per-company basis. The DAA's *transparency* principle requires that consumers receiving OBA be given "enhanced notice," providing "the ability to exercise choice regarding the collection and use of data for online behavioral advertising" via "common wording and a link/icon that consumers will come to recognize" [36]. In 2010, the industry selected the "Advertising Option Icon" to indicate a link to enhanced notice for behavioral ads [149]. However, some advertisements still display an older icon or none at all [77]. In any case, my co-authors and I previously found that users often have misperceptions and misunderstandings about these icons and their associated opt-out pages [86].

[1] `http://www.aboutads.info/choices/`

3.3.3 Previous User Studies

Studies from the past fifteen years have examined different facets of user sentiment towards online advertising in general. In a 2002 paper, Rodgers described two studies with 106 student and 38 non-student participants that looked at interactions between user motivation for using the Internet and the effectiveness of certain types of banner ads, finding that for at least some users, “ads that complement the user’s motive may have more success at being noticed and clicked on than ads that do not” [122]. In a 2003 paper, Rettie et al. described a survey with 100 UK student participants, finding that only 13% enjoy Internet advertising. Fewer than 20% of participants found Internet ads informative or useful. Although 62% indicated that they prefer that websites not have ads, 69% agreed with accepting “ads as pay for content” [121]. In 2007, McCoy et al. described a study with 536 participants and found that online advertising caused users to report being both less likely to return to a website and less able to recall features of that website [100]. However, Campbell and Wright conducted a survey study with 97 participants and a laboratory study with 118 participants in 2008, finding that the personal relevance of ads increased users’ positive attitudes toward repetitive online advertisements [20]. Taken together, this prior work suggests that users find online advertising annoying, yet targeted ad selection may reduce annoyance.

Other studies have looked specifically at user perception of OBA and online tracking, finding significant privacy concerns about the practice. Turow et al. conducted a 2009 survey of 1,000 US adult Internet-users and discovered that 68% of Americans “definitely would not” and 19% “probably would not” allow advertisers to track them online, even anonymously [142]. In a study published in 2010 that included 14 in-person interviews and an online survey of 314 participants, McDonald and Cranor found that just one-fifth of their online respondents preferred targeted ads to random ads, and 64% thought targeted ads were “invasive.” The study found that “people understand ads support free content, but do not believe data are part of the deal” [101]. In a 2009 online study of 2,604 participants, Hastak and Culnan found that 46% of respondents were uncomfortable with the identities of the websites they visit being used to target ads, although this number decreased to 30% of participants when the practice was transparent and offered participants the choice not to receive targeted ads [61]. A 2012 Pew telephone survey of 2,253 participants found that 68% of respondents were “not okay with targeted advertising because [they] don’t like having [their] online behavior tracked and analyzed” [118].

Stakeholders from both the privacy-services and advertising industries have also surveyed consumers about OBA. TRUSTe conducted a 2011 survey with 1,004 United States residents, asking about perceptions of OBA. 53% of participants agreed that online privacy is “a really important issue that I think about often,” and another 41% agreed that it is “a somewhat important issue that I think about sometimes.” Over a third of participants agreed with the statement: “I know how to protect my personal information online and consistently take the necessary steps to do so.” Over half of participants indicated that they definitely or probably would not share their browsing behavior with advertisers, and only 15% indicated being willing or probably willing to consent to being tracked online for relevant ads. Only 8% of participants indicated liking OBA, and only 5% showed awareness of the Advertising Option Icon [139]. A 2011 marketing survey of 9,600 individuals across 31 countries found that 90% of respondents expressed concerns about the privacy of their personally identifiable information, yet 62% were willing to allow online advertisers to track their web usage "under the right circumstances" [81]. More recently, my co-authors and I investigated what types of information collected for OBA actually matter to users [88].

Much of this past work has employed surveys to gauge the attitudes of a large number of participants. However, surveys are inherently limited in that they don't provide a way for consumers to discuss ideas and thoughts outside the questions asked, or for followup questions to be asked. As a result, surveys alone cannot fully explain how different nuances of attitude are connected to each other. The work I report in the "Smart, Useful" study (Section 4.6) fills in the gap of understanding how consumers are simultaneously privacy-concerned and willing to have their information collected.

3.4 Privacy in the Financial Industry

The final case study in my thesis focuses on privacy notice and choice in the financial industry. In this section, I thus discuss both relevant aspects of the U.S. financial industry, as well as privacy notice and choice in general. In particular, I describe the privacy provisions of the U.S. Gramm-Leach-Bliley Act (GLBA), some criticisms of those provisions, and the regulatory development of an optional standardized format for financial institutions' privacy disclosures. Finally, I highlight efforts to improve privacy notices beyond the financial industry, including the creation of formal specifications, standardized formats, and usable privacy notices.

3.4.1 Federal Laws' Financial Privacy Provisions

In the "Financial" study (Section 4.8), I rely on financial institutions' annual privacy disclosures that are mandated by GLBA, which was signed into law on November 12, 1999 [57]. GLBA's primary purpose was to encourage competition in the financial services industry by removing barriers that prevented common ownership (affiliation) between commercial banks, investment banks, and insurance businesses [94, 129, 156].

Affiliation between different types of financial services companies presented an opportunity for newly affiliated companies to share information. In response to concerns about the privacy of consumer information, Congress included Title V, known as the Privacy Rule, in GLBA. This rule requires financial institutions to provide annual notices of their privacy policies and practices (15 U.S.C. §§ 6802–6803). The rule also mandates that customers have the right to opt out of data sharing with nonaffiliated companies. However, the Privacy Rule provides a "joint marketing exception" to the opt-out requirements, allowing nonaffiliated financial companies to share information without offering an opt-out when there exists a formal agreement for marketing financial products or services to a consumer [26].

Although GLBA's Privacy Rule does not give consumers a general right to opt out of all data sharing, the Fair Credit Reporting Act (FCRA) does give consumers that right for certain types of credit information. The FCRA, which regulates the use and distribution of consumer information, exempts from its definition of a consumer report any communication between affiliates. However, this exemption only applies if the communication is "clearly and conspicuously disclosed to the consumer . . . and the consumer is given the opportunity, before the time that the information is initially communicated, to direct that such information not be communicated among such persons" (15 U.S.C. § 1681a(d)(2)(A)(iii)). In other words, consumers must be able to opt out of data sharing about their creditworthiness between affiliates.

The Fair and Accurate Credit Transactions Act of 2003 (FACTA) [44] amended the FCRA to further restrict the use of information shared between affiliates. The rule, called the "Affiliate Marketing Rule," prohibits companies that receive information that would be considered a consumer report if not for § 1681a(d)(2)(A)(iii) from using that information for marketing unless the consumer is given notice and the opportunity to opt out (15 U.S.C. § 1681s-3(a)).

The provisions of GLBA, the FCRA, and FACTA combine to establish three contexts in which financial institutions must provide notice and the opportunity to opt out. GLBA's Financial Privacy Rule applies to the sharing of consumer financial information with non-affiliates, the FCRA restricts sharing consumer report information between affiliated companies, and FACTA limits when consumer report information shared between affiliates may be used for marketing [99].

3.4.2 Criticisms of GLBA's privacy provisions

The privacy protections offered by GLBA have prompted a range of criticisms. Some critics feel that GLBA offers incomplete or too few privacy protections. For instance, in an examination of GLBA privacy provisions, Janger et al. conclude that GLBA "leaves the burden of bargaining on the less informed party, the individual consumer" [67]. Schiller also argues that the notice provisions provided by GLBA do not go far enough toward providing privacy protections [125]. She recommends that GLBA further restrict information sharing among affiliates. Freeman similarly concludes that GLBA was a good start, yet "need[s] further refinement" [51], arguing that the "opt-out" provision has made it unlikely that many customers will take the active steps needed to protect their confidential data" [51]. Nojeim also argues that GLBA is incomplete because it does not prevent the flow of personal information among affiliates and uses an opt-out approach, failing to require consumers' active consent [110].

Other critics feel that the protections offered by GLBA are an impediment to the free market. Some economists have claimed that "efforts to protect privacy in the financial services industry (and elsewhere) are obstacles to the functioning of optimally efficient markets" [136]. Lacker, for example, argues that in a perfectly competitive market, financial privacy would be determined by economic forces regardless of the choice mechanisms offered [85]. Furletti and Smith claim that the open sharing of consumer information makes the market more efficient and benefits both financial institutions and consumers. They further claim that laws like the Fair Credit Reporting Act provide sufficient privacy protections for consumers [52]. In counterpoint, Swire argues that inappropriate disclosure of personal information can easily lead to a "misallocation of resources" [136].

Investigations conducted around the time GLBA came into effect studied the act's initial impact on financial institutions' privacy disclosures. Sheng et al. performed a longitudinal study of fifty financial institutions' privacy policies. They found that although privacy policies became more complete and contained more detailed information about sharing practices after GLBA, the amount of sharing among affiliates and nonaffiliates increased [128]. Antón et al. examined privacy statements from nine financial institutions covered by GLBA and concluded that these statements did not comply with the GLBA requirements of conspicuousness and clarity. They suggested the use of a standardized vocabulary to improve the readability of financial institutions' privacy policies [5].

3.4.3 Development of the model privacy form

A few years after GLBA was enacted, eight U.S. regulators[2] jointly noted wide variations in the privacy notices financial institutions were sending to consumers. They found these notices "difficult to compare, even among financial institutions with identical practices" and questioned "whether such notices comply with the requirement that they be clear and conspicuous." As a result, regulators started a process to create a standard model for privacy notices that "consumers could more easily use and understand" [112]. Financial institutions, researchers, and communications firms took part in this process.

The process of developing a standardized notice began in the summer of 2004. The regulators retained a communications firm, Kleimann Communication Group, to develop a prototype of a standardized notice. To this end, the firm conducted two ten-participant focus groups and 46 individual interviews, releasing a report of their findings in February 2006 [72]. Notably, the main goal of the prototype notice was to help consumers understand financial institutions' sharing practices, not necessarily to provide a comprehensive list of the types of personal information that financial institutions collect. In March 2007, the regulators issued the prototype for public comment [112].

Following public comments on the proposed model form, the regulators commissioned a quantitative survey designed to evaluate the effectiveness of the revised model form. The survey, which was conducted in the spring of 2008, tested comprehension and usability of the model form as compared with three other styles of notice. Notices from three fictitious banks with different sharing practices were tested among 1,032 consumers recruited from five US cities. The prototype outperformed the alternative styles tested [95].

In December 2008, Levy and Hastak submitted a report to the regulators analyzing the results of the usability testing [90]. Although participants who tested the proposed prototype better understood the differences in sharing practices, Levy and Hastak found that participants experienced problems understanding how to exercise their opt-out rights. The report proposed improvements to reduce the length of the disclosure table and to increase the clarity of opt-out choices. The regulators revised the model form again based on both the Levy-Hastak report and public comments received after publishing the survey results.

The regulators again commissioned Kleimann Communication Group to conduct validation testing. The firm conducted a seven-participant study and concluded in its February 2009 report that the improvements suggested by Levy and Hastak improved the clarity of opt-out choices without affecting understanding of sharing practices [73]. Garrison et al. give a more detailed account of the user testing behind the model forms [53].

[2] The Office of the Comptroller of the Currency; the Board of Governors of the Federal Reserve System; the Federal Deposit Insurance Corporation; the National Credit Union Administration; the Federal Trade Commission; the Securities and Exchange Commission; the Office of Thrift Supervision; and the Commodity Futures Trading Commission

In December 2009, the regulators released the final model privacy form, shown in Figure 3.1 and Figure 3.2. Although use of the model privacy form is voluntary, financial institutions may rely on this model privacy form as a safe harbor to provide privacy disclosures [112], potentially spurring its adoption. Notably, this model privacy form is the basis of one of the first widespread uses of a standardized format for privacy disclosures, facilitating our large-scale analysis.

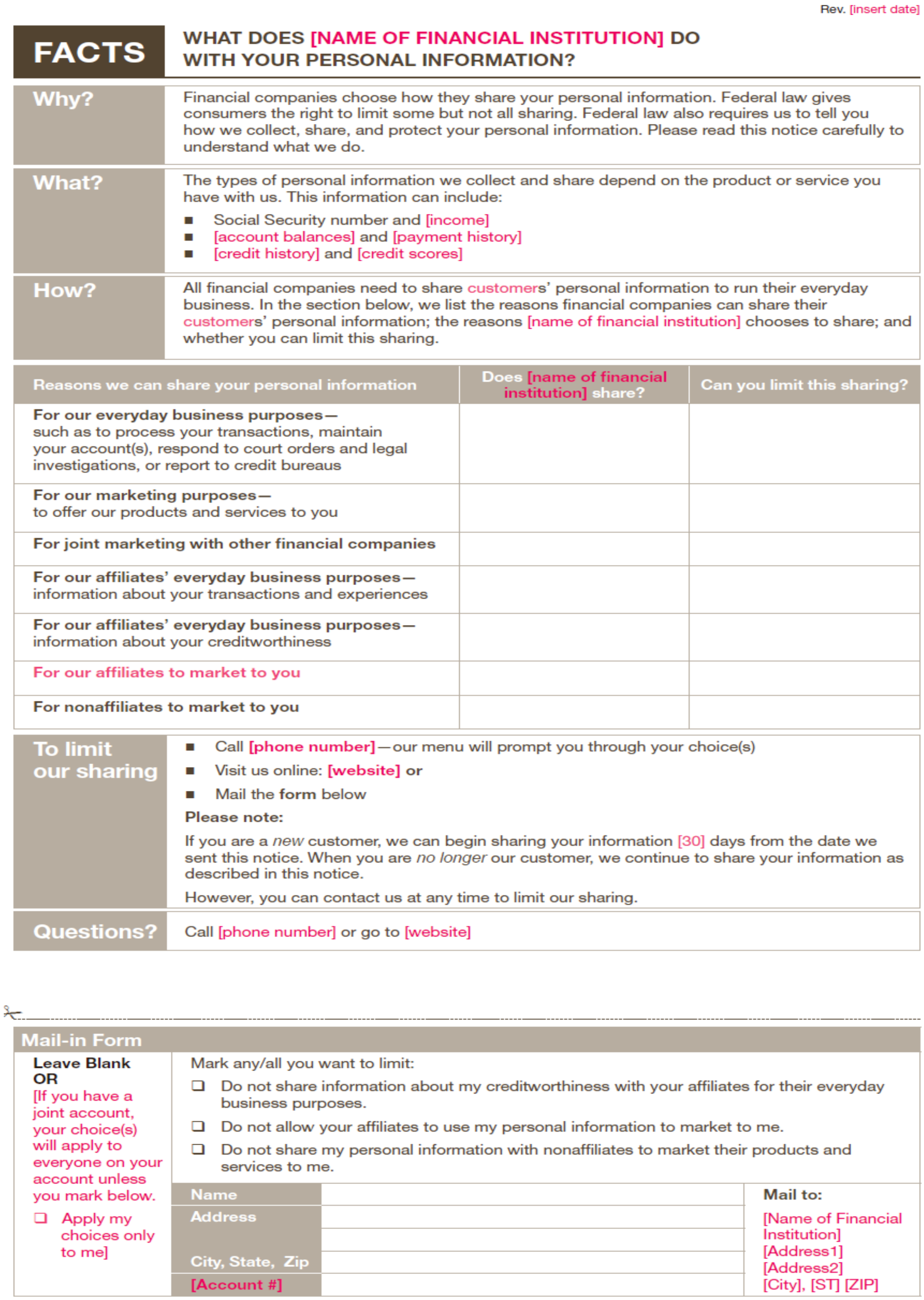
3.4.4 Privacy policies

The idea that consumers should receive clear notice about privacy is a core principle of many privacy frameworks, including the OECD's 1980 privacy guidelines [113] and the U.S. Federal Trade Commission's Fair Information Practice Principles (FIPPs) [46]. Privacy notice is often presented to consumers in the form of a privacy policy. Overall, privacy notice has been found to impact trust and promote social welfare. For instance, in a study of retail websites, Tang et al. found that the clarity and credibility of privacy notices were crucial for influencing consumer trust [137].

Unfortunately, a number of issues negatively impact the usability of current privacy policies. Privacy policies are generally written at a very high reading level. For instance, in a study of health websites, Graber et al. found the average privacy policy to require two years of college education to comprehend [56]. Similarly, Jensen and Potts examined 64 privacy policies and found that many were difficult to find and read [68]. The reading level of privacy policies is not the only barrier to comprehension; privacy policies are sometimes unavailable in a user's language, even when the rest of a website is available in that language [147]. McDonald and Cranor examined the length of privacy policies, estimating that a user would need to spend hundreds of hours a year to read all of the privacy policies relevant to their browsing [102].

Well-designed, standardized formats for privacy notice can overcome many of these obstacles. Furthermore, privacy notices can be compared easily if they are presented in a standardized format. Researchers have examined methods for presenting privacy policies in a standardized, usable manner. For example, Kelley et al. found that displaying privacy policy information in a tabular "nutrition label" format made it easier for users to find information [70].

Standardized privacy notices—whether human-readable or machine readable—help facilitate large-scale comparison and evaluation [29]. For instance, the Platform for Privacy Preferences (P3P) is an XML-based W3C standard for machine-readable privacy policies that specifies what data will be collected and how it will be used [28]. Cranor et al. conducted a study of several hundred computer-readable privacy policies encoded using P3P. They used automated tools to analyze the data collection, use, and sharing practices encoded in each policy. [30]. Unfortunately, P3P has not been widely adopted [29]. In a different study, Cranor et al. found high rates of syntax errors among the P3P policies they examined [30]. Furthermore, Leon et al. found a number of websites misrepresenting their privacy practices through erroneous or misleading P3P compact policies, which are short strings designed to summarize privacy practices associated with cookies [89].



Rev. [insert date]

FACTS **WHAT DOES [NAME OF FINANCIAL INSTITUTION] DO WITH YOUR PERSONAL INFORMATION?**

Why?	Financial companies choose how they share your personal information. Federal law gives consumers the right to limit some but not all sharing. Federal law also requires us to tell you how we collect, share, and protect your personal information. Please read this notice carefully to understand what we do.
What?	The types of personal information we collect and share depend on the product or service you have with us. This information can include: ■ Social Security number and [income] ■ [account balances] and [payment history] ■ [credit history] and [credit scores]
How?	All financial companies need to share customers' personal information to run their everyday business. In the section below, we list the reasons financial companies can share their customers' personal information; the reasons [name of financial institution] chooses to share; and whether you can limit this sharing.

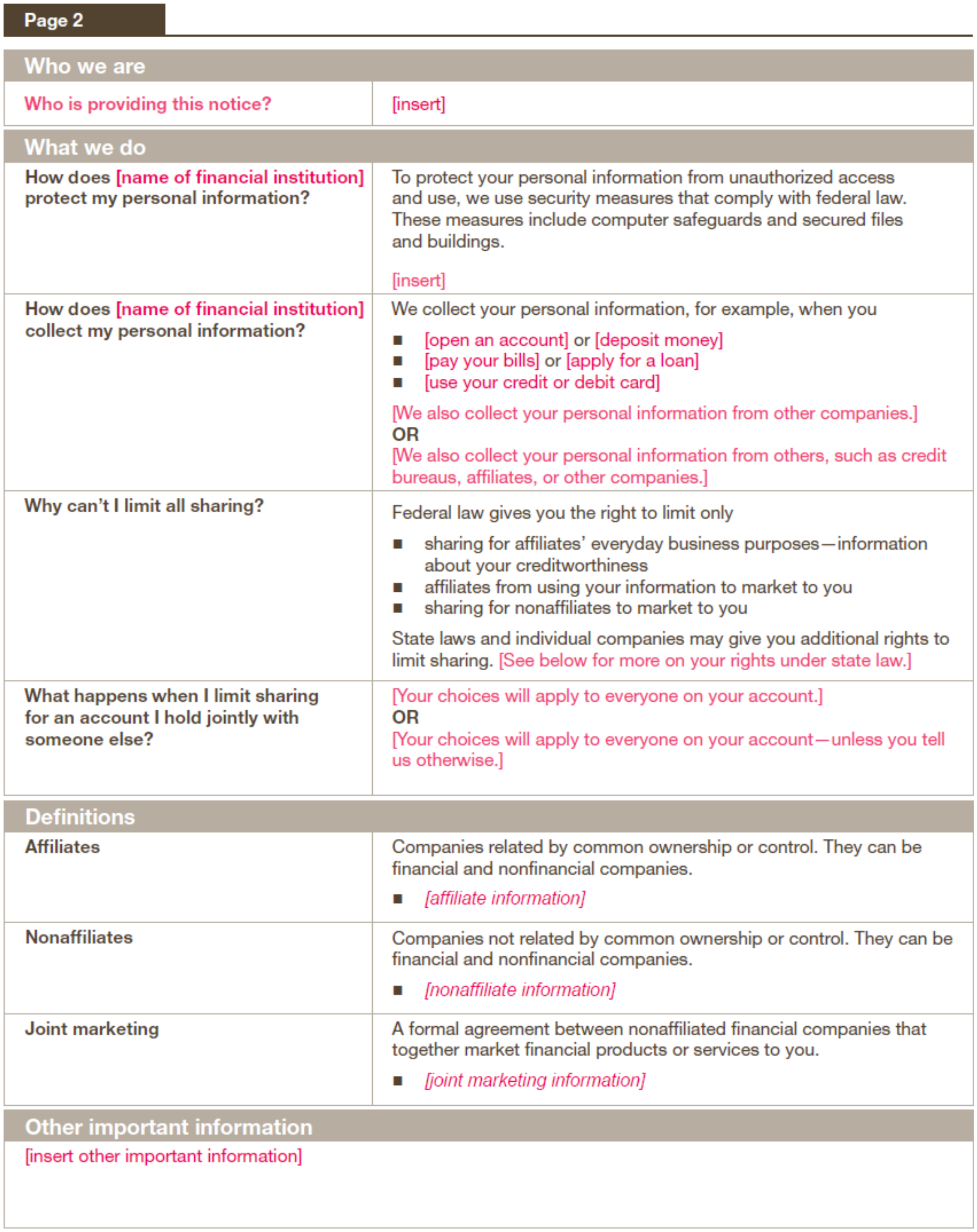
Reasons we can share your personal information	Does [name of financial institution] share?	Can you limit this sharing?
For our everyday business purposes— such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus		
For our marketing purposes— to offer our products and services to you		
For joint marketing with other financial companies		
For our affiliates' everyday business purposes— information about your transactions and experiences		
For our affiliates' everyday business purposes— information about your creditworthiness		
For our affiliates to market to you		
For nonaffiliates to market to you		

To limit our sharing	■ Call **[phone number]**—our menu will prompt you through your choice(s) ■ Visit us online: **[website] or** ■ Mail the **form** below **Please note:** If you are a *new* customer, we can begin sharing your information [30] days from the date we sent this notice. When you are *no longer* our customer, we continue to share your information as described in this notice. However, you can contact us at any time to limit our sharing.
Questions?	Call [phone number] or go to [website]

Mail-in Form

Leave Blank OR [If you have a joint account, your choice(s) will apply to everyone on your account unless you mark below.] ❑ Apply my choices only to me]	Mark any/all you want to limit: ❑ Do not share information about my creditworthiness with your affiliates for their everyday business purposes. ❑ Do not allow your affiliates to use my personal information to market to me. ❑ Do not share my personal information with nonaffiliates to market their products and services to me.		
	Name		**Mail to:** [Name of Financial Institution] [Address1] [Address2] [City], [ST] [ZIP]
	Address		
	City, State, Zip		
	[Account #]		

Figure 3.1: The first page of the model privacy form [112]. I extracted and analyzed what information is collected, how information is shared, including whether consumers can limit any type of sharing, and how consumers may limit sharing. The sharing table and text in pink need to be filled in by the financial institution.



Page 2

Who we are	
Who is providing this notice?	[insert]
What we do	
How does [name of financial institution] protect my personal information?	To protect your personal information from unauthorized access and use, we use security measures that comply with federal law. These measures include computer safeguards and secured files and buildings. [insert]
How does [name of financial institution] collect my personal information?	We collect your personal information, for example, when you ■ [open an account] or [deposit money] ■ [pay your bills] or [apply for a loan] ■ [use your credit or debit card] [We also collect your personal information from other companies.] **OR** [We also collect your personal information from others, such as credit bureaus, affiliates, or other companies.]
Why can't I limit all sharing?	Federal law gives you the right to limit only ■ sharing for affiliates' everyday business purposes—information about your creditworthiness ■ affiliates from using your information to market to you ■ sharing for nonaffiliates to market to you State laws and individual companies may give you additional rights to limit sharing. [See below for more on your rights under state law.]
What happens when I limit sharing for an account I hold jointly with someone else?	[Your choices will apply to everyone on your account.] **OR** [Your choices will apply to everyone on your account—unless you tell us otherwise.]
Definitions	
Affiliates	Companies related by common ownership or control. They can be financial and nonfinancial companies. ■ *[affiliate information]*
Nonaffiliates	Companies not related by common ownership or control. They can be financial and nonfinancial companies. ■ *[nonaffiliate information]*
Joint marketing	A formal agreement between nonaffiliated financial companies that together market financial products or services to you. ■ *[joint marketing information]*
Other important information	
[insert other important information]	

Figure 3.2: The second page of the model privacy form [112]. From this page, I extracted and analyzed how information is collected, as well as the list of affiliates, nonaffiliates, and joint marketing partners.

Chapter 4

Completed and Proposed Studies

My thesis consists of two main parts. The first part of the thesis encompasses five studies on analyzing passwords for the purpose of providing data-driven feedback to users about password creation. The second part of the thesis encompasses three studies on data-driven privacy. Two of these privacy studies are in the domain of online behavioral advertising, while the third study involves presenting consumers with data-driven information about financial institutions' privacy practices at the time a consumer is considering switching financial institutions.

In more detail, through four initial studies on aspects of passwords, I build towards a fifth study that integrates the insights from the prior studies in directly helping users make better passwords. In the "Meters" study (USENIX Security 2012 [143]), I demonstrate that password-strength meters do influence both the characteristics and the guessability of passwords users create. I also find that the way passwords are scored has far greater influence on these factors than the visual design of the password-strength meter. Although this study demonstrates that meters can lead users to change their behavior, the heuristics of password length and character-class usage that underlie common password-strength meters are often only weakly correlated to strong passwords. For example, Google's password-strength meter gives "aa1aa!AA1AA!" its maximum rating of "strong" even though the password is very predictable.

Through three subsequent studies, I delve into the following three aspects that are crucial for helping users make better passwords: understanding which passwords are actually secure; better understanding semantic and structural patterns hidden in passwords; and uncovering how users' perceptions of what features make a password secure match with, or diverge from, these features' actual impact on security. In the first of these studies, the "Biases" study (USENIX Security 2015 [146]), I evaluate the guessability of diverse sets of passwords to 100 trillion (10^{14}) guesses by a professional attacker and many configurations of four different automated password-cracking approaches. I find that a particular password's minimum guess number across the four automated approaches can serve as a conservative upper bound on a password's guessability by professionals, at least through 100 trillion guesses. In the second study, the "Art" study (completed but not yet published), I instead delve into password semantics and structures by applying crowdsourcing best practices to have human crowdworkers manually "reverse engineer" passwords, undoing transformations and separating each password into its component sections. Using the reverse-engineered passwords, I document the frequencies of different character substitutions, the use of dictionaries, the prevalence of multi-word phrases from different corpora, and the ways in which users modify candidate passwords to comply with password-composition policies. In the “Perceptions” study (piloted, but not yet conducted), I instead focus on users’ perceptions of security, particularly how this perception aligns with reality. In addition, I also gauge study participants’ models of attackers and password threats.

In the fifth study in the security portion of my thesis, the “Password Feedback” study, I will combine the insights from the four previous studies on passwords into a data-driven tool that provides users password feedback. Through both a small-scale laboratory study and an online study, I will then test the effectiveness of different types of feedback in helping users create more secure passwords for a security-critical account. This culminating project directly examines my hypothesis that detailed data provided just in time when a user is creating a password will help the user create a more secure, yet equally memorable, password.

Through three additional studies, I also explore data-driven decision making in two different aspects of privacy. The first area, online behavioral advertising (OBA), involves the collection of a user’s web browsing data by third-party companies for the purpose of targeting advertising to the user’s individual interests. To lay the groundwork for data-driven decision making in this area, in the “Smart, Useful” study (SOUPS 2012 [144]) I take a qualitative approach to understanding users’ mental models of online behavioral advertising. Through interviews of 48 privacy-concerned Internet users, I find that most participants have inaccurate models of how OBA works. As a result, these users are left confused and uncertain about what data is collected for OBA purposes. Furthermore, they are uncertain how to balance their desire to receive relevant advertising with their privacy concerns. To help users better understand the mechanics of OBA, I will visualize for users what information advertisers know about their browsing in my follow-up “Visualizing OBA” study. I will use a browser plugin to locally “track the trackers,” recording on what webpages different advertising companies have tracked the user and estimating what they may have inferred about the user’s interests.

The third study in the privacy section of my thesis falls within the domain of consumer decision-making about privacy. In particular, for the “Financial” study, I will rely on data I have already gathered [**?**, 31] from the standard-format privacy notices used by many U.S. financial companies. Using this data, I will investigate whether the availability of comparative privacy information, augmented with consumer-specific data about which banks have nearby branches, influences consumers’ stated willingness to consider switching to a more privacy-protective institution.

Below, I detail each of the studies I have completed or am proposing as part of this thesis.

4.1 “Meters” study [USENIX Security 2012]

My first study of password-strength meters laid the groundwork for my future work in using data to help users make better passwords by documenting that password-strength meters can influence the passwords users create. My co-authors and I conducted a between-subjects study of 14 different password-strength meters. We designed our baseline meter to reflect typical practices we observed in a survey we conducted of the password-strength meters on Alexa's 100 most visited global sites. In particular, this meter used a bar metaphor to display password strength, provided suggestions about how to make a password better (e.g., "consider adding a digit"), and scored a password based on the heuristics of the length of the password and the number of character classes it contains. As we saw in our survey of popular sites, this type of heuristic scoring is very widely used in the wild, likely because it is very easy to implement.

The other 13 conditions each varied in one or more dimensions from the baseline meter. For most of the conditions, we used the same underlying scoring mechanism, yet changed the visual presentation. For example, one condition had a very large meter, while another had a very small meter. In another condition, we made the meter monochromatic, as opposed to using color as a reiteration of the password-strength score. In addition to conditions that varied in visual presentation, we tested conditions that scored passwords differently in two different ways. The first of these ways scored passwords using the same heuristics of length and character-class usage, yet was more stringent in converting these characteristics into a final score. In particular, we tested always giving passwords half and one-third the score they would have received with the baseline meter. The two conditions in our second class of alternate scoring approaches focused on nudging users towards a particular type of password. In particular, we tried encouraging users to make increasingly long passwords, as well as passwords with increased character-class complexity.

We had 2,931 participants each create a password in one of these conditions, assigned round-robin. We measured the characteristics of the passwords (e.g., length and character-class usage), characteristics of password creation (e.g., the time it took to create a password), and users' self-reported sentiment on the degree to which password creation was annoying, fun, or difficult. As our security metric, we calculated the guessability of each password under Weir et al.'s probabilistic context-free grammar (PCFG) approach to password cracking [155], as implemented in Kelley et al.'s guess-calculator framework [71].

Although we found that meters with a variety of visual appearances led users to create longer passwords, we only achieved significant increases in security using meters that scored passwords stringently. These stringent meters also led participants to include more digits, symbols, and uppercase letters. Password meters also affected the act of password creation. Participants who saw stringent meters spent longer creating their password and were more likely to change their password while entering it, yet they were also more likely to find the password meter annoying. However, the most stringent meter caused participants to place less importance on satisfying the meter. Participants who saw more lenient meters tried to fill the meter and were averse to choosing passwords a meter deemed "bad" or "poor."

4.2 "Biases" study [USENIX Security 2015]

Much of the passwords work in my thesis, as well as passwords research in general, relies on being able to estimate the security of a password. To that end, my "Biases" study analyzes the guessability of diverse sets of passwords to an attack by professionals. I then calculate the guessability of these same passwords against many configurations of two software tools (Hashcat and John the Ripper) popular in adversarial cracking, as well as multiple configurations of two password-cracking approaches popular in academia (Markov models and a probabilistic context-free grammar).

For many years, the security of a set of passwords was determined by calculating the entropy of the set. In recent years, traditional entropy metrics have fallen out of favor because they do not reflect how easily a password can be cracked in practice [12,71,154]. It has instead become common to measure password strength by running or simulating a particular cracking algorithm, parameterized by a set of training data [13, 71, 154]. This approach has two main advantages. First, it calculates the guessability of each password individually, enabling data-driven strength estimates during password creation [22, 75]. Second, it estimates real-world security against existing, rather than idealized, adversarial techniques. A disadvantage of this approach is that the (simulated) cracking algorithm may not be configured or trained as effectively as by a real attacker, leading to inaccurate estimates of password strength.

Our analysis is the first study of how various cracking approaches used by researchers compare to real-world cracking by professionals, as well as how the choice of approach biases research conclusions. We contracted a computer security firm specializing in password recovery to crack a set of passwords chosen for their diversity in password-composition policies. We then computed the guessability of these passwords using four popular approaches. We tested many configurations of two well-known password-cracking toolkits: John the Ripper [115] and oclHashcat [132]. We also tested two approaches popular in academia: Weir et al.'s probabilistic context-free grammar (PCFG) [155] and Ma et al.'s Markov models [93].

Unsurprisingly, a professional attacker updating his or her strategy dynamically during cracking outperformed fully automated, "fire-and-forget" approaches (henceforth simply referred to as *automated*), yet often only once billions or trillions of guesses had been made. We found that relying on a single automated approach to calculate guessability underestimates a password's vulnerability to an experienced attacker, but using the earliest each password is guessed by any automated approach provides a realistic and conservative approximation.

We found that each approach was highly sensitive to its configuration. Using more sophisticated configurations than those traditionally used in academic research, our comparative analysis produced far more nuanced results than prior work. These prior studies found that Markov models substantially outperform the PCFG approach [37, 93], which in turn substantially outperforms tools like John the Ripper [34,154,160]. We found that while Markov was marginally more successful at first, it was eventually surpassed by PCFG for passwords created under typical requirements. Furthermore, the most effective configurations of John the Ripper and Hashcat were frequently comparable to, and sometimes even more effective than, the probabilistic approaches.

Both the differences across algorithms and the sensitivity to configuration choices are particularly notable because most researchers use only a single approach as a security metric [22, 25, 38, 98, 127, 143, 154]. In addition, many researchers use adversarial cracking tools in their default configuration [24, 32, 33, 43, 50, 59, 76, 159]. Such a decision is understandable since each algorithm is very resource- and time-intensive to configure and run. This raises the question of whether considering only a single approach biases research studies and security analyses. For instance, would substituting a different cracking algorithm change the conclusions of a study?

We investigate these concerns and find that for comparative analyses of large password sets (e.g., the effect of password-composition policies on guessability), choosing one cracking algorithm can reasonably be expected to yield similar results as choosing another.

However, more fine-grained analyses—e.g., examining what characteristics make a password easy to guess—prove very sensitive to the algorithm used. We find that per-password guessability results often vary by orders of magnitude, even when two approaches are similarly effective against large password sets as a whole. This result has particular significance for efforts to help system administrators ban weak passwords or provide customized guidance during password creation [22, 75].

To facilitate the analysis of password guessability across many password-cracking approaches and to further systematize passwords research, we introduce a Password Guessability Service [21] for researchers. I rely on this service for calculating password guessability in other parts of my thesis.

4.3 "Art" study [Completed; not yet published]

Whereas the "Biases" study focuses on estimating the security of passwords, the "Art" study is a deep analysis of user behavior in creating passwords. The "Art" study, which has been mostly completed but not yet published, builds towards providing users data-driven feedback on password creation by using numerous, complementary data-analysis techniques to uncover hidden semantic and structural patterns in passwords. Our investigation is enabled by applying crowdsourcing and programmatic techniques to reverse engineer more than 45,000 passwords into a representation illuminating their structure and semantics. To reveal otherwise obscured password elements, crowdworkers on Amazon's Mechanical Turk reverse engineered each password into semantic "chunks" and undid character substitutions. For example, the password *˜Cowscomehom3* became *till the cows come home*; the crowdworkers realized the tilde stood for "till the." To understand how patterns we discovered corresponded to attackers' ability to guess passwords, we also modeled each password's vulnerability to two major password-guessing approaches.

Building on this preprocessing phase, we offer three main contributions that collectively provide new insights into users' habits and reveale subtle, yet common, patterns that should be discouraged during password creation.

First, we provide a large-scale quantification of the use of character substitutions in passwords. Between 5% and 18% of passwords, depending on the source, contain substitutions, and the top twenty mappings account for more than 77% of all substitutions. This ground-truth data can enable proactive password checking to account for substitutions.

Second, we report on password semantics, identifying choices that users should avoid. We compare frequencies of words across different password sets and natural language. For example, we find that up to 5% of words used in passwords are contained in a 247-word list of pet names and many others appear in small dictionaries. In addition, roughly half of the passwords we examined that contained any content from a dictionary contained a common multi-word phrase. We use Wikipedia's categorization system to further analyze semantics, discovering previously unreported patterns.

Finally, we delve into the individual steps of password creation. We analyze sequences of attempts to comply with a password-composition policy to understand how users modify passwords across attempts. Unexpectedly, forcing users to comply with strict policies sometimes reduces security; over 20% of users who generated a completely new password after their original attempt was rejected made a less secure password. In contrast, certain types of small modifications nearly always made a password harder to guess. We also examine the impact of suggestions (e.g., "Add a digit") during password creation, finding that users do follow these suggestions.

In the "Password Feedback" portion of this thesis, I draw heavily on these insights about password semantics and structure to design data-driven feedback for users.

4.4 "Perceptions" study [Not completed, but piloted]

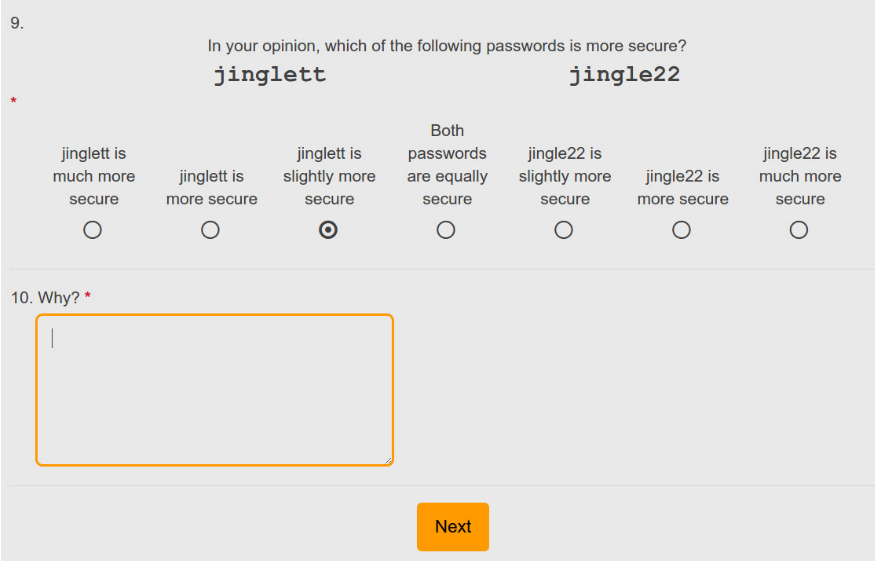
Whereas the previously described studies shed light on both the security and the semantic and structural composition of passwords, they did not directly examine the critical aspect of how users perceive the security impact of different characteristics. Users making a predictable password becomes particularly problematic from a security standpoint if they do not realize that password is weak. My proposed "Perceptions" study builds off our recent, in-lab study in which we had 49 participants "think aloud" while creating three passwords each [145]. In that study, we found that participants often had trouble estimating the security of common "microdecisions" made in the course of password creation, such as adding a symbol to the end of a password (e.g., "password!"), using dictionary words as part of a password, using common phrases, and substituting numbers or symbols for letters. My proposed study further builds off of, and adapts methods from, Aviv and Fichter's study of users' perceptions of the security of Android graphical unlock patterns [6]. They showed participants two unlock patterns for an Android phone, and they had participants pick which of the two patterns they believed to be more secure.

My proposed "Perceptions" study will be an online study that gauges both how users perceive, and misperceive, the security impact of different password characteristics. The study will also document users' attacker models. For this study, I will recruit 500 workers on Amazon's Mechanical Turk platform. As described below, the study takes approximately thirty minutes.

To investigate the first question on security perceptions, I will show participants 25 pairs of passwords. For each password pair, they will rate the comparative strength of these passwords on an implicit 7-point scale, as shown in Figure 4.1 (1 = the first password is much stronger; 4 = the passwords are of equal strength; 7 = the second password is much stronger).

Each of these password pairs will test a different hypothesis. The following hypotheses are examples of those I will test:

1. Participants will think that a password of length 8 will be less secure than a password of length 12.
2. Participants will think that a password using two character classes will be more secure than a password using one class.
3. Participants will think that a password using three character classes will be more secure than a password using two classes.
4. Participants will think that using a word including the name or function of a site will be less secure than using a random word.



9.

In your opinion, which of the following passwords is more secure?

`jinglett` `jingle22`

*

jinglett is much more secure	jinglett is more secure	jinglett is slightly more secure	Both passwords are equally secure	jingle22 is slightly more secure	jingle22 is more secure	jingle22 is much more secure
○	○	●	○	○	○	○

10. Why? *

Figure 4.1: An example password comparison in the "Perceptions" study.

5. Participants will think that using a single "uncommon" word will be more secure than using two common words.
6. Participants will think that using a common keyboard pattern will be more secure than using a dictionary word.
7. Participants will think that using a common keyboard pattern will be less secure than a "random" password without a discernible pattern.
8. Participants will think that using any special character (at the beginning, middle, or end) will be more secure than using a digit in any of those places.
9. Participants will think that using a number with no semantic significance in the middle of the password will be more secure than using it at the beginning or end.
10. Participants will think that passwords using common structures are equally secure to those using infrequent structures.
11. Participants will think that a password using all special characters will be more secure than one using all numbers.
12. Participants will think that a password using all digits will be more secure than one using just lowercase letters.
13. Participants will think that a password using a birthday that is not theirs is more secure than one using their birthday.
14. Participants will think that a password using random digits is more secure than a password of equal length that is a date.
15. Participants will think that a password that has a lowercase letter replaced by a number will be more secure than a lowercase letter replaced by an uppercase letter.
16. Participants will think that a password with a letter replaced with a number will be more secure than a password with no replacements.
17. Participants will think that a password beginning with a digit will be more secure than one ending with a digit.

18. Participants will think that a password beginning with a digit will be more secure than one beginning with an uppercase letter.
19. Participants will think that a password of length 8 that uses three character classes will be more secure than a password of length 12 that only lowercase letters.
20. Participants will think that a password consisting of a commonly misspelled word is more secure than a password consisting of an easy-to-spell word of equal length.
21. Participants will think that a password that has an exclamation point at the end is much more secure than the same password without the exclamation point.

I developed these hypotheses through a combination of analyzing the results of our laboratory study of password creation [145], pilot testing this protocol with many of these hypotheses, and manually examining passwords over the last few years. For each hypothesis, I will test three example pairs of passwords. I will select at least one password in each pair from leaks of real passwords. To ensure that the passwords in each pair vary as minimally as possible from each other while enabling us to test the hypothesis, I will create most of the second passwords in each pair by manually modifying the leaked password in each pair. For all passwords, both leaked passwords and the passwords that I modify from leaked passwords, I will calculate their actual guessability using our Password Guessability Service framework [146]. For each participant, I will randomly select one of the three password pairs for each hypothesis, and I will also randomly select which password in the pair appears on the left.

Beyond examining these targeted hypotheses, I will also investigate how participants rate passwords overall, which will enable me to build a regression model to identify what characteristics users associate with strong passwords. In greater detail, each participant will assign a score on a 7-point scale (from "Completely Insecure" to "Completely Secure") to 20 passwords randomly selected from a subset of 8-character passwords taken from the breach of the RockYou gaming website. I will build an ordinal regression model in which the dependent variable is the score a participant assigned the password, while the independent variables will be characteristics like the length, the number of character classes, whether the password ends with a symbol, whether the password contains a character substitution, and similar characteristics. Because each participant will be rating multiple passwords, I will use a mixed model in which the ratings of individual passwords are grouped hierarchically under each participant.

A crucial component underlying a users' ratings of password strength is their attacker model, and thus their definition of what it means for a password to be secure. The final section of the study for each participant explores these questions. The following questions are examples of what I will ask in this part of the study, and each will refer to their primary email account (e.g., Gmail):

1. In your opinion, what does it mean for a password to be secure?
2. In your opinion, what characteristics make a password easy for an attacker to guess?
3. Please describe the type of attacker (or multiple types of attackers), if any, whom you worry might try to guess your password.
4. As far as you know, how do attackers try to guess your password?
5. How many guesses (by an attacker) would a password need to be able to withstand for you to consider it secure?

On a subsequent page, I will also have participants respond on a 7-point Likert scale ("strongly disagree" to "strongly agree") to statements about how guessing attacks may or may not work. These statements will cover targeted attacks (e.g., "looking at my Facebook profile" or "looking me up in a public directory, like a phone book"), brute-force attacks, and other large-scale guessing attacks (e.g., "using phrases found in the lyrics of popular songs," "adding digits and symbols to the end of dictionary words").

4.5 "Password Feedback" study [Not completed]

Building on the previously described studies, the passwords section of my thesis culminates in the construction and evaluation of a data-driven tool that provides users real-time feedback on the password they are creating. This study, the "Password Feedback" study, encompasses both a laboratory study and a follow-up online study. I have chosen this two-part design in order to gather initial, qualitative feedback on the design of the tool, but then to test it on a large-scale so that I can perform statistically meaningful comparisons.

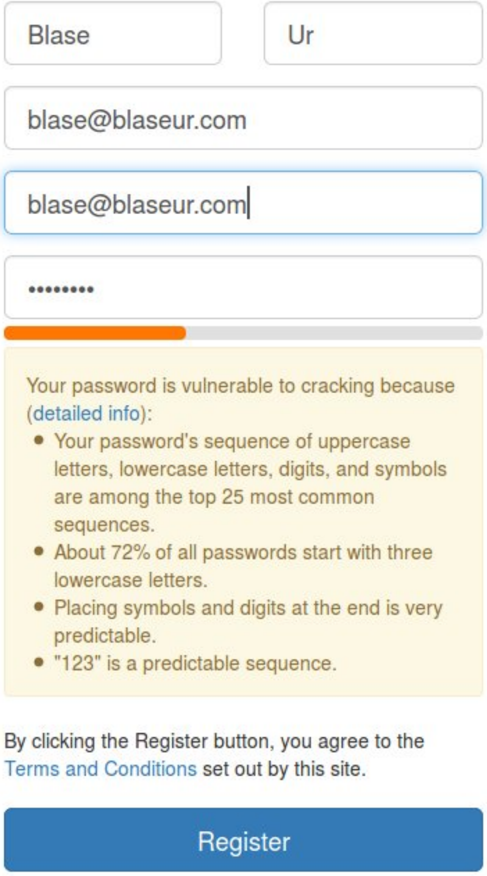
The password-feedback tool I am in the process of designing aims to provide users real-time feedback on the following characteristics of passwords:

- The use of dictionary words
- The use of common phrases
- The use of their own name/username
- The use of keyboard patterns
- The use of sequences (e.g., "abcdefg" or "helloolleh")
- The use of dates and years
- Using few unique characters in the password (e.g., "akakakakak")
- Adding digits or symbols to the end of a password
- Using a common password structure
- The use of common character substitutions

In addition, the tool will also warn specifically against other misconceptions we uncover as part of our "Perceptions" study. The tool will do so with the aim of disabusing users of those misconceptions.

The overriding question in this study, and in my thesis itself, is whether providing just-in-time data about security (or, in other cases, privacy) can help users make decisions. The series of follow-up questions about what precise representations provide value, and which do not, are also crucial. Through my two-part approach using different methods, I will capture data on both the overriding and follow-up questions in this study.

The first part of my approach will be a laboratory study of 10–20 participants. I intend to use this first section of the study to gather rich qualitative data on how participants perceive the data-driven tool, as well as how they perceive the framing of the study's scenario. Because I am particularly interested in studying how users create passwords when they expect that the password will likely be subjected to a large-scale guessing attack, I plan to depart from the protocol our passwords research group has employed for the last few years. In that protocol, participants make a new password for an online account. Instead, I will adopt



Blase

Ur

blase@blaseur.com

blase@blaseur.com

•••••••••

Your password is vulnerable to cracking because (detailed info):

- Your password's sequence of uppercase letters, lowercase letters, digits, and symbols are among the top 25 most common sequences.
- About 72% of all passwords start with three lowercase letters.
- Placing symbols and digits at the end is very predictable.
- "123" is a predictable sequence.

By clicking the Register button, you agree to the Terms and Conditions set out by this site.

Figure 4.2: An early prototype of the tool I will finish building and test in the "Password Feedback" study.

a scenario inspired by SpiderOak[1], Truecrypt, and similar tools that encrypt files under a key derived from a password because these sorts of passwords can always be subjected to a large-scale guessing attack. I will present participants with a short description of potential attacks against a computer with disk encryption that is unlocked using a password-derived key. I will then ask them to create a password for such a scheme, role-playing that this password will be used to secure their own personal computer containing their financial data, emails, and files.

As in our recent laboratory study of password creation [145], I will ask participants to think aloud as they create a password. I will also ask targeted questions about whether the meter gave any feedback that surprised them, as well as what they believe to be characteristics of passwords that are hard to guess or easy to guess. I hypothesize that the data-driven feedback from the meter will help participants learn what those characteristics are. As in the "Perceptions" study, I will also ask them to describe how they expect attackers to try to guess passwords. In the final section of the laboratory study, I will invite participants to freely test any additional passwords they are curious about using the meter, again having them think aloud to describe what they are testing. I will again ask them to describe what

[1] https://www.spideroak.com

characteristics make a password strong or weak. After the participant goes home, I will test whether they remember the password they created in the study. One week after the in-person study, participants will receive an email instructing them to enter their password on a website I will design to look like a computer log-in screen. They will have five attempts to enter the password correctly, and I will log their keystrokes as they attempt to log in. They will then answer a short survey on how they remembered their password, as well as provide their final thoughts on the data-driven password-feedback tool.

Although the portion of this overall study that uses a laboratory study will provide rich qualitative data, I also wish to individually test the impact of different aspects of the password-feedback tool at a scale that allows me to compare the security of passwords participants create in a statistically meaningful way. As a result, I will conduct a 1,600-participant, between-subjects online study in which participants again create a password in the presence of one of eight conditions specifying a variant of my data-driven password-feedback tool. The eight conditions are as follows:

1. (Control) No feedback whatsoever
2. The full-featured tool used in the laboratory study, revised to fix any shortcomings uncovered in the laboratory study
3. **Four conditions** representing variants of the tool in which one feature is removed to test its effectiveness. These variants will be chosen based on the outcomes of the laboratory study
4. The full-featured tool, but without any text explanations. That is, the same scoring algorithm will be used, but without giving the users any explanation as to why the password was scored that way
5. A traditional password-strength meter that scores passwords based on length and character-class heuristics

As with my previous study on password meters [143], I will log users' keystrokes and ask them our traditional battery of usability questions. Similar to the laboratory portion of this study, I will invite participants to return and re-enter their passwords one week later. At that time, I will also ask them questions about what characteristics of a password make the password hard for an attacker to guess or easy for an attacker to guess. I will also give them a second opportunity to test any passwords they want using the password-feedback tool, after which I will ask them to explain what new information, if any, they learned about password security as a result of using the tool. This study will conclude the passwords section of my thesis.

4.6 "Smart, Useful" study [SOUPS 2012]

I will then move to the privacy portion of my thesis. To gain an initial understanding of users' perceptions of online behavioral advertising (OBA) analogous to the "perceptions" study on passwords, I turned to purely qualitative methods in the "Smart, Useful" study [144]. In particular, my co-authors and I conducted 48 semi-structured interviews about OBA. We investigated non-technical users' attitudes about and understanding of OBA, using participants' expectations and beliefs to explain their attitudes. Participants found OBA to be simultaneously useful and privacy invasive. They were surprised to learn that browsing history is currently used to tailor advertisements, yet they were aware of contextual targeting.

Our results identify mismatches between participants' mental models and current approaches for providing users with notice and choice about OBA. Participants had strong concerns about data collection, and the majority of participants believed that advertisers collect personally identifiable information. Overall, few participants felt confident that they understood precisely what information was being collected by advertisers, or how. For example, many participants thought advertisers were accessing users' credit card information that would be stored in cookies on their computer, and they did not understand how behavioral profiling might work. These misunderstandings are not surprising, however. The process of behavioral targeting is very convoluted to explain, beginning with storing a unique, randomly generated token on a user's computer in the form a persistent cookie, protected by the same-origin policy. As a user visits a page, potentially dozens of third-party HTTP requests will be generated in the background, and the unique identifier that is automatically sent as an HTTP header enables advertisers to correlate a user's visits to different webpages and then use machine-learning techniques to infer the user's interests and subsequently decide which advertisements to show.

In addition, participants misunderstood the role of advertising networks, basing their opinions of an advertising network on that company's non-advertising activities. Participants' attitudes towards OBA were complex and context-dependent. While many participants felt tailored advertising could benefit them, I found in this study that existing notice and choice mechanisms are not effectively reaching users, suggesting the need for data-driven tools to provide users greater insight into data collection for OBA purposes.

4.7 "Visualizing OBA" study [Not completed]

Although a number of privacy tools (e.g., Ghostery, Lightbeam, and Privacy Badger) can help users control OBA, average users are left utterly confused about OBA [144] even after using such tools [87]. I propose moving beyond these existing tools, which alert users to tracking occurring at the current moment, by designing and testing a tool that takes a data-driven, personalized approach to privacy awareness. I will conduct a 90-participant, 2-week field trial comparing different visualizations of personalized tracking data, as I detail below.

I hypothesize that users can better understand OBA and resultant privacy threats if equipped with a tool that visualizes instances of them being tracked over time. That is, I hypothesize that showing "Doubleclick knows you visited the following 82 pages" or showing "Doubleclick has concluded you like 'ice hockey' based on the following 15 pages you have visited" will be more effective than current tools' approach of notifying the user about tracking happening at the moment (e.g., "Doubleclick is a third party on the current page"). Studies have shown benefits in notifying users about the collection of data by smartphone apps [3, 7, 60]. My proposal translates these insights to the OBA domain, yet makes further intellectual contributions by exploring the impact of presenting different abstractions and granularities of the information tracked.

My proposed tool and accompanying user study build on my hypothesis that integrating personalized behavioral data, such as a user's own browsing, into a privacy tool can better equip users to understand risks and make privacy decisions they are more confident about. In particular, our tool will "track the tracking" that occurs as an individual user browses the web, storing *which companies* have tracked the user on *what* webpages (and *when*) in a local database on the user's computer. When the user encounters these third-party trackers in future browsing, the tool will provide personalized, longitudinal information about that particular company's tracking. One of the representations of this history that we will test is a layered view of the different websites (and individual pages) a company knows the user has visited. We will also design summary visualizations of this information and integrate into the tool the visualizations with the potential to be most illuminating.

While we hypothesize surfacing the browing information advertisers collect can help users understand OBA, we believe it is even more crucial to show users our best guesses of the inferences advertisers have made about the user's interests. Doing so will help users understand what advertisers *do with* the information they collect, a critical step in privacy awareness [157]. Our tool will scrape inferred information that some advertising companies provide to users through "privacy dashboards." Since most companies are not transparent about their inferences, however, our tool will use the rough estimates described below to make an educated guess about inferences. The tool will show users these best cases about potentially inferred interests.

We have conducted initial pilot studies investigating what kinds of visual representations of longitudinal tracking information users might find useful. Whereas only 44% of pilot participants felt that current representations of tracking (which companies are tracking you on the current website) are useful, two-thirds of participants stated that a plugin showing tracking over time would be useful, and over two-thirds of participants stated that a plugin showing the inferences companies had made about them would be useful. We hypothesize that these two visualizations will help mitigate users' common misconceptions about what kinds of data are actually collected for OBA purposes [87,144] and what inferences could be made about this data [157], partially bridging the large disconnect between privacy leakage and protection measures [82].

The visualizations supported by existing privacy tools are far more limited. Tools like Ghostery and Disconnect provide pop-ups during browsing listing what third-party companies are tracking the user on that page, yet lack longitudinal information. Futhermore, these tools do not attempt to show what those companies might be inferring. Mozilla Lightbeam instead graphs the interrelationships between third-party advertisers and first-party websites, yet only on a very abstract level.

I will create a Google Chrome plugin that "tracks the trackers" locally and also locally makes an educated case about what an advertiser may have inferred about the user. I will then investigate the impact of presenting the user with different abstractions of when he or she has been tracked, as described below. To investigate this question through study participants' actual browsing, our study will take the form of a 90-participant, 2-week field trial. Participants will participate remotely over the Internet, downloading our plugin on their own computer.

All participants will install our Chrome plugin and use it in the background for one week

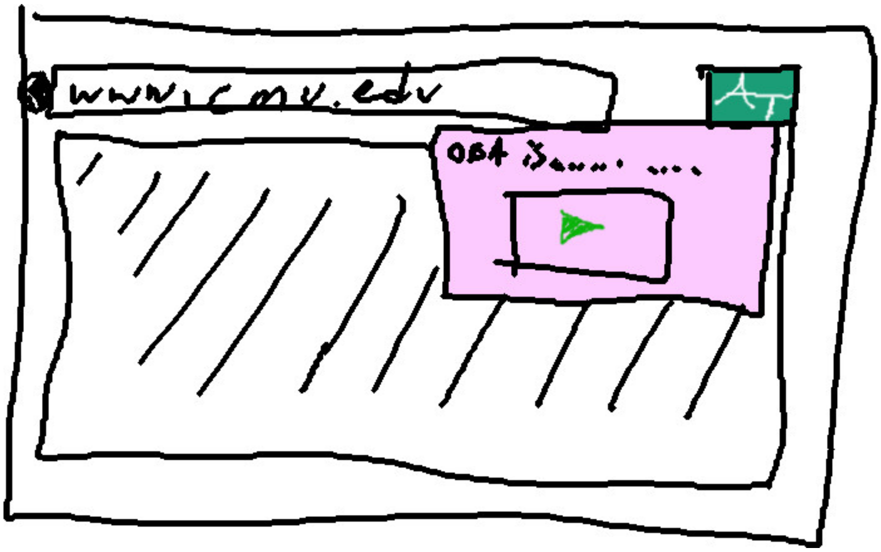


Figure 4.3: A lo-fi mockup of Condition 1, which shows a generic description of OBA and embedded informational video about OBA when the plugin's icon is clicked.

while the plugin collects data for bootstrapping the tool. During this first week, the plugin's logo will be visible in the browser toolbar, but there will be no pop ups. If the user clicks on the logo, only informational text about the study will be visible. At the start of the second week of the study, the tool will inform the participant that he or she is welcome and encouraged to view information from the tool in the course of their browsing, and I will instrument the tool to record these interactions. The tool is also capable of blocking third-party tracking, and I will enable participants to enable this blocking if they so choose, recording if they do so.

In the final (fourteenth) day of the study, the tool will force all participants to interact with it, regardless of their previous interaction with the tool. Following this interaction, the participant will take a survey. The survey will measure each of the following aspects I hypothesize might be impacted by the data visualized: accurate knowledge about OBA; awareness of potential privacy leaks; OBA privacy attitudes; the likelihood of blocking third-party tracking even if it means no longer receiving relevant ads; and perceptions of the plugin's utility. The data will include responses on 7-point Likert scales, augmented by open-ended questions for clarification. The plugin will also automatically collect information about users' interactions with the different interface elements of the tool.

I will assign participants round-robin to one of the following six conditions specifying what the tool does:

1. (Control; generic information) The participant will watch a short video about OBA at the beginning of the study and use a plugin that provides only a generic informational pop-up about OBA when the user clicks on it. There will be no pop-up windows except whern the user clicks on the tool. See Figure 4.3.
2. (Condition 1 + Connections) This condition augments the video and generic OBA information from Condition 1 with a Mozilla Lightbeam-inspired graph showing how tracking companies and websites are connected and correlated. See Figure 4.4.
3. (Condition 1 + Current trackers) This condition includes the video and generic de-

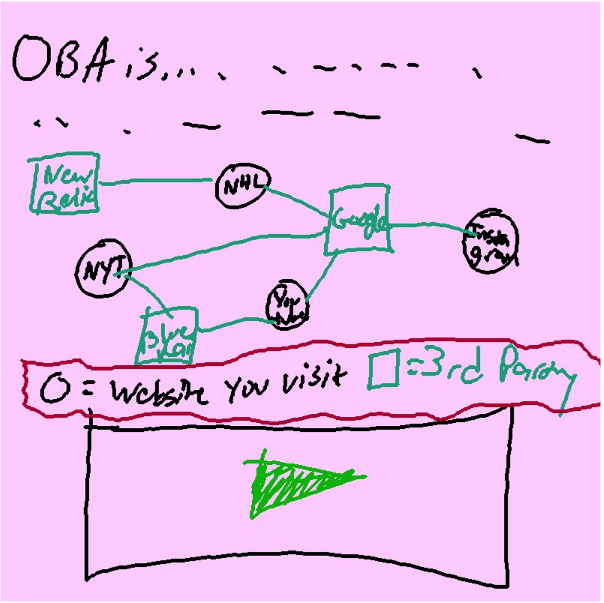


Figure 4.4: A lo-fi mockup of the windows that appears in Condition 2 when the plugin's icon is clicked. The graphic that shows the connections between advertisers and websites is based on Mozilla Lightbeam.

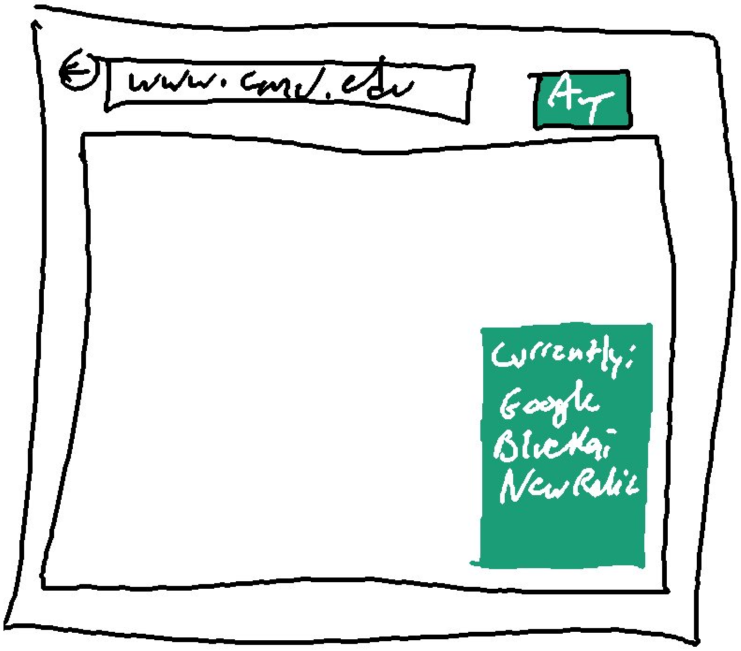


Figure 4.5: A lo-fi mockup of Condition 3, which lists the companies tracking the user at the current moment.

scription of OBA from Condition 1. Analogous to tools like Ghostery and Disconnect, the browser plugin will also provide a small pop-up on each site the user visits listing the advertising companies tracking them on that site. See Figure 4.5.

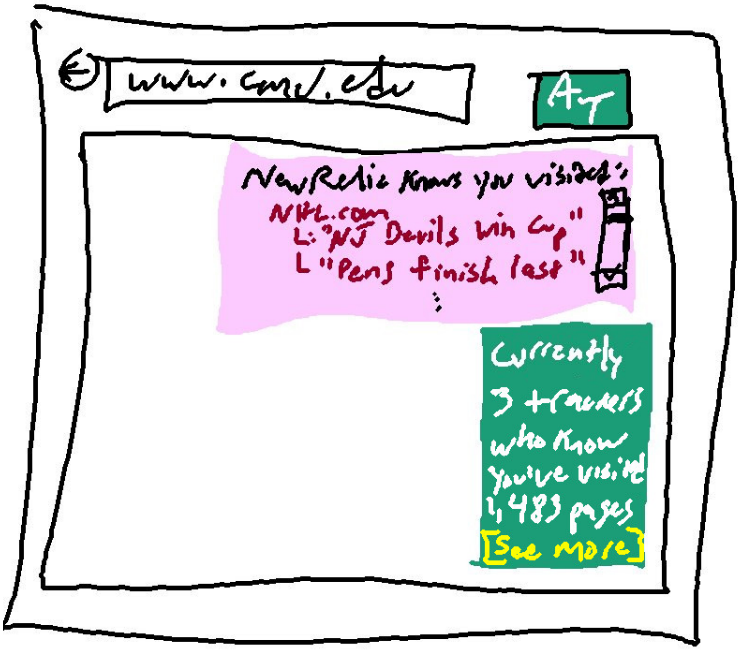

www.cmu.edu
AT
NewRelic knows you visited:
NHL.com
L: "NJ Devils win cup"
L "Pens finish last"
Currently 3 trackers who know you've visited 1,483 pages
[See more]

Figure 4.6: A lo-fi mockup of Condition 4, which pops up a message like "Currently 3 trackers who know you've visited 1,483 pages [see more]." If the user clicks "see more" or clicks on the plugin's icon, they will see the annotated data of what companies have tracked.

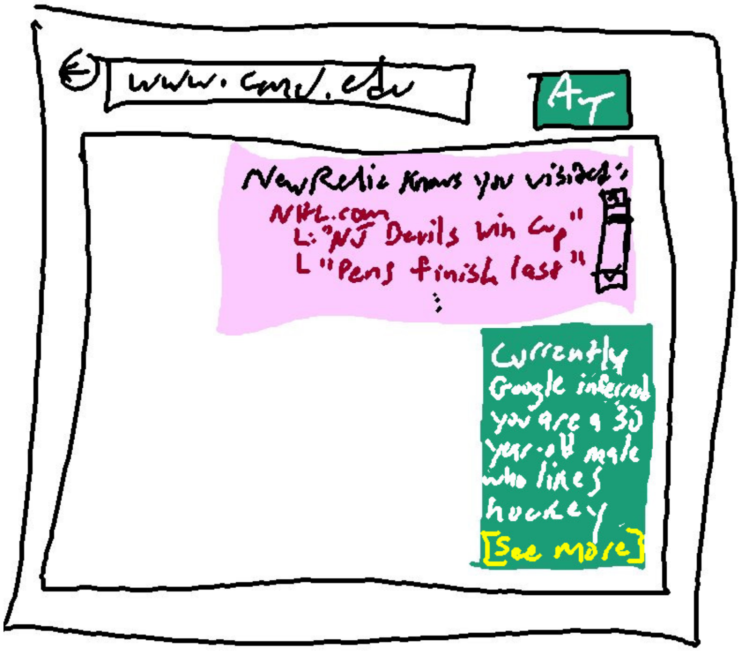

www.cmu.edu
AT
NewRelic knows you visited:
NHL.com
L: "NJ Devils win cup"
L "Pens finish last"
Currently Google inferred you are a 30 year-old male who likes hockey
[See more]

Figure 4.7: A lo-fi mockup of Condition 5, which is similar to Condition 4, yet also shows inferences (both demographic and interest-based) scraped from the ad-interest managers of companies like Google.

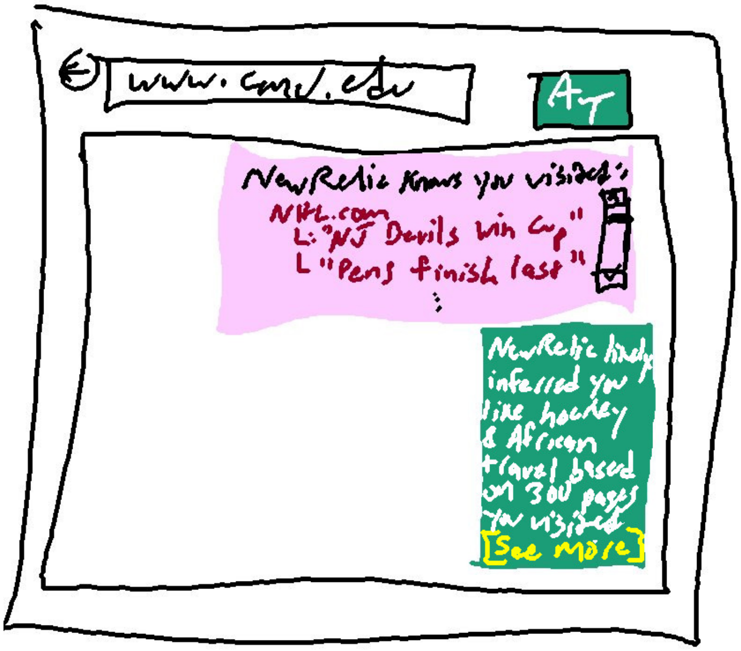

www.cmu.edu
AT
NewRelic knows you visited:
NHL.com
L: "NJ Devils win Cup"
L "Pens finish last"
NewRelic likely inferred you like hockey & African travel based on 300 pages you visited
[See more]

Figure 4.8: A lo-fi mockup of Condition 6, which is similar to Condition 4, yet also shows our rough approximation of the interests an advertising company may have inferred about a user based on the pages the user has visited.

4. (Condition 1 + List of trackers, annotated) This condition includes the video and generic description of OBA from Condition 1. In addition, the plugin will present detailed visualizations, in addition to access to the underlying data, of how each company had tracked a user over time when the user clicks on the tool. For example, the user could see that DoubleClick had tracked them on 53 different websites in the past month, as well as the precise pages tracked. On each site the user visits, an informational pop-up will summarize the data that will be shown if the user clicks on the tool's icon. See Figure 4.6.
5. (Condition 4 + companies' conclusions) This condition includes everything from Condition 4. In addition, the plugin will show users data scraped from the ad-interest managers of the companies that offer such things (e.g., Google, Yahoo, and BlueKai). See Figure 4.7.
6. (Condition 4 + our estimated conclusions) While a few companies provide limited transparency about their inferences, most do not. In this condition, we use heuristics to estimate what those companies may have concluded. This condition includes everything from Condition 4. Using the same underlying data, the plugin will also estimate the interests an advertiser might have inferred based on the user's browsing. The plugin will present pop-ups on different websites showing these inferences drawn from our own estimated inferences, with the full data from Condition 4 available when the user clicks on the tool's icon. See Figure 4.8.

I am currently working with two interns in our lab to develop the data-driven privacy tool that will enable this study. We are currently developing the plugin for Google Chrome because Chrome is the most popular browser worldwide, although we may also port the plugin to Mozilla Firefox time permitting. Conditions 1–2 are relatively trivial to implement, although we have not yet implemented them. In addition, in its current form, the prototype tool already mostly supports Condition 3–5, above.

Each time a user visits a webpage, our prototype tool currently writes to Chrome's local storage each first-party URL the user visits, along with the time stamp. In addition, using a list from the open-source Disconnect tool[2] mapping the domains of URLs to different tracking companies, the prototype tool already writes to Chrome local storage which third-party tracking companies spawned requests on that first-party URL. These features of our prototype tool enable Conditions 3-4.

While helping users explore on precisely which webpages different companies have tracked them is relatively straightforward to implement, I also hypothesize that users will find it valuable to know what those companies may have inferred about their interests. Currently, our Chrome plugin periodically opens in a background tab the ad-interest managers from Google and Yahoo, screen-scraping the interests those ad-interest managers report and saving them in local storage. We have also been exploring ad-interest managers from BlueKai, Exelate, and Lotame, though we have found those to be less comprehensive and less accurate in the interests and demographics they report. Time permitting, we may add support for those three ad-interest managers.

Approximating what interests a user may have based on his or her browsing is less straightforward, though currently in progress. Unfortunately, advertising companies are generally not forthcoming about how they make inferences about users' interests based on browsing data, so we are resigned to make our best guess. To approximate interests, I developed a set of keywords for each possible interest category as detailed below. The plugin will then scan the body text (stripping HTML) of each webpage a user visits for those keywords, noting the frequency of the relevant keywords observed by each third-party tracking company.

In more detail, to determine all possible interest categories, we first scraped the full list of hundreds of interest categories from the source code of Google's ad-interest manager. To determine in a principled way what keywords to associate with each interest category, I am relying on Term Frequency-Inverse Document Frequency (*TF-IDF*) [124], a metric used in computational linguistics to determine each unique word's importance to a particular document in a corpus of documents. For each interest category, I will manually identify one or more Wikipedia articles that describe that interest category (e.g., the Wikipedia articles "India" and "Tourism in India" for the interest category Travel:India). I will consider all words above a constant TF-IDF threshold (to be determined heuristically via experimentation) on those pages to be keywords for that interest category. In the course of a user's browsing, if he or she has visited pages containing a threshold proportion (to be detemined heuristically via experimentation) of different keywords for a particular interest category, the prototype tool will register that they likely have that interest. Because the prototype tool already determines which third-party trackers are aware the user has visited a given website, the tool can determine which third-party trackers have likely inferred different interest categories.

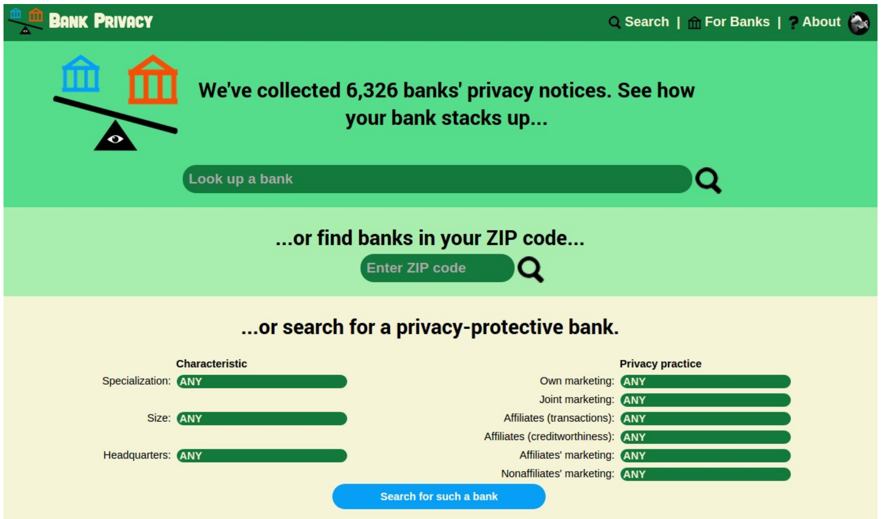
Measuring changes in users' awareness of OBA practices and willingness to act based

[2] https://disconnect.me/

on this knowledge is critical because users recognize both benefits in receiving relevant ads and drawbacks to the tracking of their browsing, leaving them confused and conflicted [144]. Existing privacy tools can lead to inaccurate models of OBA; users often believe third parties collect either far more or far less information than they actually do [87]. My proposed study aims to give users more transparency about OBA.

4.8 "Financial" study [Not completed]

The final section of my thesis investigates whether just-in-time information about financial institutions' privacy practices, as well as user-specific information about institutions that have local branch locations, would impact consumers' willingness to switch to a more privacy-protective institution. Over the past few years, my co-authors and I have been analyzing the privacy practices of U.S. financial institutions [**?**, 31]. In particular, I wrote code that automatically searches for, downloads, and parses PDF privacy notices that follow the model privacy form developed in the wake of the Gramm-Leach-Bliley Act (GLBA) [112]. Federal regulators worked with privacy consultants to develop these standardized notices, which provide a safe harbor for financial institutions to make the annual privacy disclosures to consumers required under the GLBA. Previously, we used the data set of over 6,000 institutions' privacy practices that I extracted from these notices to map the landscape of privacy practices across the U.S. financial industry, as well as to search for correlations between institutions' demographic characteristics and their privacy practices [**?**, 31].



We've collected 6,326 banks' privacy notices. See how your bank stacks up...

...or find banks in your ZIP code...

...or search for a privacy-protective bank.

Characteristic

Specialization: ANY

Size: ANY

Headquarters: ANY

Privacy practice

Own marketing: ANY

Joint marketing: ANY

Affiliates (transactions): ANY

Affiliates (creditworthiness): ANY

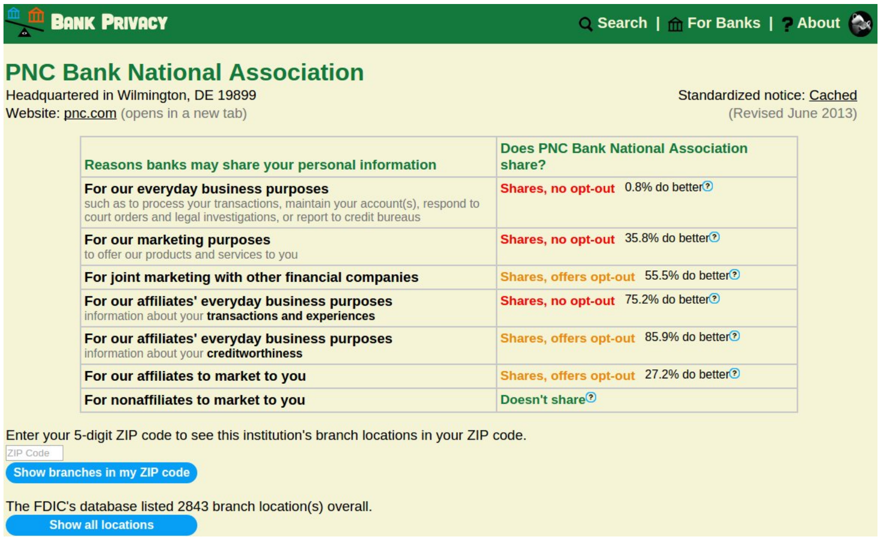
Affiliates' marketing: ANY

Nonaffiliates' marketing: ANY

Figure 4.9: The Bank Privacy website.

In support of my proposed study, I have used the data set I built to construct the CMU "Bank Privacy" interactive website.[3] I designed this site (Figure 4.9) to provide consumers easy access to data on the privacy practices of financial institutions. For instance, a consumer can search for their financial institution by name and see how its privacy practices compare to the other institutions in our data set (e.g., "First Bank shares for its marketing purposes,

[3] `http://cups.cs.cmu.edu/bankprivacy/`



PNC Bank National Association

Headquartered in Wilmington, DE 19899
Website: pnc.com (opens in a new tab)

Standardized notice: Cached
(Revised June 2013)

Reasons banks may share your personal information	Does PNC Bank National Association share?
For our everyday business purposes such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus	Shares, no opt-out 0.8% do better?
For our marketing purposes to offer our products and services to you	Shares, no opt-out 35.8% do better?
For joint marketing with other financial companies	Shares, offers opt-out 55.5% do better?
For our affiliates' everyday business purposes information about your **transactions and experiences**	Shares, no opt-out 75.2% do better?
For our affiliates' everyday business purposes information about your **creditworthiness**	Shares, offers opt-out 85.9% do better?
For our affiliates to market to you	Shares, offers opt-out 27.2% do better?
For nonaffiliates to market to you	Doesn't share?

Enter your 5-digit ZIP code to see this institution's branch locations in your ZIP code.

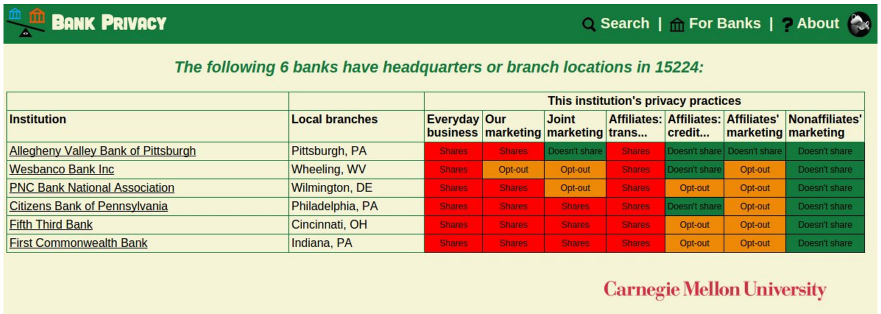
ZIP Code

Show branches in my ZIP code

The FDIC's database listed 2843 branch location(s) overall.

Show all locations

Figure 4.10: An example financial institution's page on the Bank Privacy website showing comparative information about the institution's privacy practices.



The following 6 banks have headquarters or branch locations in 15224:

		This institution's privacy practices						
Institution	**Local branches**	**Everyday business**	**Our marketing**	**Joint marketing**	**Affiliates: trans...**	**Affiliates: credit...**	**Affiliates' marketing**	**Nonaffiliates' marketing**
Allegheny Valley Bank of Pittsburgh	Pittsburgh, PA	Shares	Shares	Doesn't share	Shares	Doesn't share	Doesn't share	Doesn't share
Wesbanco Bank Inc	Wheeling, WV	Shares	Opt-out	Opt-out	Shares	Doesn't share	Opt-out	Doesn't share
PNC Bank National Association	Wilmington, DE	Shares	Shares	Opt-out	Shares	Opt-out	Opt-out	Doesn't share
Citizens Bank of Pennsylvania	Philadelphia, PA	Shares	Shares	Shares	Shares	Doesn't share	Opt-out	Doesn't share
Fifth Third Bank	Cincinnati, OH	Shares	Shares	Shares	Shares	Opt-out	Opt-out	Doesn't share
First Commonwealth Bank	Indiana, PA	Shares	Shares	Shares	Shares	Opt-out	Opt-out	Doesn't share

Carnegie Mellon University

Figure 4.11: Example search results on the Bank Privacy website.

yet 35.8% of banks do better from a privacy perspective"), as shown in Figure 4.10. Similarly, a consumer can enter their ZIP code and see banks with branches in his or her ZIP code or nearby ZIP codes (Figure 4.11). In its public release, the Bank Privacy website provides search results ordered beginning with those banks that share consumers' personal information for the fewest reasons.

A major remaining question, however, is whether providing consumers with large-scale, comparative information about financial institutions' privacy practices will actually impact their behavior. To examine this question, I will conduct a study in which participants use a copy of the Bank Privacy website, renamed "Bank Search," to learn more about local banks by completing a search task that I designed to be a common scenario in which privacy decisions are not the primary focus of the task. I detail this task below. Following the task, I will present the participant with survey questions covering whether the presence of comparative privacy information impacts participants' stated willingness to consider switching banks, the considerations that impact their willingness, and the degree to which a desire to find a more privacy-protective bank does or does not outweigh the logistical barriers to switching.

In more detail, I will recruit 800 participants online using Amazon's Mechanical Turk platform. I will screen for participants age 18+ from the United States because the Bank

Privacy website only supports the United States. Participants will be assigned round-robin to one of the following four conditions for the Bank Search website:

- (Control condition) The participant will use a variant of the Bank Search website that does not include any privacy information. Search results will be ordered alphabetically.
- The participant will use a variant of the Bank Search website that includes privacy information. Search results will be ordered alphabetically.
- The participant will use a variant of the Bank Search website that includes privacy information. All search results will be ordered from most privacy-friendly to least privacy-friendly institution.
- The participant will use a variant of the Bank Search website that includes privacy information, but all comparative privacy information (e.g., "35.8% of banks do better") will be removed. Furthermore, search results will not provide an overview of privacy information, but participants can see each bank's privacy information by viewing that particular bank. Search results will be ordered alphabetically.

Participants will first name their current primary banking institution, as well as provide other demographic information (e.g., age range, occupation). They will then rate, on a 5-point Likert scale from "not at all important" to "very important," how much different factors matter when they are choosing a financial institution. These factors will encompass privacy practices, as well as non-privacy factors (e.g., whether the local branch is open on weekends, the number of ATM withdrawals permitted each month without extra fees, the security of the bank's website).

I will then ask participants to use Bank Search, in the variant specified by their condition, to complete the following two-part search task. First, I will have the participant use the Bank Privacy website and their financial institution's website (linked directly from the Bank Privacy site) to search for the Saturday opening hours (if any) and address of that institution's closest location, as well as the bank's current interest rate for savings accounts (if posted). I will then ask the participant to choose another local financial institution that seems potentially compelling. The participant will then repeat those tasks for the competing institution.

I will split the Bank Search website into two sections, corresponding to these two tasks. The first section will only allow the user to search by institution name or website URL, whereas the second section will allow the user to search only by ZIP code. In all conditions, I will instrument the Bank Search website to log the pages participants visit, in order.

After participants complete the first task using the Bank Search website, I will ask them if they learned anything new about their financial institution from the site, hypothesizing that some aspect of the bank's privacy practices, how their institution's privacy practices compare to other institutions, or the location of a nearby branch would be the most likely answers. I will also accept as an answer that they learned nothing new.

After participants complete the second task, I will have them rate, on a 5-point Likert scale ("not at all willing" to "very willing") how willing or unwilling they would be to consider switching financial institutions. If they state they would be at least somewhat willing (2–5 on the 5-point scale), I will have them explain in a free response what factors would make them willing to switch. If they state they would not be willing to switch institutions (1 on the 5-point scale), I will have them explain in a free response why they would not be willing to switch institutions. I will have two coders perform qualitative coding on these free responses to map out what factors plan into decisions to switch, or not to switch, institutions. I will give special attention to factors related to privacy and security. In both cases, on a subsequent page I will have participants rate on 5-point Likert scales the expected impact of different hypothesized barriers to switching financial institutions.

Chapter 5

Logistics

5.1 Thesis Outline

INTRODUCTORY CHAPTERS

- **Chapter 1**: Introduction [to be written in **January 2016**]
- **Chapter 2**: Related work on using data in security and privacy. [Draft in this proposal; to be revised in **January 2016**]

SECURITY OF PASSWORDS

- **Chapter 3**: Introduction to the password part of the thesis and related work about passwords [to be written in **January 2016**]
- **Chapter 4 ("Meters")**: The impact of password-strength meters [**USENIX Security 2012**]
- **Chapter 5 ("Biases")**: Understanding Biases in Modeling Password Cracking [**USENIX Security 2015**]
- **Chapter 6 ("Art")**: Art of Password Creation: Semantics and Structures [to be completed and submitted to TOCHI in **August 2015**]
- **Chapter 7 ("Perceptions")**: Perceptions of Password Security [to be completed in **August 2015**]
- **Chapter 8 ("Password Feedback")**: The impact of targeted feedback on what's wrong with your password [to be completed in **December 2015**]

ONLINE PRIVACY

- **Chapter 9**: Introduction to the privacy part of the thesis and related work about OBA and financial privacy [to be written in **February 2016**]
- **Chapter 10 ("Financial")**: The impact of merged privacy information on financial decisions [to be completed in **September 2015**]

- **Chapter 11 ("Smart, Useful")**: User perceptions of OBA: Smart, Useful, Scary, Creepy [**SOUPS 2012**]
- **Chapter 12 ("Visualizing OBA")**: The impact of visualizing OBA [to be completed in **November 2015**]

CONCLUSION

- **Chapter 13**: Conclusions and future directions [to be completed in **April 2016**]

5.2 Monthly Timeline

July 2015

- Apply for "Financial" IRB approval.
- Develop "Financial" website and study infrastructure.

August 2015

- Finish developing "Financial" website and study infrastructure.
- Apply for "Visualizing OBA" IRB approval.
- Recruit participants and collect/analyze data for "Perceptions" study.
- Write paper for "Perceptions" and submit it to a conference.
- Revise "Art" paper and submit it to a journal.

September 2015

- Recruit participants and collect/analyze data for "Financial" study.
- Write paper for "Financial" and submit it to a conference.

October 2015

- Complete development of "Visualizing OBA" plugin.
- Recruit participants and collect data for "Visualizing OBA" study.
- Apply for "Password Feedback" IRB approval.
- Write academic job application materials.
- Begin applying for academic jobs.

November 2015

- Write paper for “Visualizing OBA” and submit it to a conference.
- Finish developing infrastructure for “Password Feedback.”
- Continue applying for academic jobs.

December 2015

- Recruit participants and collect data for “Password Feedback” paper.
- Write paper for “Password Feedback” and submit it to a conference.
- Write the chapter introducing the security part of the thesis and covering related work on passwords (Chapter 3).
- Continue applying for academic jobs.

January 2016

- Write the introduction chapter (Chapter 1).
- Write the related work chapter (Chapter 2).
- Integrate previously published (“Meters,” “Biases,”) and submitted (“Art,” “Perceptions,” “Password Feedback”) papers about passwords into the thesis. (Chapters 4–8)
- Write the chapter introducing the privacy part of the thesis and covering related work on privacy (Chapter 9).

February 2016

- Integrate previously published (“Smart, Useful”) and submitted (“Financial,” “Visualizing OBA”) privacy papers into the thesis. (Chapters 10–12)
- (Hopefully) interview for academic jobs.

March 2016

- Write the conclusions chapter (Chapter 13) of the thesis.
- (Hopefully) interview for academic jobs.

April 2016

- Defend the thesis.

May 2016

- (Hopefully) graduate.

5.3 Acknowledgments

In recognition of their excellent feedback and support, I gratefully acknowledge my five committee members: Alessandro, Jason, Lorrie, Lujo, and Mike.

I also thank the many additional co-authors (and, often, friends) who have made the work I report in this thesis proposal possible: Maung Aung, Jonathan Bees, Adam Buchinsky, Nicolas Christin, Adam L. Durity, Phillip (Seyoung) Huh, Kelly Idouchi, Patrick Gage Kelly, Saranga Komanduri, Darya Kurilova, Joel Lee, Pedro Giovanni Leon, Michael Maass, Stephanos Matsumoto, Michelle Mazurek, William Melicher, Rupal Nahar, Timothy Passaro, Sean M. Segreti, Richard Shay, Manya Sleeper, Ashwin Srinivasan, Timothy Vidas, and Yang Wang.

Bibliography

[1] Saeed Abu-Nimeh, Dario Nappa, Xinlei Wang, and Suku Nair. A comparison of machine learning techniques for phishing detection. In *Proc. APWG eCrime Researchers Summit*, 2007.

[2] Anne Adams and Martina Angela Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, 1999.

[3] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Cranor, and Yuvraj Agarwal. Your location has been shared 5,398 times! A field study on mobile app privacy nudging. In *CHI*, 2015.

[4] Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. Usable transparency with the data track: A tool for visualizing data disclosures. In *Proc. CHI Extended Abstracts*, 2015.

[5] Annie I Antón, Julia B Earp, Qingfeng He, William Stufflebeam, Davide Bolchini, and Carlos Jensen. Financial privacy policies and the need for standardization. *IEEE Security & Privacy*, 2(2):36–45, 2004.

[6] Adam J. Aviv and Dane Fichter. Understanding visual perceptions of usability and security of android's graphical password pattern. In *Proc. ACSAC*, pages 286–295, 2014.

[7] Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor, and Carolyn Nguyen. "Little brothers watching you:" Raising awareness of data leaks on smartphones. In *Proc. SOUPS*, 2013.

[8] Rebecca Balebako, Pedro G. Leon, Richard Shay, Blase Ur, Yang Wang, and Lorrie Faith Cranor. Measuring the effectiveness of privacy tools for limiting behavioral advertising. In *Proc. W2SP*, 2012.

[9] Francesco Bergadano, Bruno Crispo, and Giancarlo Ruffo. Proactive password checking with decision trees. In *CCS*, 1997.

[10] André Bergholz, Gerhard Paa, Frank Reichartz, Siehyun Strobel, and Jeong-Ho Chang. Improved phishing detection using model-based features. In *Proc. CEAS*, 2008.

[11] Matt Bishop and Daniel V. Klein. Improving system security via proactive password checking. *Computers & Security*, 14(3):233–249, 1995.

[12] Joseph Bonneau. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *Proc. IEEE Symp. Security & Privacy*, 2012.

[13] Joseph Bonneau. Statistical metrics for individual password strength. In *Proc. WPS*, 2012.

[14] Joseph Bonneau and Sören Preibusch. The password thicket: technical and market failures in human authentication on the web. In *Proc. WEIS*, 2010.

[15] Joseph Bonneau and Stuart Schechter. Towards reliable storage of 56-bit secrets in human memory. In *Proc. USENIX Security*, 2014.

[16] Joseph Bonneau and Ekaterina Shutova. Linguistic properties of multi-word passphrases. In *Proc. USEC*, 2012.

[17] Joseph Bonneau and Rubin Xu. Of contraseñas, sysmawt, and mìmă: Character encoding issues for web passwords. In *Proc. W2SP*, 2012.

[18] Kay Bryant and John Campbell. User behaviours associated with password security and management. *Australasian Journal of Information Systems*, 14(1), 2006.

[19] William E. Burr, Donna F. Dodson, and W. Timothy Polk. Electronic authentication guideline. Technical report, NIST, 2006.

[20] Damon E. Campbell and Ryan T. Wright. Shut-up I don't care: Understanding the role of relevance and interactivity on customer attitudes toward repetitive online advertising. *Journal of Electronic Commerce Research*, 9(1):62–76, 2008.

[21] Carnegie Mellon University. Password guessability service. `https://pgs.ece.cmu.edu`, 2015.

[22] Claude Castelluccia, Markus Dürmuth, and Daniele Perito. Adaptive password-strength meters from Markov models. In *Proc. NDSS*, 2012.

[23] Sonia Chiasson, Alain Forget, Elizabeth Stobert, P. C. van Oorschot, and Robert Biddle. Multiple password interference in text passwords and click-based graphical passwords. In *Proc. CCS*, 2009.

[24] Hsien-Cheng Chou, Hung-Chang Lee, Hwan-Jeu Yu, Fei-Pei Lai, Kuo-Hsuan Huang, and Chih-Wen Hsueh. Password cracking based on learned patterns from disclosed passwords. *IJICIC*, 2013.

[25] Yiannis Chrysanthou. Modern password cracking: A hands-on approach to creating an optimised and versatile attack. Master's thesis, Royal Holloway, University of London, 2013.

[26] Federal Trade Commission. Privacy of consumer financial information; final rule. Federal Register, May 2000.

[27] Sunny Consolvo, Jaeyeon Jung, Ben Greenstein, Pauline Powledge, Gabriel Maganis, and Daniel Avrahami. The wi-fi privacy ticker: improving awareness & control of personal information exposure on wi-fi. In *Proc. Ubicomp*, 2010.

[28] Lorrie Faith Cranor. *Web privacy with P3P.* O'Reilly, 2002.

[29] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *Journal on Telecommunications and High Technology Law*, 10:273–307, 2012.

[30] Lorrie Faith Cranor, Serge Egelman, Steve Sheng, Aleecia M McDonald, and Abdur Chowdhury. P3P deployment on websites. *Electronic Commerce Research and Applications*, 7(3):274–293, 2008.

[31] Lorrie Faith Cranor, Kelly Idouchi, Pedro Giovanni Leon, Manya Sleeper, and Blase Ur. Are they actually any different? Comparing thousands of financial institutions' privacy practices. In *Proc. WEIS*, 2013.

[32] Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. The tangled web of password reuse. In *Proc. NDSS*, 2014.

[33] Xavier de Carné de Carnavalet and Mohammad Mannan. From very weak to very strong: Analyzing password-strength meters. In *Proc. NDSS*, 2014.

[34] Matteo Dell'Amico, Pietro Michiardi, and Yves Roudier. Password strength: An empirical analysis. In *Proc. INFOCOM*, 2010.

[35] Martin M.A. Devillers. *Analyzing Password Strength.* PhD thesis, Radboud University Nijmegen, 2010.

[36] Digital Advertising Alliance. Self-Regulatory Principles for Online Behavioral Advertising. `http://www.aboutads.info/resource/download/seven-principles-07-01-09.pdf`, July 2009.

[37] Markus Dürmuth, Fabian Angelstorf, Claude Castelluccia, Daniele Perito, and Abdelberi Chaabane. OMEN: Faster password guessing using an ordered markov enumerator. In *Proc. ESSoS*, 2015.

[38] Markus Dürmuth, Abdelberi Chaabane, Daniele Perito, and Claude Castelluccia. When privacy meets security: Leveraging personal information for password cracking. *CoRR*, 2013.

[39] Peter Eckersley. How unique is your web browser? EFF report, Electronic Frontier Foudation, 2009.

[40] Serge Egelman, Lorrie Faith Cranor, and Abdur Chowdhury. An analysis of p3p-enabled web sites among top-20 search results. In *Proc. ICEC*, 2006.

[41] Serge Egelman, Andreas Sotirakopoulos, Ildar Muslukhov, Konstantin Beznosov, and Cormac Herley. Does my password go up to eleven? the impact of password meters on password selection. In *Proc. CHI*, 2013.

[42] William Enck, Peter Gilbert, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N. Sheth. Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. In *Proc. OSDI*, 2010.

[43] Sascha Fahl, Marian Harbach, Yasemin Acar, and Matthew Smith. On The Ecological Validity of a Password Study. In *Proc. SOUPS*, 2013.

[44] Fair and Accurate Credit Transactions Act. Pub. L. No. 108-159, 117 Stat. 1952, 2003.

[45] Ayman Farahat and Michael Bailey. How effective is targeted advertising? *WWW 2012*.

[46] Federal Trade Commission. Privacy online: A report to Congress, June 1998.

[47] Federal Trade Commission. Self-regulatory principles for online behavioral advertising. `http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf`, 2009.

[48] Dinei Florêncio and Cormac Herley. A large-scale study of web password habits. In *Proc. WWW*, 2007.

[49] Dinei Florencio, Cormac Herley, and Paul C. van Oorschot. Password portfolios and the finite-effort user: Sustainably managing large numbers of accounts. In *Proc. Usenix Security*, 2014.

[50] Alain Forget, Sonia Chiasson, P. C. van Oorschot, and Robert Biddle. Improving text passwords through persuasion. In *Proc. SOUPS*, 2008.

[51] Edward H. Freeman. Privacy notices under the Gramm-Leach-Bliley Act. *Information Systems Security*, 12(2):5–9, 2003.

[52] Mark Furletti and Stephen Smith. Financial privacy: perspectives from the payment cards industry. *Payment Cards Center Discussion Paper*, 2003.

[53] Loretta Garrison, Manoj Hastak, Jeanne M Hogarth, Susan Kleimann, and Alan S Levy. Designing evidence-based disclosures: A case study of financial privacy notices. *Journal of Consumer Affairs*, 46(2):204–234, 2012.

[54] Shirley Gaw and Edward W. Felten. Password management strategies for online accounts. In *Proc. SOUPS*, 2006.

[55] Dan Goodin. Anatomy of a hack: How crackers ransack passwords like "qeadzcwrsfxv1331". *Ars Technica*, May 2013.

[56] Mark A. Graber, Donna M. D'Alessandro, and Jill Johnson-West. Reading level of privacy policies on Internet health web sites. *Journal of Family Practice*, 2002.

[57] Gramm-Leach-Bliley Act. Pub. L. No. 106-102, 113 Stat. 1338, 1999.

[58] Beate Grawemeyer and Hilary Johnson. Using and managing multiple passwords: A week to a view. *Interacting with Computers*, 23(3), June 2011.

[59] S.M. Taiabul Haque, Matthew Wright, and Shannon Scielzo. A study of user password strategy for multiple accounts. In *Proc. CODASPY*, 2013.

[60] Marian Harbach, Markus Hettig, Susanne Weber, and Matthew Smith. Using personal examples to improve risk communication for security & privacy decisions. In *Proc. CHI*, 2014.

[61] Manoj Hastak and Mary J. Culnan. Online behavioral advertising "icon" study. `http://futureofprivacy.org/final_report.pdf`, January 2010. Unpublished Report.

[62] Chris Hoofnagle, Ashkan Soltani, Nathan Good, Dietrich Wambach, and Mika Ayenson. Behavioral advertising: The offer you cannot refuse. *Harvard Law & Policy Review*, 6(2):273–296, 2012.

[63] Troy Hunt. The science of password selection. Blog Post, July 2011. `http://www.troyhunt.com/2011/07/science-of-password-selection.html`.

[64] Philip Inglesant and M. Angela Sasse. The true cost of unusable password policies: password use in the wild. In *Proc. ACM CHI*, 2010.

[65] InsidePro. PasswordsPro. `http://www.insidepro.com/eng/passwordspro.shtml`, 2003-.

[66] Markus Jakobsson and Mayank Dhiman. The Benefits of Understanding Passwords. *Mobile Authentication*, 2013.

[67] Edward J. Janger and Paul M. Schwartz. Gramm-Leach-Bliley Act, information privacy, and the limits of default rules, the. *Minnesota Law Review*, 86, 2001.

[68] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 471–478, 2004.

[69] Ari Juels and Ronald L. Rivest. Honeywords: Making password-cracking detectable. In *Proc. CCS*, 2013.

[70] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "nutrition label" for privacy. In *Proc. SOUPS*, 2009.

[71] Patrick Gage Kelley, Saranga Komanduri, Michelle L. Mazurek, Richard Shay, Tim Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Julio Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *Proc. IEEE Symp. Security & Privacy*, 2012.

[72] Kleimann Communication Group Inc. Evolution of a prototype financial privacy notice, February 2006.

[73] Kleimann Communication Group Inc. A report on validation testing results, February 2009.

[74] Saranga Komanduri. *Modeling the adversary to evaluate password strengh with limited samples.* PhD thesis, Carnegie Mellon University, 2015.

[75] Saranga Komanduri, Richard Shay, Lorrie Faith Cranor, Cormac Herley, and Stuart Schechter. Telepathwords: Preventing weak passwords by reading users' minds. In *Proc. USENIX Security*, 2014.

[76] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. Of passwords and people: Measuring the effect of password-composition policies. In *Proc. CHI*, 2011.

[77] Saranga Komanduri, Richard Shay, Greg Norcie, Blase Ur, and Lorrie Faith Cranor. AdChoices? Compliance with online behavioral advertising notice and choice requirements. *ISJLP*, 7:603–721, 2012.

[78] KoreLogic. "Crack Me If You Can" - DEFCON 2013. `http://contest-2013.korelogic.com`, 2010-.

[79] KoreLogic. Pathwell topologies. *KoreLogic Blog*, 2014. `https://blog.korelogic.com/blog/2014/04/04/pathwell_topologies`.

[80] Peter Kosmala. Yes, Johnny can benefit from transparency and control. DAA Blog, `http://www.aboutads.info/blog/yes-johnny-can-benefit-transparency-and-control`, October 2011, Retrieved November 2011.

[81] KPMG International. The converged lifestlye. `http://www.kpmg.com/convergence`, 2011.

[82] Balachander Krishnamurthy, Konstantin Naryshkin, and Craig Wills. Privacy leakage vs. protection measures: The growing disconnect. In *Proc. W2SP*, 2011.

[83] Balachander Krishnamurthy and Craig Wills. Privacy diffusion on the web: a longitudinal perspective. *WWW 2009.*

[84] Naveen Kumar. Password in practice: An usability survey. *Journal of Global Research in Computer Science*, 2(5), 2011.

[85] Jeffrey M. Lacker. The economics of financial privacy: to opt out or opt in? *Economic Quarterly-Federal Reserve Bank of Richmond*, 88(3):1–16, 2002.

[86] Pedro G. Leon, Justin Cranshaw, Lorrie Faith Cranor, Jim Graves, Manoj Hastak, Blase Ur, and Guzi Xu. What do online behavioral advertising disclosures communicate to users? In *Proc. WPES*, 2012.

[87] Pedro G. Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Faith Cranor. Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proc. CHI*, 2012.

[88] Pedro G. Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. What matters to users? Factors that affect users' willingness to share information with online advertisers. In *Proc. SOUPS*, 2013.

[89] Pedro Giovanni Leon, Lorrie Faith Cranor, Aleecia M McDonald, and Robert McGuire. Token attempt: The misrepresentation of website privacy policies through the misuse of P3P compact policy tokens. In *Proceedings of the 9th annual ACM Workshop on Privacy in the Electronic Society (WPES)*, pages 93–104, 2010.

[90] Alan Levy and Manoj Hastak. Consumer comprehension of financial privacy notices. Interagency Notice Project, `http://ftc.gov/privacy/privacyinitiatives/Levy-Hastak-Report.pdf`, December 2008.

[91] Jialiu Lin, Shahriyar Amini, Jason Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In *Proc. UbiComp*, 2012.

[92] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I. Hong. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Proc. SOUPS*, 2014.

[93] Jerry Ma, Weining Yang, Min Luo, and Ninghui Li. A study of probabilistic password models. In *Proc. IEEE Symp. Security & Privacy*, 2014.

[94] Jonathan R. Macey. The business of banking: Before and after Gramm-Leach-Bliley. *Journal of Corporation Law*, 25:691, 1999.

[95] Macro International Inc. Mall intercept study of consumer understanding of financial privacy notices: Methodological report. `http://www.ftc.gov/reports/quantitative-research-macro-international-report`, September 2008.

[96] David Malone and Kevin Maher. Investigating the distribution of password choices. In *Proc. WWW*, 2012.

[97] Jonathan R. Mayer and John C. Mitchell. Third-party web tracking: Policy and technology. In *Proc. IEEE S&P*, 2012.

[98] Michelle L. Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Richard Shay, and Blase Ur. Measuring password guessability for an entire university. In *Proc. CCS*, 2013.

[99] Peter L. McCorkell and Andrew M. Smith. Fair Credit Reporting Act update—2008. *Business Lawyer*, 64(2):579–591, 2009.

[100] Scott McCoy, Andrea Everard, Peter Polak, and Dennis Galletta. The effects of online advertising. *Communications of the ACM*, 50(3):84–88, 2007.

[101] Aleecia M. McDonald and Lorrie Faith Cranor. Beliefs and behaviors: Internet users' understanding of behavioral advertising. *TPRC 2010*.

[102] Aleecia M. McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. *ISJLP*, 4:543–897, 2009.

[103] B. Dawn Medlin and Joseph A. Cazier. An Empirical Investigation: Health Care Employee Passwords and Their Crack Times in Relationship to HIPAA Security Standards. *IJHISI*, 2(3), 2007.

[104] Eric Medvet, Engin Kirda, and Christopher Kruegel. Visual-similarity-based phishing detection. In *Proc. SecureComm*, 2008.

[105] Robert Morris and Ken Thompson. Password security: A case history. *CACM*, 22(11), 1979.

[106] NAI. FAQs. `http://www.networkadvertising.org/managing/faqs.asp`. Last accessed June 2012.

[107] NAI. Learn about online behavioral advertising, privacy, cookies, and how this all works! `http://networkadvertising.org/managing/learn_more.asp`, Retrieved November 2011.

[108] NAI. 2008 NAI Principles: The Network Advertising Initiative's Self-Regulatory Code of Conduct. `http://www.networkadvertising.org/networks/2008NAIPrinciplesfinalforWebsite.pdf`, 2008.

[109] Arvind Narayanan and Vitaly Shmatikov. Fast dictionary attacks on passwords using time-space tradeoff. In *Proc. CCS*, 2005.

[110] Gregory T. Nojeim. Financial privacy. *New York Law School Journal of Human Rights*, 17:81, 2000.

[111] Gilbert Notoatmodjo and Clark Thomborson. Passwords and perceptions. In *Proc. ACIS*, 2009.

[112] OCC, Federal Reserve System, FDIC, OTS, NCUA, FTC, CFTC, and SEC. Final model privacy form under the Gramm-Leach-Bliley Act. *Federal Register*, 74:62890–62994, December 1, 2009.

[113] OECD. Guidelines on the protection of privacy and transborder flows of personal data, September 1980.

[114] Alexander Peslyak. John the Ripper. `http://www.openwall.com/john/doc/MODES.shtml`, 1996-.

[115] Alexander Peslyak. John the Ripper. `http://www.openwall.com/john/`, 1996-.

[116] PHDays. "Hash Runner"– Positive Hack Days. http://2013.phdays.com/program/contests/, 2013.

[117] Robert W. Proctor, Mei-Ching Lien, Kim-Phuong L. Vu, E. Eugene Schultz, and Gavriel Salvendy. Improving computer security for authentication of users: Influence of proactive password restrictions. *Behavior Research Methods, Instruments, & Computers*, 34(2):163–169, 2002.

[118] Kristen Purcell, Joanna Brenner, and Lee Rainie. Search engine use 2012. Technical report, March 2012.

[119] Ashwini Rao, Birendra Jha, and Gananand Kini. Effect of grammar on security of long passwords. In *Proc. CODASPY*, 2013.

[120] Rapid7. Linkedin passwords lifted, retrieved September 2012. http://www.rapid7.com/resources/infographics/linkedIn-passwords-lifted.html.

[121] R. Rettie, H. Robinson, and B. Jenner. Does Internet advertising alienate users? *Occasional Paper Series*, (52), 2003.

[122] S. Rodgers. The interactive advertising model tested: The role of Internet motives in ad processing. *Journal of Interactive Advertising*, 2(2):22–33, 2002.

[123] Norman Sadeh, Alessandro Acquisti, Travis D. Breaux, Lorrie Faith Cranor, Aleecia M. McDonald, Joel Reidenberg, Noah A. Smith, Fei Liu, N. Cameron Russell, Florian Schaub, Shomir Wilson, James T. Graves, Pedro Giovanni Leon, Rohan Ramanath, and Ashwini Rao. Poster: Towards usable privacy policies: Semi-automatically extracting data practices from websites' privacy policies. In *Proc. SOUPS Extended Abstracts*, 2014.

[124] G. Salton, A. Wong, and C.S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975.

[125] Julia C. Schiller. Informational privacy v. the commercial speech doctrine: Can the Gramm-Leach-Bliley Act provide adequate privacy protection. *Commlaw Conspectus*, 11:349, 2003.

[126] Richard Shay, Patrick Gage Kelley, Saranga Komanduri, Michelle L. Mazurek, Blase Ur, Timothy Vidas, Lujo Bauer, Nicholas Christin, and Lorrie Faith Cranor. Correct horse battery staple: Exploring the usability of system-assigned passphrases. In *Proc. SOUPS*, 2012.

[127] Richard Shay, Saranga Komanduri, Adam L. Durity, Phillip (Seyoung) Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, Lorrie Faith Cranor, and Nicolas Christin. Can long passwords be secure and usable? In *Proc. CHI*, 2014.

[128] Xinguang Sheng and Lorrie Faith Cranor. An evaluation of the effect of us financial privacy legislation through the analysis of privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, 2:943, 2005.

[129] Bernard Shull. Banking, commerce and competition under the Gramm-Leach-Bliley act. *Antitrust Bulletin*, 47:25, 2002.

[130] Supriya Singh, Anuja Cabraal, Catherine Demosthenous, Gunela Astbrink, and Michele Furlong. Password sharing: implications for security design based on social practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 895–904, New York, NY, USA, 2007. ACM.

[131] Andreas Sotirakopoulos, Ildar Muslukov, Konstantin Beznosov, Cormac Herley, and Serge Egelman. Motivating users to choose better passwords through peer pressure. *SOUPS Poster*, 2011.

[132] Jens Steube. Hashcat. `https://hashcat.net/oclhashcat/`, 2009-.

[133] Jens Steube. Mask Attack. `https://hashcat.net/wiki/doku.php?id=mask_attack`, 2009-.

[134] Jens Steube. Rule-based Attack. `https://hashcat.net/wiki/doku.php?id=rule_based_attack`, 2009-.

[135] Elizabeth Stobert and Robert Biddle. The password life cycle: User behaviour in managing passwords. In *Proc. SOUPS*, 2014.

[136] Peter P. Swire. Efficient confidentiality for privacy, security, and confidential business information. *Brookings-Wharton Papers on Financial Services*, 2003(1):273–310, 2003.

[137] Zhulei Tang, Yu (Jeffrey) Hu, and Michael D. Smith. Gaining trust through online privacy protection: Self-regulation, mandatory standards, or caveat emptor. *Journal of Management Information Systems*, 24(4):153–173, 2008.

[138] Terms of Service; Didn't Read. `http://tosdr.org/`.

[139] TRUSTe. 2011 consumer research results: Privacy and online behavioral advertising. `http://www.truste.com/ad-privacy/TRUSTe-2011-Consumer-Behavioral-Advertising-Survey-Results.pdf`, July 2011.

[140] Trustwave. 2014 business password analysis. *Password Research*, 2014.

[141] Janice Y. Tsai, Serge Egelman, Lorrie F. Cranor, and Alessandro Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2):254–268, June 2011.

[142] Joseph Turow, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy. Americans reject tailored advertising and three activities that enable it. `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214`, 2009.

[143] Blase Ur, Patrick Gage Kelly, Saranga Komanduri, Joel Lee, Michael Maass, Michelle Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. How does your password measure up? The effect of strength meters on password creation. In *Proc. USENIX Security*, August 2012.

[144] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proc. SOUPS*, 2012.

[145] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. "I added '!' at the end to make it secure": Observing password creation in the lab. In *Proc. SOUPS*, 2015.

[146] Blase Ur, Sean M. Segreti, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Saranga Komanduri, Darya Kurilova, Michelle L. Mazurek, William Melicher, and Richard Shay. Measuring real-world accuracies and biases in modeling password guessability. In *Proc. USENIX Security*, 2015 (forthcoming).

[147] Blase Ur, Manya Sleeper, and Lorrie Faith Cranor. {Privacy, Privacidad, Приватност} policies in social media: Providing translated privacy notice. *I/S: A Journal of Law and Policy for the Information Society*, 9(2), 2013.

[148] Ashlee Vance. If your password is 123456, just make it hackme. New York Times, 2010.

[149] Tanzina Vega. Ad group unveils plan to improve web privacy. New York Times, `http://www.nytimes.com/2010/10/04/business/media/04privacy.html`, October 2010, retrieved March 2011.

[150] Rafael Veras, Christopher Collins, and Julie Thorpe. On the semantic patterns of passwords and their security impact. In *Proc. NDSS*, 2014.

[151] Rafael Veras, Julie Thorpe, and Christopher Collins. Visualizing semantics in passwords: The role of dates. In *Proc. VizSec*, 2012.

[152] Emanuel von Zezschwitz, Alexander de Luca, and Heinrich Hussmann. Survival of the Shortest: A Retrospective Analysis of Influencing Factors on Password Composition. In *Proc. INTERACT*. 2013.

[153] Kim-Phuong L. Vu, Robert W. Proctor, Abhilasha Bhargav-Spantzel, Bik-Lam (Belin) Tai, and Joshua Cook. Improving password security and memorability to protect personal and organizational information. *Int. J. of Human-Comp. Studies*, 65(8):744–757, 2007.

[154] Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proc. CCS*, 2010.

[155] Matt Weir, Sudhir Aggarwal, Breno de Medeiros, and Bill Glodek. Password cracking using probabilistic context-free grammars. In *Proc. IEEE Symp. Security & Privacy*, 2009.

[156] Lawrence J White. The Gramm-Leach-Bliley Act of 1999: A bridge too far—or not far enough. *Suffolk University Law Review*, 43:937, 2009.

[157] Craig E. Wills and Can Tatar. Understanding what they do with what they know. In *Proc. WPES*, 2012.

[158] Craig E. Wills and Mihajlo Zeljkovic. A personalized approach to web privacy: awareness, attitudes and actions. *Information Management & Computer Security*, 19(1):53–73, 2011.

[159] Shiva Houshmand Yazdi. Analyzing password strength and efficient password cracking. Master's thesis, The Florida State University, 2011.

[160] Yinqian Zhang, Fabian Monrose, and Michael K Reiter. The security of modern password expiration: An algorithmic framework and empirical analysis. In *Proc. CCS*, 2010.

[161] Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. Password advice shouldn't be boring: Visualizing password guessing attacks. In *Proc. eCRS*, 2013.

[162] Moshe Zviran and William J. Haga. Password security: an empirical study. *J. Mgt. Info. Sys.*, 15(4), 1999.