Prioritizing Security over Usability: Strategies for How People Choose Passwords

Rick Wash[*†]
Media and Information
Michigan State University
East Lansing, MI 48824
USA
wash@msu.edu

Emilee Rader[*]
Media and Information
Michigan State University
East Lansing, MI 48824
USA
emilee@msu.edu

Word Count: 13986 words

[*]The authors wish it to be known that, in their opinion, both authors should be regarded as joint First Authors.

[†]Corresponding Author

Abstract

Passwords are one of the most common security technologies that people use everyday. Choosing a new password is a security decision that can have important consequences for end users. Passwords can be long and complex, which prioritizes the security-focused aspects of a password. They can also be simple—easy to create, remember and use—which prioritizes the usability aspects of the password. The tradeoff between password security versus usability represents competing constraints that shape password creation and use. We examined an ecologically valid dataset of 853 passwords entered a total of 2533 times by 134 users into 1010 websites, to test hypotheses about the impact of these constraints. We found evidence that choices about password complexity reflect an emphasis on security needs, but little support for the hypothesis that users take day-to-day ease of use of the password into account when creating it. There was also little evidence that password creation policies drive password choices.

1 Introduction

User interfaces for password creation and entry are the most common security mechanisms that modern users of technology encounter; they exist on traditional desktop computers, on smartphones and tablets, at ATMs and at payment terminals (in the form of PINs), and on car doors and office doors. Hundreds of thousands of websites and applications require users to enter passwords on a regular basis. Although security experts have been predicting for decades that passwords will soon be replaced by biometrics (e.g., "Passwords could be past tense by 2002" [1]), passwords are still an essential part of computer security, and are the most common form of authentication. Most users have 4-8 different passwords [2] that they enter into 10 or more websites per month. Sasse et al. describe how users feel "authentication fatigue" from entering their password so often [3].

Good passwords have two goals that are very difficult to simultaneously meet: they must be sufficiently complex, unique, and difficult to guess that attackers cannot crack them, even using brute force (the security goal); and they must be sufficiently simple and straightforward that the user can easily remember them and enter them when they need to (the usability goal) [4]. There are wide variety of strategies that end users have identified for achieving these goals. Common strategies include writing down complex passwords [5], re-using the same complex password across multiple different accounts [2], using simple, "throwaway" passwords for accounts that aren't important [6], or using password manager software to store and remember passwords [7].

In most computing systems, end users are empowered to choose their own passwords. User-chosen passwords are usually much easier to remember, and thus users are more accepting of user-chosen passwords [8]. However, this means that the tradeoff between the security and usability requirements of passwords is primarily the responsibility of end users, and they may value these goals differently than administrators or security professionals. In this paper, we use a dataset from 134 end users that includes all of the passwords they entered into websites over a 6 week period to examine patterns in password choices. How do people decide what password to use for each account?

Historically, password research has examined a user's choice of password as an independent decision for each account. Most technical designs for new password systems focus on helping users create a single password for a single account. Research on existing passwords has primarily focused on the password choices that many different individuals make for the same website (due to the use of leaked password datasets) [9]. However, passwords exist in an ecosystem [8]; each user has multiple passwords that they use on multiple accounts [2, 10]. Users have to make choices that not only work for the individual account on which they are creating a password, but that also fit into the larger ecosystem. Remembering a single password is difficult, but remembering 10 or 50 passwords is even more difficult [8]. But, reusing a password can create vulnerabilities across accounts. A better understanding of how users choose passwords and the constraints they face can help technologists create better methods to help users choose better passwords that fit with what they are already doing.

Reviewing the literature, we identify four possible strategies that users might follow to choose passwords that have been identified by prior research: 1. Reusing existing passwords, 2. Focusing on constraints on passwords imposed by websites during password creation, 3. Focusing on the day-to-day usage of the password, 4. And focusing on the security needs of the account.

We logged data from the computers and web browsers of 134 people for six weeks, recorded data about every instance when they entered a password to capture evidence of their password choices, and looked for patterns in the characteristics of their passwords. We conducted a series of tests following the principle of *strong inference* [11], trying to find evidence to falsify our hypotheses [12]. Traditional, weak inference derives hypotheses after seeing the data and tests them against uninformative null hypotheses. Instead, we collected data specifically to test these four preexisting hypotheses, and tested them against each other rather than against uninformative null hypotheses. Our work follows this explicit strategy to try to distinguish between these four hypotheses by examining a single set of passwords and trying to determine which hypothesis (or hypotheses) best explain how those passwords were chosen. This process helps avoid confirmation bias in scientific research [11] that may have been present in prior studies, and compares conflicting prior findings directly against each other to see which strategies are dominant.

We ruled out all but the fourth strategy as the predominant strategy: users commonly take the security needs of the websites they use into account, by choosing passwords that are perceived to be stronger on websites believed to have higher security needs, and passwords that are perceived to be weaker on websites with less need for security. This suggests that users seek a balance between usability and security, but also make distinctions between types of websites that are reflected in their password choices. In other words, most users may be voluntarily adopting a strategy that prioritizes security needs over usability. This means that, generally, the goals of users and of security professionals are not inherently at odds, and that opportunities exist to design systems that support users in their own security goals.

2 Literature Review

2.1 Password Complexity

There has been much research analyzing how users choose what password to use on a system. Passwords are often evaluated on their *complexity*: how many characters are in the password (its length), and how many different types of characters (letters? numbers? symbols?) does the password include. Much of the security advice that end users receive about passwords is focused on its complexity; users regularly hear that their passwords should be at least N characters long (though possibly with different N's), and that passwords should include symbols or numbers [7].

Researchers have measured password complexity in a number of different ways. Frequently, password complexity is measured by calling upon the concept of Shannon En-

tropy [5]. Shannon measured entropy, or the amount of information, by taking the logarithm base 2 of the options weighted by their likelihood. This is a theoretical concept that relies on the idea that a password is chosen from some set of possible passwords, and thus Shannon's concept mostly applies to the set rather than an individual password [13].

The U.S. National Institute for Standards and Technology (NIST) used Shannon's concept of entropy to come up with multiple ways of measuring the complexity of a password [14]. One measure, which they called "random password entropy", has since become a common measurement. This measure assumes that all passwords with similar characters are equally likely, and thus measures complexity with the logarithm of the number of possible passwords. This measure is often called "Shannon entropy" in the security literature, though Bonneau describes this use of the term as "imprecise" [5].

Many researchers use character classes as a way of measuring random password entropy. For example, if a password is entirely lowercase English letters, then each character must be one of 26 options, and if uppercase letters are included then it must be one of 52 options. The focus on character classes can be traced back to Microsoft Windows NT Service Pack 2, which was one of the first systems to enforce a password policy based on character classes [13]. Passwords that include characters from multiple classes have a larger set of options for each character, and longer passwords have more characters. In a supplemental guide, NIST tried to estimate the random password entropy for passwords chosen from different character classes [15], though Shay et al. question the accuracy of those estimates [16].

Guessability Much of the research looking at password choice examines datasets of passwords that a large number of users have chosen on a single system. This work often comes from cracking a password database that researchers have legitimate access to [17] or analyzing a leaked dataset of passwords from the hack of a popular web service [9]. This research repeats a common theme: a large number of users choose the same, obvious passwords as each other [9]. This insight then led researchers to use these patterns in password popularity to improve password guessing attacks by guessing more popular passwords first. Bonneau formalized a new measure of password security – guessability – that measures how difficult it is to brute-force guess a password, guessing more popular passwords first [9, 18]. NIST refers to this measure as "guessability entropy" [14], acknowledging that this is another measure inspired by Shannon's concept of entropy. This measure depends heavily on the database of passwords used to order the guesses; however, most common databases in use today have roughly similar guessability scores [19].

There has been much debate in the password research community about the difference between password complexity (measured by random password entropy), and password guessability (usually measured using brute-force search or an approximation to such a search [20]). The consensus is emerging that password guessability is a much better measurement of the actual, real-world security of a password. However, most end users do not have a good method of determining guessability, because they do not have visibility into the set of passwords other people have chosen. Instead, end user mental models of pass-

words are often focused on characteristics of passwords such as how long they are (number of characters), whether they include numbers, letters, and special characters, and whether they include common words. That is, end user mental models of passwords are usually more focused on the characteristics of passwords that are part of complexity measurements like random password entropy than they are on guessability [21, 22].

2.2 Choosing Passwords

There is little direct research about how users choose passwords in real-world settings. Much of the existing research comes from users self-reporting strategies for choosing passwords. One consistent finding is that users do not seem to use the same strategy for all passwords, but instead choose different passwords for different accounts.

One strategy that has been reported by users is to create "stronger" passwords on websites that have more "sensitive" content [23, 24]. Notoamojo and Thormobson found that 70% of users reported having at least one password reserved for high importance websites [24]. They also reported that users believe passwords that are difficult to recall are more secure. Haque et al. suggested that users might treat different categories of websites differently when choosing passwords [25]. Their diary study found that users commonly used words to classify websites as "financial", "sketchy", or "content" as ways of distinguishing different types of websites. A very similar strategy emerged from a series of diary studies of password users by Duggen et al. In these studies, users chose weaker, more memorable passwords on non-sensitive sites because the information on the site wasn't important to them [26]. In all of these studies, users reported choosing different passwords for different accounts, and intentionally and thoughtfully making these decisions.

Steves et al. conducted a diary study of password use by US Government employees [27]. They found that users use passwords for a wide variety of authentication purposes, including access to email (both work and personal), access to specific software systems, physical access to buildings, access to devices like mobile phones and wifi networks, and to accomplish goals like encryption and making purchases. They report that users described "authentication fatigue": that they had to authenticate too often to too many different places, and that remembering all of those passwords was very difficult. Most users utilized multiple memory aids to help them remember all of their passwords.

Stobert and Biddle described this decision-making process as a "lifecycle" process [8]. Using both interview and survey data, they described how passwords are initially chosen, lived with for a time, and then changed to accommodate a variety of different influences on password choices, including day-to-day use and security concerns. Both experts and non-experts reported taking both usability and security into account when choosing or updating passwords on accounts, devoting more attention toward accounts that they felt were more important [8].

2.3 Password Strategies

In order to better understand and analyze passwords, we make a distinction between *password choice* and *password strategy*. To do so, we borrow the concept of strategy from the field of Game Theory [28].

Game theory puts a formal, mathematical structure on situations with uncertain outcomes (which it calls "games"), and makes an important conceptual distinction between a *strategy* and an *choice*. In a given situation, the choice that a person makes can be called that person's action. Game theorists use different words to describe this, include "choice", "action", "move", and (confusingly) "pure strategy". When observing someone in that situation, it is usually possile observe the choice that they end up making.

A strategy, though, is different; it is a higher level "plan of action" [29]. Strategies are, roughly, *how* a person goes about choosing which action to take in a situation. Strategies can include randomization (randomly choosing among possible actions, a so-called "mixed strategy"); they can include contingency plans for what to do after learning more information; they also can include reasoning for why the plan is a good idea, which can help deal with unexpected situations. Rubenstein discusses the complexities of what a strategy is, and quotes Shubik in defining a strategy to be "a complete description of how a player intends to play a game, from beginning to end" [29].

For the case of passwords, the actual password used is evidence of the choice that the user made in that specific situation. However, as Stobert and Biddle argue [8], when users need multiple passwords for multiple different purposes (websites, apps, etc.), users do not choose those passwords independently. Instead, they form higher-level plans (strategies) to help them manage their "ecosystem" of passwords. According to this definition, then, we define *password strategy* to be a guideline or plan for choosing multiple passwords across a range of different websites, apps, and services.

Strategies are difficult to empirically observe; they often include unobservable plans for situations that do not actually occur (such as random choices by participants or contingency plans) [30]. However, understanding the underlying strategies is critical when designing for future situations. Axelrod conducted a number of simulations of the commonly-studied Prisoner's Dilemma game and showed that, even when the observed past actions are similar, if you change the rules of the game (e.g. real-world policies), then players react differently to the new rules based on their strategies, and the outcomes depend more on the strategies used than on the past actions [31].

Most of the past research that empirically examines passwords has focused on choices – which passwords the users actually chose. In this paper, we focus on trying to understand the *strategies* that users employ to choose those passwords. Following Axelrod's example, we believe that in order to design new password systems and password policies, it is more important to understand user strategies, so we can better estimate how users will react in new situations.

Empirically examining strategies is difficult. As we mentioned above, we can often directly observe actions that people take, but we almost never can observe the strategies that

they used to choose those actions. Our approach follows Popper's idea of falsification [12]: rather than trying to measure which strategy users are following, we instead identify a number of potential strategies (hypotheses) and then collect data that potentially allows us to falsify those hypotheses, showing that it cannot be the strategy that users follow.

One example from game theory involves the game *Matching Pennies*, which is surprisingly close to the situation faced by soccer (football) players doing penalty kicks. Game theory predicts that the only equilibrium in Matching Pennies involves the use of a randomized strategy. Chiaporri et al. collected empirical evidence from penalty kicks in real games and looked for evidence that players were not using the predicted randomized strategy [30]. They were unable to rule out this strategy, thus concluding that the randomized strategy is a reasonable description of how players make choices in penalty kick situations. We seek to do a similar task with passwords; we examine patterns in actual password choices, compare those patterns to hypotheses about password strategies, and then (hopefully) rule out some hypotheses as inconsistent with the data about password choice.

2.4 Hypotheses About Password Choice

Passwords allow the end user to make different security/usability tradeoffs for different accounts. Users can choose different passwords that are either more secure or more usable for different websites, depending on properties of those websites. Summarizing the existing literature, we posit four high-level classes of strategies that people can use for choosing passwords that represent different ways of making the security/usability tradeoff. Each class of strategies focuses on a different constraint that users face when choosing passwords. We then pose hypotheses that each class is commonly followed by users, and examine data that can help differentiate which of these strategies are most commonly being followed by users. Due to limitations in the data, we do not distinguish between between patterns across different users of a website, and longitudinal patterns of a single user's choices across different sites in our hypotheses.

2.4.1 Reuse Focused Password Strategy

To begin, we start with the simplest possible strategy: always choose the same password for every website. Rather than choosing a (potentially) different password for each website a user encounters, users can simply reuse the same or a substantially similar password as previously used on other websites. Password re-use is a very common strategy for many end users [2, 8, 32]. However, reusing passwords across websites poses an important security risk. If an attacker learns a password for an account one website, he or she can then also use that same password to log into similar accounts at all places where that password was reused [33]. Therefore, this practice by users creates interdependencies between websites [34].

However, password reuse is an important strategy for improving the usability of passwords in general across the password ecosystem [8]. By reusing passwords, users have to

memorize fewer passwords, get more regular reminders of what the password is (because they have to enter it in more frequently), and can log in from anywhere even if they don't have access to their written down passwords (because reused passwords are more likely to be memorized). Von Zeschwitz et al. [32] found that over 50% of their interviewees reported reusing passwords, and they claimed this was because it would be too hard to remember passwords if they did not.

Wash et al. [2] found that users frequently reuse passwords. Among their college student sample, they found their users' most-reused passwords were reused across an average of 9 different websites. Pearman et al. similarly found that their more diverse, non-student sample reused approximately 80% of their passwords [10]. Participants in Inglesant and Sasse's diary study reported that good passwords are a "resource" to return to when creating new accounts [35].

Reusing passwords is not entirely straightforward, however. Most websites have different policies about the minimum requirements for a password [36]. These policies require different features of passwords; for example, some websites may require the use of special characters, and other websites do not allow them. Some websites have a minimum password length, which may be above another website's maximum password length. Pearman et al. speculate that stronger passwords may be easier to reuse because they satisfy the policy requirements of a larger number of websites [10].

Hypothesis 1 *(Reuse Focused) Users primarily choose a single complex password that meets most security requirements, and reuse that password across as many websites as will allow it.*

A number of researchers have observed that users often have passwords that are slight variants of each other; one password may replace a letter with a symbol or add a number at the end of the password [10]. That is, users often *partially* reuse their passwords. Prior research has suggested at least two reasons how these variants might arise: 1) users want to reuse a password from a different website, but that password does not meet the new website's requirements, so they make a minor modification so that it does [16]; or 2) users are forced to change their password after a certain amount of time, and to make it easier to remember the new password they simply make a minor modification to their old password [37]. This can lead to different variants in use on different websites if the original password was reused.

Hypothesis 1 only covers *exact* reuse of a password. When a person uses variants, they (by definition) have more than one variant to choose from. Using variants still leaves the strategic question open about which websites receive variants, and if so, which variant should be used? That is, there is still a security/usability tradeoff involved in choosing which variant to use. We do not explicitly have separate hypotheses about variants; all of our hypotheses are valid hypotheses about which variant gets chosen. We avoid trying to classify passwords as similar enough to be a "variant", or different enough to be "unique"; instead we focus on the choice that the users make about which to use.

Next, we posit three hypotheses about how users make password choices that lead to the use of different passwords (or variants) on different websites.

2.4.2 Creation Focused Password Strategies

If users do not simply reuse the same password across all websites, then how do they choose which password (or password variant) to use on which website? One possible strategy to choose is responding to the most obvious constraint: the website's policy for what passwords need to look like [37]. When users do this, they focus their password choices on the act of creating the password and the usability concerns that arise during creation.

Hypothesis 2 *(Creation Focused) Users primarily choose passwords by focusing on ease of creating a password.*

We already listed one strategy that users can employ that makes both password creation and password use easier: reuse existing passwords across sites. For Hypothesis 2, we suggest that users may choose to use different passwords on different websites, but that the choice of which password to use on which website is driven mostly by concerns at the time of creation rather than use. For example, if a user wants to reuse a password but that password is not allowed by the website's policy, then the user may create a variant that meets the policy [10].

This hypothesis, however, is too high-level, and is not a detailed, specific strategy. It represents a class of strategies. One concrete way that users can accomplish this is to take this to an extreme: always choose the simplest password that they can. This is equivalent to using creation usability as the sole criterion for password choice, and completely ignoring security needs. We do not believe that this is a realistic password choice strategy; however, we will analyze this strategy by comparing passwords to the minimum required password to determine if the data supports it as a commonly employed strategy. Also note that this strategy is still incomplete; even if users want to choose the simplest allowed password, that does not help users decide which password among the simplest allowed should be chosen.

Hypothesis 2.1 *Users primarily choose passwords by choosing the simplest password allowed by the website.*

Other than choosing the simplest possible password, there are other ways that password policies can influence password choices. Users may, for example, use the policy as an indicator and choose more complex passwords for websites that have more complex policies. We will not analyze these strategies individually, but we will look for broad evidence that policies are affecting password choices.

Hypothesis 2.2 *User choice of passwords is influenced by the password policy of the website, with more complex passwords used on websites with policies that require more complex passwords.*

Florencio et al. analyzed password policies from 75 websites and found that websites that rely on voluntary use or rely on ads as part of their business model tended to have lower password requirements [36]. They found little relationship between website-related security concerns and password policies. If users are focusing on meeting password policy constraints when creating passwords, then they are not using more complex passwords on websites with greater security concerns, because security concerns and password creation policies are not necessarily related.

2.4.3 Usage Focused Password Strategies

If users choose different passwords for each website, then a reasonable strategy would be for users to choose weaker passwords for accounts where that password has to be entered frequently. Having a complex 30 character password that you enter once a year is fine, but having to enter it multiple times a day to unlock your computer is extremely burdensome. The more frequently the password needs to be entered and used, the simpler the password should be. We call this strategy the *usage focused* strategy because the primary concern of users is day-to-day use of the password. In situations where usability really matters (frequent entry, or mobile keyboard entry), usability is the primary concern of end users, who choose simple passwords in these situations.

Hypothesis 3 *(Usage Focused) Users primarily choose passwords by making a security/usability tradeoff and focusing on the usability needs of using the website*

In focusing on usage, there are two possible types of uses that users can focus on. First, the most logical use would be for users to focus on entering the password; the more often they have to enter the password, the simpler the password should be to make it easier to enter:

Hypothesis 3.1 *Users choose different passwords for websites, and primarily choose passwords by focusing on how frequently they enter the password into the website with more frequent password entry leading to simpler passwords.*

However, password entry is a relatively infrequent activity. Many websites have "remember me" style functionality that enables users to enter a password once and remain logged in for days or weeks. This may make it difficult for users to think about how frequently they enter passwords, and instead may lead them to focus on how frequently they visit and use a site. Sites that are visited more often may get simpler passwords.

Hypothesis 3.2 *Users choose different passwords for websites, and primarily choose passwords by focusing on how frequently they visit the website, with more frequent visits leading to simpler passwords.*

2.4.4 Security Focused Password Strategy

Interviews with users about password creation have suggested that users choose more complex passwords for accounts that are very important to them, and choose simple, easy-to-remember "throwaway" passwords for accounts at transient websites. Notoamojo and Thormobson, for example, report that end users describe using this strategy for choosing passwords [24]. In addition, Hanamsagar et al. [38] found that participants' passwords for websites they rated as important were longer and less guessable than their passwords for non-important websites. Importance in their study was correlated with the perception that the participant would experience negative consequences if a stranger were to gain access to their account.

We call this strategy the *security focused* strategy because the primary concern in this strategy is how important the account is, and therefore the security needs of the account. Accounts that need greater security get more complex passwords, and users can choose simpler passwords for accounts that don't need strong security.

Hypothesis 4 *(Security Focused) Users primarily choose passwords by making a security/usability tradeoff and focusing on security needs of the website.*

Prior research has suggested at least two different ways that end users can evaluate the security needs of a website. Haque et al. suggest that users might look at the kind of website – financial website, social media, "sketchy" websites, etc. – and use those logical categories as a way of determining whether to use a complex password or a simple password [25].

Hypothesis 4.1 *Users choose different passwords for websites, and primarily choose those passwords by focusing on the security needs of the category of website.*

Another option that arises from users' self-reported password choice strategies: users consider how "important" the website is to them, and choose stronger passwords for websites that are considered more important [24].

Hypothesis 4.2 *Users choose different passwords for websites, and primarily choose those passwords by focusing on whether they consider the website to be important to them in some way.*

Both of these previous classes of strategies (*security-focused*, and *usage-focused*) recognize the need for making a security / usability tradeoff where some passwords are more secure and others are more usable; they differ in exactly how this tradeoff is made. These two strategies are not entirely mutually exclusive, and they make similar predictions in many situations. For example, both strategies would predict that users would chose a complex password for TurboTax, an income tax preparation software that is used only once a year (low usability needs) but contains large amounts of important financial data (high security needs).

However, for many accounts the predictions of these strategies diverge. The most common divergent case is an organizational single sign-on account, such as an employer or

school login. These accounts are often very important to users because they provide access to a large number of different systems and information (high security needs); however, they frequently need to be entered multiple times a day, and sometimes on a variety of different devices or computers (high usability needs). In this situation, these two hypotheses would make opposite predictions about how users would choose passwords. Users following a *security-focused* strategy would choose complex passwords, prioritizing the security needs of the website, while users following a *usage-focused* strategy would choose less complex passwords to prioritize the day-to-day usability of the website.

3 Methods

Testing hypotheses like these is not straightforward. While we can directly observe the individual password *choices* made by the users, we cannot observe the higher-level *strategies* that led to these choices. The four hypotheses are all hypotheses about these higher-level strategies. As Popper argues, it is impossible to prove a hypothesis to be true. However, it is possible to explicitly look for evidence that could prove the hypothesis to be false [12]. We take that approach here; we explicitly look for evidence about password choices from real users that has the potential to falsify each of our hypotheses. For example, while we cannot prove that people are intentionally trying to reuse a single password everywhere, if we find that they are using different passwords on different websites, then we can declare the hypothesis of a single reused password to be false. This approach has been used in the past to test hypotheses about game theoretic strategies in real world settings where only individual choices can be observed (e.g. [30]).

In order to empirically study password *strategies*, we needed data with two properties not commonly found in leaked password data. First, since strategies are normally enacted by an individual person, we needed a relatively comprehensive set of passwords chosen by that person. We can then look at patterns in each person's password choice and use those patterns to rule out possible strategies that were hypothesized above. Second, we needed additional information that users might incorporate into their strategies, such as password creation policies or information about the websites those passwords are used on. This allows us to examine strategies that involve intentionally choosing different passwords on different websites, which is commonly reported in past research.

Platt argues that while analyzing individual hypotheses is reasonable and scientifically valid, science progresses more quickly when we analyze sets of competing, related hypotheses. Rather than separately testing hypotheses against uninformative null hypotheses, he argues for the principle of strong inference [11]: Most scientific hypotheses make a number of similar predictions, but do not make the same predictions in all situations. The best place to look for evidence to distinguish between competing hypotheses, then, is to design "critical experiments" that create situations where the hypotheses make easy-to-distinguish, different predictions. This process helps avoid confirmation bias in scientific research.

In this paper, we are analyzing data from real-world password use; instead of creating "critical experiments", we instead look for data representing "critical situations": situations where our hypotheses make different, competing predictions. For example, we explicitly look at data about what type of website a password is used on because some types of website are exactly these critical situations where our hypotheses make different, competing predictions.

Additionally, we are not seeking to identify any individual person's strategy. Instead, our hypotheses are about common strategies. If a given strategy (or class of strategies) is common, then certain patterns should logically appear in the data, and other patterns should not appear in the data. We look for those patterns to identify which strategies are commonly used.

3.1 Data

Our primary dataset of passwords comes from a study conducted in the Spring of 2015. We invited a sample of students at a large midwestern university in the USA to participate in a research study about computer security. Students from Computer Science and Engineering were not eligible to participate. We first asked participants to fill out a survey about attitudes and intentions for computer security. Results from this survey are reported elsewhere and are not used in this paper. Second, we asked participants to install a custom software application that collected data from their computers. This application consisted of a Windows service that collected system logs on a regular basis, and a browser plugin (that works on both Mozilla Firefox and Google Chrome browsers) that collected data about the participant's web browsing. Participants were asked to leave this application running for six weeks, and were compensated US$10 per week via Amazon.com gift card. Finally, at the end of the six weeks, participants were asked to fill out another survey (also reported elsewhere).

134 participants completed the study and provided valid web browsing data. Our sample is fairly representative of the population of the university, excluding Computer Science and Engineering students. Almost all participants were in the 18-29 age range. Close to the demographics of the student population, our sample was 53% female and 77% white. Approximately 73% of the participants were undergraduates, while the remaining were graduate students. Only 4 of the 134 participants had children. Table 1 has more details, and more information about the sample can be found in a prior paper that analyzed this data [2].

The browser plugin watched all webpages for instances where users entered a password. It primarily looked for the "password" form element, although through testing we identified that this does not capture all passwords and we added a number of special cases to catch a larger number of password entries. When the plugin identified a password, it computed some statistics about the composition of the password, and then sent a hash of the password along with the computed statistics back to our server. This allowed us to compare the hashed passwords against each other and identify instances where the exact same

Demographic	#	%
Man	61	46%
Woman	71	53%
18–29 years old	127	95%
30–49 years old	7	5%
High School Diploma / Undergraduate student	98	73%
Bachelors degree / Graduate student	36	27%
Have children	4	3%
No children	130	97%
White	103	77%
Asian	13	10%
African American	4	3%
Hispanic	6	5%

Table 1: Demographics of our main sample, whose password choices we are analyzing

password was used by the same user on multiple different websites. We never collected the raw passwords, for privacy reasons. The password characteristics we measured included complexity, and a check for whether the password appeared on a list of common passwords. From this data, we were able to identify each time a user entered a password into a website, what website that password was entered into, and some basic summary statistics about that password. Following Wash and Rader [2], we analyzed the password entries to separate out incorrect passwords from correct passwords, and identified a "likely correct" password for each participant on each website where they entered a password.

To measure password complexity, the browser plugin looked at the password that was entered and decided how many different character classes were represented from the following classes: lowercase letters, uppercase letters, numbers, symbols, extended symbols. Each class represents a number of possible options for that character (26 letters, for example). Our password complexity measure was the logarithm (base 2) of the total number of possible options in the represented character classes raised to the number of characters in the password. This measure approximates how past research on passwords has measured "random password entropy" [14], sometimes imprecisely called "Shannon entropy" [5]. This measure was reported to us by the browser plugin, but the original password was not.

As described above, there are many ways to measure the complexity of a password. We chose this measure because it more closely aligns with user beliefs about password complexity, which we provide evidence for below. This measure is based mostly on character classes and length, which are commonly believed to result in more complex passwords [16, 22]. While past evidence suggests that this type of "random password entropy" does not measure real-world resistance to password guessing as well as guessability measures [9], we believe it is a better approximation of user perceptions of the complexity of a

password.

To be able to test strategies that include choosing different passwords for the websites that users were entering passwords into, we conducted three additional data collections. These three datasets about websites are available in the supplemental materials online: `https://osf.io/a28q9/`

First, we identified the minimum password requirements for each website. In the Spring of 2017, we manually visited all of the websites that at least two participants had entered passwords into, to identify the minimum password requirements for each website. We did this in two ways: 1) we tried to create a number of different passwords on each website in order to determine which passwords were acceptable and which were prohibited; and 2) we used the Google search engine and browsed around the website to look for a written password policy or set of minimum requirements. From these requirements, were were able to identify the minimum complexity password for each website. To account for the fact that this data collection happened approximately two years after the original data collection, we also used the Internet Archive Wayback Machine [39] to look up written password policies for these websites from the time period of the original data collection.

Second, we grouped websites into a set of conceptual categories. To do this, we used the Webshrinker online categorization API [40]. This API has a set of approximately 39 categories that are assigned mostly based on a proprietary machine learning algorithm. We used this system to categorize every domain name that was ever visited by participants in the original data collection.

Third, we wanted to know how "important" each website was to the user who entered their password. We conducted a study in February 2017 that surveyed a new sample using the same sampling frame as the initial data collection: a random sample of undergraduate students in the same large midwestern university excluding Computer Science and Engineering students. In this survey, we presented the participants with 10 randomly selected domains from the set of domains that at least two original participants had entered passwords into. For each domain, we asked a series of Likert-scale questions about the importance of that website to them. While this does not allow us to know how important the website was to the user who chose the password, it does allow us to know whether that website is important in general for members of the same population (students at the university).

This approach is similar to research that uses third-party raters to evaluate texts online (e.g., Twitter posts) for subjective perceptions of aspects of the text. The ratings are then used as ground truth in training a model. Here, we use the website importance ratings not as a proxy for what each individual participant who entered a password on a site might have thought about the website, but rather as an aggregate evaluation of baseline website importance in the same population. Therefore, we are not arguing a direct causal relationship between password choices and website importance. Rather, we're examining importance as a characteristic of websites that may be correlated with password complexity.

Finally, in the Fall of 2018, we conducted a survey in which we measured the relationship between how people perceive the security of a password and the the complexity of

the password, which we report below in Section 4. We recruited participants both from Amazon Mechanical Turk (MTurk), from the Qualtrics panel service, and combined these surveys together into one dataset. We did this because MTurk participants can be more tech-savvy than the general population, and we wanted a more diverse sample on that characteristic. Overall, our sampling frame was similar to that of the previously described studies: regular computer users who did not have specific technical or computer security training. However, the MTurk and Qualtrics participants were slightly older than the university students; the modal age group was 30-49 years old.

Following McShane et al. [41], primarily focus on effect direction and size when interpreting statistical results. We present results of null hypothesis tests and confidence intervals when available, but treat them as secondary indicators.

3.2 Ethical Considerations

Collecting both website visit information and password information from participants is highly private information, that could cause them harm if our research data were to become compromised in some way. We worked to protect participants' privacy during and after this study. At any time, participants in the passwords study could pause data collection using a control panel we supplied with the data collection tool. Additionally, we did not collect any data from the web browsers while in incognito mode (Chrome) or private browsing mode (Firefox), and we informed participants of this and provided instructions on how to use these features. All participants provided informed consent to participate in the study, and we pre-tested the consent form to try to ensure that participants understood what data was being collected about them. All participants in the original study received $10 per week of data collection, plus an additional $10 for filling out surveys, for a total compensation of $70 for participation in the study.

All studies reported here were approved by our institution's IRB.

4 Perceptions of Password Security

Many of the strategies hypothesized above involve the user's perception of a password's security. To test these, we must identify a dependent variable that allows us to measure users' perceptions of how secure a password is. We propose that password complexity, as measured by random password entropy, can be used as a proxy measure for user perceptions of password security. In this section, we provide evidence that password complexity does correlate with user perceptions of password security.

4.1 Perceptions of Password Security

Previous work has shown that user mental models are often focused on the characteristics of passwords, such as length and diversity of characters, that also contribute to password complexity measures. For example, in a study conducted in 2015, Ur et al. [21]

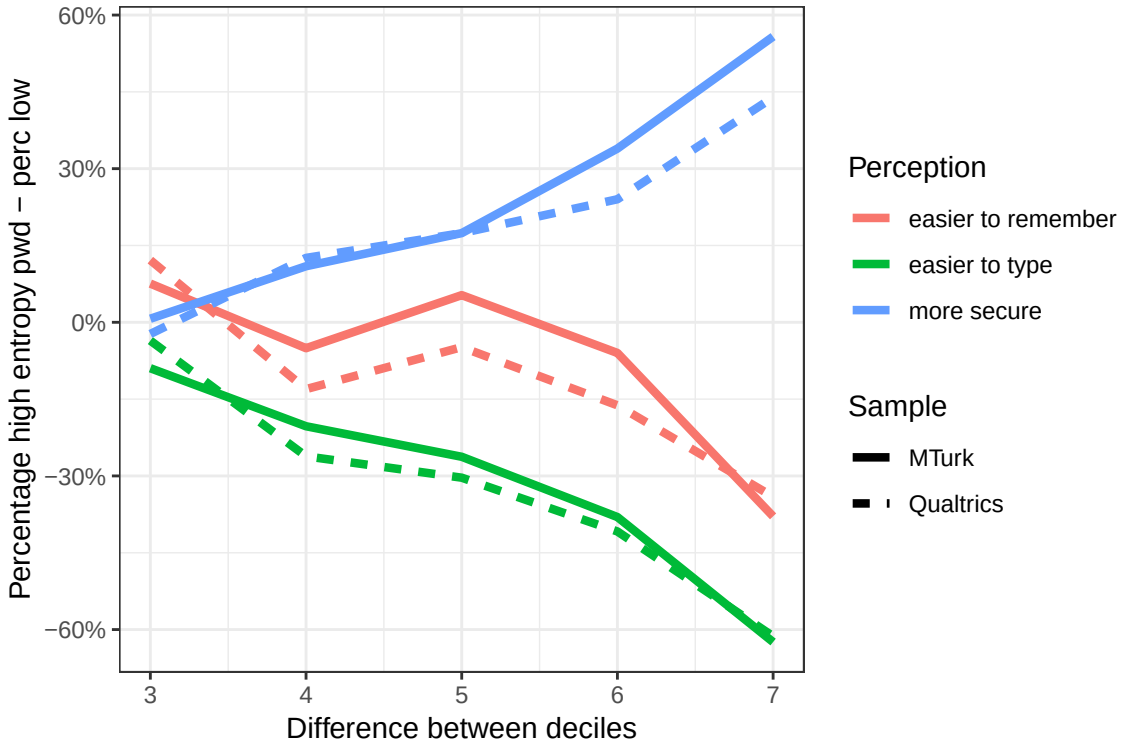

Figure 1: Difference in password entropy (x-asis) by user interpretation of that difference (y-axis). The solid line is the MTurk sample; the dotted line is the Qualtrics sample.

showed participants carefully controlled side-by-side comparisons of passwords and asked questions about their perceptions of the password. While they did not directly compare password complexity and guessability, they found that participants largely believed that adding character sets increased security, that longer passwords were better, and that users underestimated the ease with which common words and sequences of characters could be guessed [21]. Those patterns are not the same as password complexity, but are closer to complexity than to guessability, and generally reflect a focus on character sets plus length for users' mental models of password security.

The study conducted by Ur et al. used carefully selected password pairs that only varied along a single dimension (such as adding a character or replacing one character with another, similar character). We wanted to see how well random password entropy (i.e., what we are calling 'password complexity') held up as a measure of password security perceptions across passwords more generally. We took the 16 million passwords in the RockYou dataset, calculated the complexity of each password, and then grouped them into 10 deciles. We ignored the top and bottom deciles, and randomly picked pairs of passwords from the 8 middle deciles. We showed these pairs to 200 people from a Qualtrics panel and to 100 people from Amazon Mechanical Turk and asked them to choose which password was "more secure". (See the Methods section for more information about the participants.)

The blue line on Figure 1 shows the results. The y-axis is the difference between the percentage of people choosing the higher complexity password and the percentage choosing the lower entropy password. (They also had the option of choosing "both are equal".) Users in both samples believed that higher entropy passwords were more secure: 46% of the time participants chose the higher entropy password, and 23% for the time they said "both were equal". Larger differences in entropy led to a larger percentage of people choosing the higher entropy password. While no single metric captures the full range of people's mental models of password security, these results show that random password entropy—password complexity—is a reasonable proxy measure for user *perceptions* of password security.

4.2 Perceptions of Password Usability

An additional question is whether random password entropy is also a reasonable proxy measure for password usability. In considering this question, it is valuable to distinguish between many different ways a password may be considered to be a "usable" one. Tamborello and Greene describe two types of usability errors for passwords: "motor" errors that occur when typing / entering passwords, and "memory" errors that occur when misremembering passwords [42]. Following this distinction, in this paper we separate usability of passwords into two major categories: how easy a password is to remember, and how easy a password is to enter into a device when needed.

A report from NIST examined the difficulty of entering passwords, and developed a GOMS model for password entry [27]. Some of the factors that played the largest role in determining difficulty of entry were the number of characters (aka password length), and the variety of characters (aka character sets). These are exactly the features of a pass-

word that are part of the random password entropy / password complexity measure. On a day-to-day basis, the NIST report estimated that entering passwords takes more time than remembering passwords [27]. Furthermore, on mobile devices, using characters from multiple character sets often requires extra clicks or taps to change keyboards, so more character sets can make password entry extra difficult on mobile. For these reasons, we suspect that password complexity is also a reasonable metric for the perceived usability of entering a password. However, we suspect that password complexity is probably not a good measure for the usability of remembering a password. Many passwords can be measured to be highly complex but at the same time be easy to remember, or vice versa. So we suspect that password complexity is not correlated with password memorability.

In the same study where we measured perceptions of password security, we also asked participants to choose which of each pair of passwords they thought to be "easier to remember" and "easier to type". The results of these ratings are also in Figure 1. Participants perceived the password with lower complexity to be easier to type more than 53% of the time, and chose the higher complexity password as easier to type only 27% of the time. Therefore, we conclude that password complexity does appear to be inversely related to perceptions of how easy a password is to enter. However, we find little relationship between password complexity and password memorability. 43% of the time participants chose the low complexity password as easier to remember, and 38% of the time they chose the high entropy complexity. Only for very large differences in complexity did there appear to be any differences in memorability.

In this section we have shown evidence that random password entropy—password complexity—is a representation of the characteristics of a password that is also conceptually related to people's perceptions of password security. Indeed, for many years the security research community used random password entropy as a measure of security, and much of the expert security advice about password characteristics intended for end users is given in terms of the same characteristics that are used to calculate random password entropy. We found that there is an empirical correlation between the complexity of a password and how people perceive the security of that password. Password complexity is also related to one aspect of usability: higher complexity passwords are seen as more difficult to enter, but not necessarily more difficult to remember.

5 Results: Which Password Strategy?

Now we proceed to test our primary hypotheses: what strategies are commonly used by users when choosing which password to use on a given website?

For each password that was entered by a participant in the original study, our logging software calculated that password's random password entropy—a measure of how complex the password is. Password complexity is one of the few aspects of authentication systems that end users get to choose. End users can choose more complex passwords if they want to emphasize security. Or they can choose less complex passwords as a way of increasing

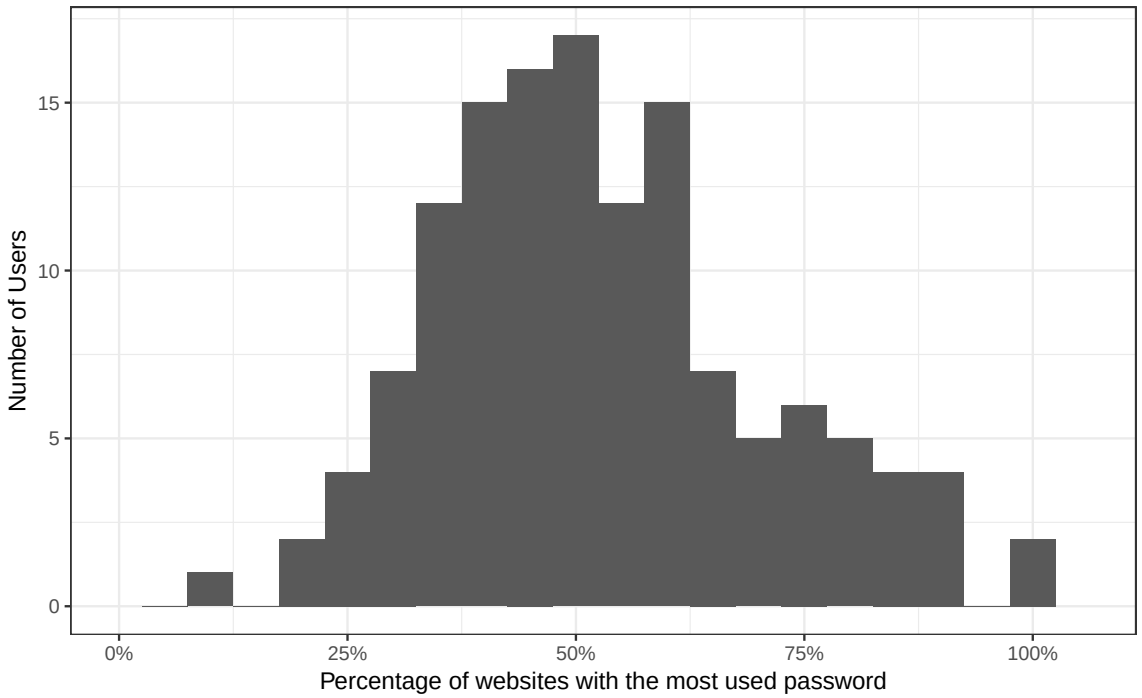

Figure 2: Histogram of often the most-common password is re-used.

usability. This choice belongs to the end user, and it is exactly this choice that we focus on in this paper. Section 4, above, provides evidence that greater password complexity is related to perceptions of password security in users' minds. We use the complexity of the passwords participants entered during our study as evidence of the strategies they used when originally creating their passwords.

5.1 Hypothesis 1: Focus on Password Reuse

We begin by examining the simplest strategy for choosing a password: always use the same password across all websites. This is a strawman hypothesis, because it is very extreme: users will only have one password that they use everywhere. However, at least one user in Sasse et al.'s diary study reported doing exactly this [3]. Once a user has more than one password that they use on different websites, then that raises the strategic questions: when does the user want to reuse a password? And when they do, which password do they reuse?

Out of the 134 users in this study, only two have exactly one password that they reuse everywhere (1.5% of users). These two users reused their single password on 10 and 18 different websites during the six weeks of the study. These two users seem to be following this strategy of exact password reuse. Thus, we conclude that while two users may be following this strategy, it is not widely used as a strategy for making a security/usability tradeoff.

The remaining 132 users all had more than one password that they had used, with different websites having different passwords. Most users had 4-8 passwords [2]. For these users, we need to look at how they choose which password to use on which website. In the past, security researchers have speculated that users try to reuse one password everywhere but cannot due to password policies, so they choose a slight variant that meets the policy in cases where the single reused password is insufficient [10].

To look for evidence for this idea of having a single, dominant password, we identified for each user the most frequently reused password (breaking ties randomly). We then looked at, out of the websites where the user entered a password, how many websites used this exact password. Figure 2 shows a histogram of the percentage of websites covered by this most frequently reused password. A large number of users do not have a single dominant password, with approximately half of our users (40%, 54 out of 134) using their most dominant password on fewer than 50% of the websites they have passwords on, and only 19 users (14%) using their most dominant password on 75% or more of websites.

If users follow a strategy of having a single, dominant password that they try to reuse everywhere, then they are not succeeding in doing so very often. They use non-dominant passwords enough that they must have some other strategy for deciding when not to use it, and what password to use instead. That is, they must have other important constraints on their choices, or are using other strategies for password choice, that lead to a variety of different passwords (or password variants) on different websites. Next, we examine some of those possible constraints and strategies to better understand password variety.

5.2 Hypothesis 2: Focus on Ease of Creation

The next hypothesis that we examine is that users choose to focus their password decision-making on one of the major and obvious constraints that passwords have: the password composition policy that the website enforces. Almost all websites have requirements about properties that a password must have in order to be used on the website, but websites often have policies that are different from each other [36, 37]. These policies constrain user choices, and often force users to choose different passwords than they otherwise would voluntarily choose. (Indeed, that is the whole point of the policies!)

Whether choosing an entirely new password or creating a new variant of an existing password, one simple, straightforward way that users can choose passwords is to choose a password that just barely meets the minimum requirements of the policy. This is Hypothesis 2.1. Policies are often, but not always, explicitly written and posted, so users can choose passwords accordingly. Websites also allow users multiple attempts when creating a password so they may slowly add complexity to a password until it is allowed by the website. Our data does not allow us to examine how or why users make this choice, only whether the outcome of the choice matches the minimum requirements.

We identified a set of websites that users in our study entered passwords into that were visited frequently. We manually looked up password policies for these 274 domains. Not all domains were still active or accessible, but we were able to identify policies for 187

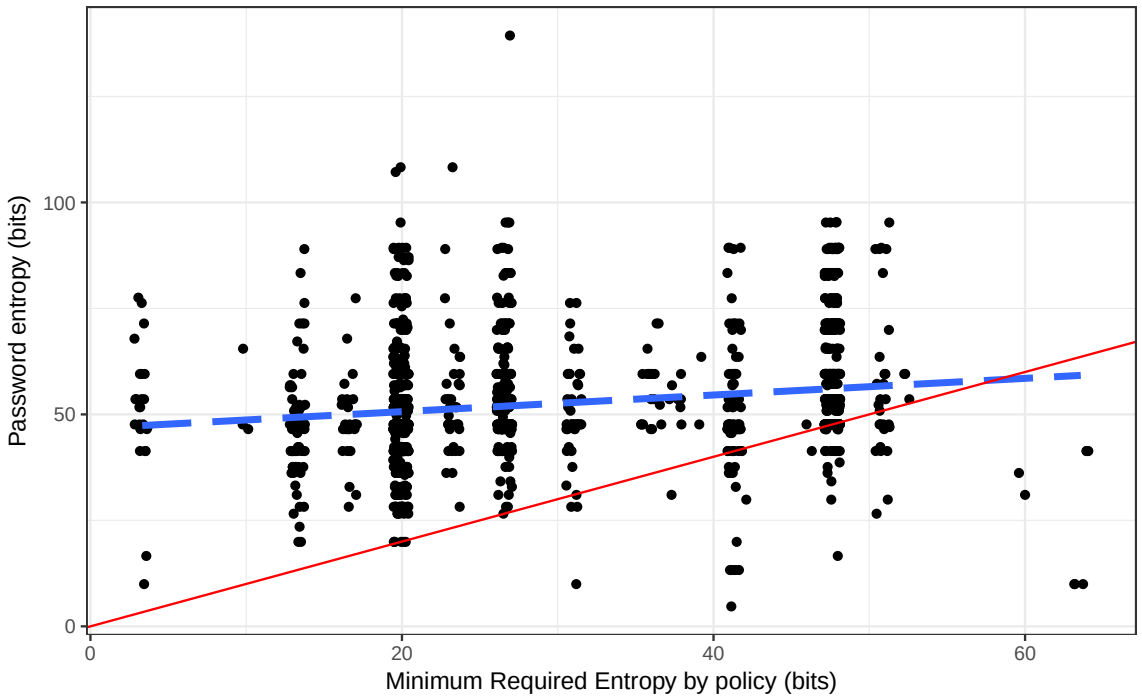

Figure 3: The relationship between password complexity (y-axis) and the minimum complexity required by the website's password policy (x-axis). The solid red line represents the minimum allowed policy ($y = x$); passwords above that line are more complex than they are required to be. The dotted blue line is a simple regression fit. Since complexity generally falls on a small number of values, points on this graph have been slightly spread to make them more visible.

of these domains. Together, these 187 domains represent 18.5% of the websites where passwords were entered, but 78% of the password entry events in our dataset. Our analysis in this subsection focuses on this subset of domains. To conduct this analysis, we calculated the minimum complexity required by the password policy and compare it to the complexity of the actual password used. Passwords can be below the minimum if the policy changed since the password was first chosen.

On average, the passwords entered into websites have 21 bits of complexity more than the password composition policy requires. This is a substantial difference; it roughly corresponds to the difference between a 7 character password (36 bits of complexity) and an 11 character password (57 bits of complexity), where the passwords are composed of lowercase letters and numbers.

Only 16.2% of passwords are at or below the minimum requirements for the website they are used in. 42% of users (56 out of 134) don't have a single password that is at or below the minimum required complexity; all of their passwords are more complex than the website where they were used requires them to be. Not a single user has all of their passwords at or below the minimum required by the policy. This suggests that users are gener-

ally not following a strategy of choosing passwords that just barely meet the requirements of the policy. This also means that users are not choosing to use the simplest password (or password variant) that meets the policy requirements.

One closely related alternative strategy is that users choose passwords by focusing on reuse, and choosing a password that just barely meets the minimum requirements of the multiple websites where it is reused. The combined set of websites may have requirements that are higher than any individual website in the set—for example, if one website requires 8 character passwords without character set requirements, and another requires only 6 character passwords but requires letters, numbers, and symbols, then the combined requirements are higher than both individual websites.

For each participant, we calculated the combined complexity requirements of all of the websites where they used each of their passwords, and compared that to the complexity of the password that was chosen. Only 30% of reused passwords are at or below this combined minimum. This means that about 70% of passwords are more complex than required on all of the websites that they are used on.e

Figure 3 illustrates the relationship between the minimum requirements of the password policy and the actual complexity of the password that was used on that website. There is a correlation between the two ($r = 0.18$, $p < 0.001$, 95% CI: $[0.09, 0.25]$). Some correlation is expected, since minimum requirements are usually technically enforced.

However, this correlation is much smaller than we would expect if password creation policies were the most influential constraint on password choice. To examine this, we simulated password policy enforcement by randomly choosing 1000 passwords for each website from the RockYou dataset[43], and then throwing them out and choosing again if the password didn't meet the policy requirements. The resulting random passwords had a $r = 0.23$ correlation with the minimum requirements. Simply enforcing password policies technically can induce a similar correlation to the one we observed, even when users do not intentionally choose passwords based on policies.

We also can look at the individual policy requirements separately. Figure 4 shows the relationship between the required length of password and the actual number of characters in the chosen password. Figure 5 shows the relationship between the number of character classes required by the policy and the number used in the password. 60% of the passwords we observed exceed both the length and character class requirements of the website where they were used.

Greene and Choong [44] suggest that there are specific parts of password policies (such as the "special characters requirement") that some end users find ambiguous and difficult to understand. Misunderstanding policies could lead to more complex passwords even if users were following the strategy in H2. If this were the case, then in our data, we should see a pattern where people match the minimum for the easy-to-understand parts (such as "at least 6 characters") and more variance about the hard-to-understand part. However, Figures 4 and 5, we still see patterns in the data that show password choices substantially longer and more complex than even misunderstood policies would produce.

These analyses provide little evidence that the users in our study are predominantly

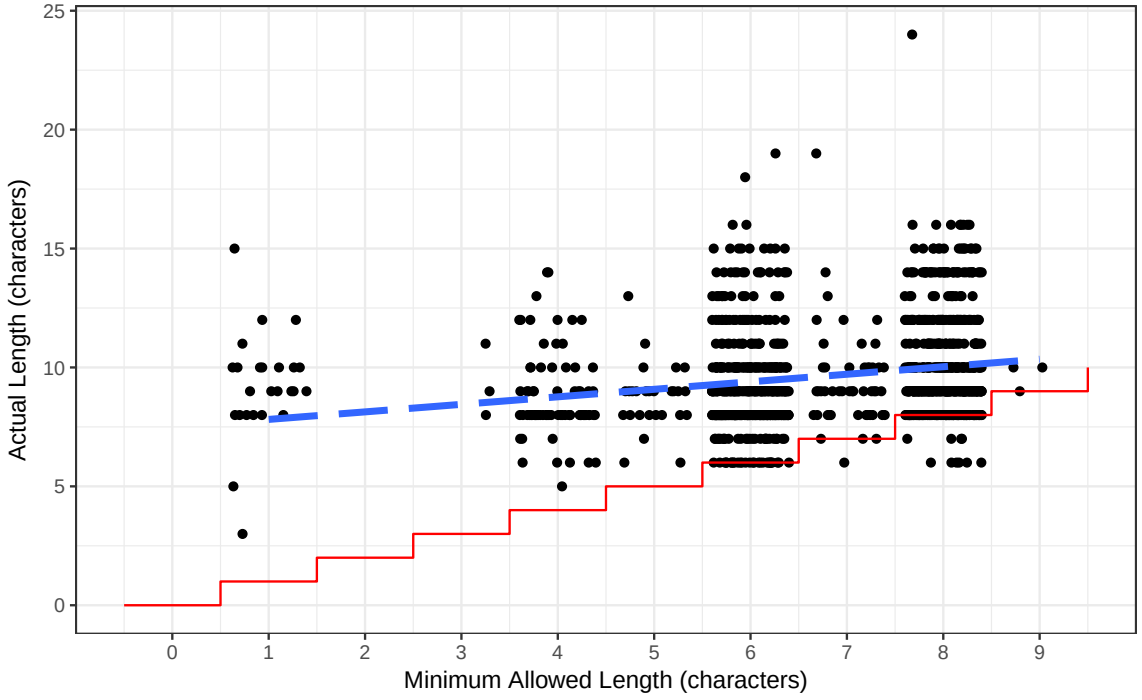

Figure 4: Required password length vs actual password length. The solid red line is the policy minimum and the dotted blue line is a simple regression fit. Values are whole numbers; points on this graph have been randomly spread horizontally to make them more visible.

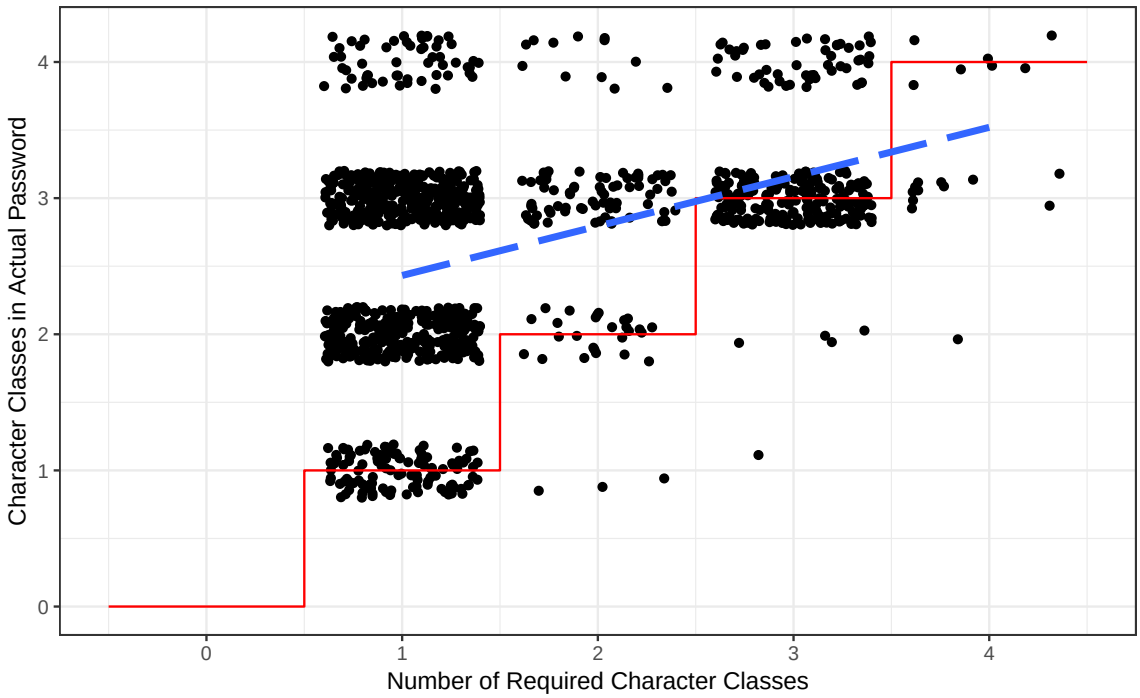

Figure 5: Required number of character classes in a password vs actual password character classes. The solid red line is the policy minimum and the dotted blue line is a simple regression fit. Values are whole numbers; points on this graph have been randomly spread out to make them more visible.

following a strategy of choosing passwords primarily based on meeting the minimum password policies of the websites where they use those passwords. The majority of passwords exceed all of the requirements of the policy, and while there is a correlation between policy and actual passwords, it is likely the result of technical enforcement of password policies. Past lab research has found that users frequently choose passwords based on policy [37]; however, those studies did not have a real-world context with real-world websites and real usability and security consequences to password choices. It is possible that in vivo usability or security concerns lead users to choose passwords that are more complex than they are required to; we analyze this next.

5.3 Hypothesis 3: Focus on Day-to-day Usability

Passwords are only chosen once, but then users have to live with that password and enter it in every time they are prompted by that website. Entering passwords can be a very time-consuming task; recent estimates suggest that it takes 10-14 seconds to enter a password every time it is needed [27], and users enter a password approximately 3.2 times per day [2], which, when combined, suggest that users are spending over 3 hours per year just entering passwords.

One way to balance security and usability is to focus on this usability need and choose passwords that are less complex, and therefore easier to enter. If this is the user's only concern, then they would likely choose the simplest allowable password; however, we showed above that this isn't happening much of the time. Instead, users may recognize that security is important and use more complex passwords for websites that they don't have to enter in very often. They then mainly use simple passwords for the everyday websites. Here we test this idea by looking for a relationship between the frequency of using a website and the complexity of the password that users chose for that website.

We begin by examining Hypothesis 3.1: users choose simpler passwords for websites that they have to enter a password into frequently. We calculated the number of times a password was entered into each website and divided it by the number of days between the first and last uses of the website by that user. This gives us an estimate for how often a password needs to be entered – the number of password entries per day. This has very little correlation with the complexity of the password used ($r(2531) = 0.00$, $p = 0.72$, 95% CI: $[-0.04, 0.03]$).

An alternative way of examining how often a password is entered is by looking at how many webpages on the site the user is able to view before being asked to log in again. If, like one of our participants, you are able to view 10,000 different pages on reddit in 6 weeks but are only asked to log in once, that may seem like a very efficient use of your time entering your password. We calculated the number of visits per password entry, but also found very little correlation with the complexity of the password used ($r(2531) = 0.03$, $p = 0.15$, 95% CI: $[-0.01, 0.06]$).

Entering a password is a relatively rare experience that many people likely don't explicitly remember doing. It is possible that instead of thinking about how often they enter the

	Estimate	*95% CI*	
(Intercept)	52.09***	51.37	52.83
Password entries per day	0.24	-0.30	0.80
Webpage visits per day	-0.08*	-0.14	-0.01
Webpage visits per password entry	0.003*	0.000	0.005
R^2	0.00		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 2: OLS Regression of the complexity of the password (random password entropy, measured in bits) for a website based on how often that password is entered into the website or how often that website is visited.

password, users simply think about how often they visit a website. This is Hypothesis 3.2: users choose less complex passwords when they visit a website more often. We also find a very small correlation here ($r(2531) = -0.03$, $p = 0.16$, 95% CI: $[-0.06, 0.01]$) between the number of webpage visits per day on a website and the complexity of the password used on that website.

Table 2 shows an OLS regression that includes all three of these usage effects together in a regression. This regression shows the relationship between password complexity and three usage indicators. The intercept—the overall average password complexity—is 52 bits. Additional password entries or webpage visits have a very small effect on the complexity of a password.

Overall, we find very little evidence that people take frequency of use into account when choosing a password. This makes sense; often users are asked to choose password very early in their use of a website, before they know how often they will use the site or how often the site will ask them to enter their password.

5.4 Hypothesis 4: Focus on Perceived Security

Choosing a password can be seen as making a security/usability tradeoff: for this website, do I want my password to be more secure, or easier to use? One possible way for users to make this tradeoff is to focus on security concerns: choose a more complex password on websites that the user believes need more security, and use a less complex password for websites where security is perceived to be less important. In this section, we examine the extent to which password choices seem to be following this strategy.

5.4.1 Security by Category

We begin by examining one way that users might determine whether a website has high security needs. Hypothesis 4.1 suggests that people look at the logical category of a website – financial website vs. travel website, for example – and uses that category to decide whether they should use a more complex password or a less complex password.

Category	*Complexity*	*# Users*
university	57.2	134
economy and finance	55.1	56
chat and messaging	55.1	18
news and media	54.3	11
information tech	54.2	110
media sharing	52.2	37
education	51.7	123
business	51.7	134
streaming media	50.3	44
shopping	50.0	112
social networking	50.0	100
entertainment	49.6	69
uncategorized	48.2	105
games	46.9	22
sports	45.8	19
travel	43.9	17

Table 3: Average complexity (in bits) of passwords used on websites, by logical category of the website. Only showing the 16 categories that at least 10 users visited; analyses were conducted with all 23 categories. Across all passwords, the average complexity was 49.5 bits. Each user may visit multiple websites in each category.

To examine this, we use the automated classification system Webshrinker [40] to classify all of the websites that the users in our study entered passwords into. Note that some websites are classified under more than one category, when appropriate. Since all of the participants in our study were students at the same university, we separated websites for that university out into its own category.

Table 3 shows the average complexity of passwords in different categories. There is a difference between categories; approximately 15 bits of complexity difference between the top category (university websites) and the category with the least complex passwords (travel websites). This is approximately the difference between an 8 character password with only letters (45 bits of complexity) and a 10 character password with both letters and numbers (60 bits of complexity). Statistically, we are able to rule out the null hypothesis of no differences between categories; some differences are statistically significant (one-way ANOVA, $F(22, 2080) = 6.62$, $p < 0.001$; $R^2 = 0.07$).

The hypothesis states that categories with higher security needs should have more complex passwords, but doesn't really state which categories represent higher security needs. University websites are likely important since all users in this study are students at the university. Financial websites are also commonly seen as important [36]. And those two

categories have the most complex passwords in our dataset. On the other end of the spectrum, travel websites, sports websites, and games have the least complex passwords in our dataset. This lends support to this hypothesis; there are moderately sized differences in complexity between categories, and the categories with approximately more security needs have more complex passwords.

However, it is also interesting to see places where this pattern doesn't hold. For example, social networking websites like Facebook and LinkedIn generally have lower than average password complexity, despite having lots of personal data. The Webshrinker category "news and media" seems like it has complex passwords, but that category mostly consists of reddit.com (since our participants did not seem to log into traditional news sites, and instead read them without logging in). It seems like websites that explicitly collect and store sensitive information (like financial websites) are more likely to have complex passwords, where websites that don't directly or explicitly ask users for sensitive information have less complex passwords.

5.4.2 Security by Importance

The type of website, and consequently the type of data on the website, are only one way of evaluating whether a website has high security needs. We asked a sample of 231 undergraduate students at the same university as the original study participants to rate 213 different domains as to how "important" that website was to them.

To calculate importance, we asked the survey respondents 19 survey questions that were all variations on "this website is important", such as "It would be bad if personal information that I enter into this website were stolen" and "Information about whether I visit this website is sensitive." Averaging multiple questions allows us to measure this construct with greater fidelity and to capture different aspects of what people might mean by "importance" [45, 46]. Questions were on a 5-point Likert scale. We averaged the responses to these 19 questions to create an overall measure of *importance* of a website. Each website was randomly presented to a different number of respondents (each respondent was only asked about 10 websites, to limit survey fatigue). Websites were rated by at least 5 respondents, and average 10.2 respondents per website. Survey instrument and raw data are available in the supplemental materials online: `https://osf.io/a28q9/`

Hypothesis 4.2 states that websites that were rated as subjectively more important will have more complex passwords than websites rated as less important. We found a positive correlation between these importance ratings and the complexity of passwords used on that website ($r(1517) = 0.09$, $p < 0.001$, 95% CI: $[0.04, 0.14]$). Table 4 shows a breakdown of password complexity grouped by importance. There is about 3.5 bits of complexity increase between the bottom quintile and the top quintile, which is approximately equivalent to replacing one letter with a symbol in a 9-character all letter password.

We also used the different importance questions to try to identify what aspect of websites was most strongly associated with higher complexity passwords. We grouped the importance questions into four categories: those about the website overall; those about

Quintile	*Complexity*	*Importance*
Low	51.0	2.65
	51.6	2.99
Medium	52.2	3.29
	54.2	3.48
High	54.6	3.65

Table 4: Password complexity grouped by the quintile of importance, with larger numbers indicating higher importance. The importance numbers are the average importance for that quintile on a 5-point Likert scale, averaged across all 19 importance questions.

Category	Correlation		95% CI	
Website Overall	0.11	***	0.06	0.16
Content on the Website	0.09	***	0.04	0.14
Whether You Visit the Website	0.06	*	0.01	0.11
Personal Information Entered into Website	0.02		-0.03	0.07

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5: Relationships between password complexity and importance, grouped by different categories of importance.

the content on the website (such as comments, posts, or photos); those about whether the respondent visits the website; and those about personal information the respondent enters into the website.

Table 5 shows the correlation between each category of importance and the complexity of passwords used. The strongest relationship was with questions about the importance of the website overall (and particularly the question "I am required to use this website"), and also with the importance of the content on the website. Surprisingly, websites that respondents rated as important because of personal information they enter showed very little relationship with password complexity.

Overall, we find a small relationship between the aggregate evaluation of importance of a website and complexity of passwords used on that website. The relationship is consistently in the expected direction – more complex passwords are used on websites seen as more important.

5.4.3 A Critical Situation: University Passwords

Following the principle of *strong inference*, we specifically looked for situations that can help us distinguish between competing hypotheses. In Section 2.4.4 above, we described such a situation: an organization's single sign on. A single sign on password both needs to be entered frequently (high usability requirements) and also protects a large amount

of important information (high security requirements). In such a situation, Hypotheses 3 and 4 make different predictions. If users are primarily following a strategy emphasizing usability (H3), they most should choose a low-complexity password that is easy to enter. If they primarily are following a strategy emphasizing security (H4), then they should choose a high-complexity password.

All users in our study were associated with a single university, and as such, all of them had a password with that university. The complexity of these passwords is labeled "university" in Table 3. These passwords were, on average, the highest complexity passwords of any category of website. Following the logic of *strong inference*, this data suggests that strategies following H4 are more likely to be followed by users than strategies from H3.

We can also use this situation to compare against Hypothesis 4 and Hypothesis 2 by looking at the password policy. At the time of the study, the minimum complexity for passwords at that university was 47 bits. This is relatively complex password policy, but it is still over 10 bits lower than the actual passwords. That is, the actual university passwords were on average 10 bits of entropy more complex than they were required to be, suggesting that users were not necessarily following Hypothesis 2 either.

University passwords are only a single website, but they represent a "critical experiment" that allows us to conduct strong inference because our three hypotheses make different predictions in this situation [11]. H2 and H3 both predict low complexity passwords, and H4 predicts high complexity passwords. The data is most consistent with Hypothesis 4, that users are following a strategy of choosing more complex passwords on websites with higher security needs.

6 Limitations

This study was conducted primarily with a population of non-technical students—young adults attending a large public university in the Midwest region of the U.S. with an undergraduate acceptance rate for Fall 2018 of greater than 70%[1]. All of our subjects are likely WEIRD[47]: members of Western, Educated, Industrialized, Rich, and Democratic societies, and it is unclear whether our results would generalize to other humans.

Students in general are an interesting population with which to study passwords. Prior research has found that password habits form early and are relatively stable over time [48], which suggests that patterns in the way they choose passwords will likely continue for much of these students' lives. However, a couple of non-peer-reviewed studies (from CSID and Research Now in 2012 [49], and from Digital Guardian in 2018 [50]) found that there appear to be generational differences in password use, with 18-to-24 year olds being the most likely to reuse passwords. It isn't clear whether older generations make password choices in the same way as the students in this study. The Digital Guardian study did find that across age groups, people self-reported prioritizing security over convenience approximately 65% to 35%, which supports a similar conclusion to the actual password data we

[1] `https://www.usnews.com/best-colleges/michigan-state-2290`

present here. However, there is some research that suggests older users choose stronger passwords [9]. Additionally, university students are generally more highly educated than the general population.

When we collected passwords, we sought to collect them as naturally as possible, with little intervention that would cause the participants to change their everyday behavior. For this reason, we were not able to ask our participants about their opinions of the webpages where they entered passwords because such an intervention would interrupt their web browsing and change their behavior. Instead, we asked a different sample of people (from the same sampling frame) about the importance of websites, two years later. And we asked a different sample about complexity and perceptions of password strength (Section 4) a year after that. Website and password opinions have been slowly changing over time [50], and it is possible that these changes over time limit how accurate our measurements are for the initial sample.

Password managers are a technology that can dramatically change the way passwords are chosen and used. We explicitly measured whether our participants had a password manager plugin installed in the web browser. We found that 26 of the participants (19%) had a third party password manager installed that was capable of remembering passwords (all participants used a web browser that had this capability also) [2], but only 8 participants (6%) used a password manager capable of generating passwords. Our sample has few users of password managers, which aligns with other recent research [51]. A prior analysis of this data found that the use of a password manager had no effect on password reuse across websites [2]. Pearman et al. [51] also found that even when people do use password managers, they usually do not use them to generate passwords. Very few participants used a password manager.

7 Summary and Discussion

Each website that requests a password is an opportunity for users to make a tradeoff between security and usability, and users can make this tradeoff differently for different websites. We identified four possible hypotheses about how users make this tradeoff, collected a dataset of user password choices, examined data to attempt to falsify those hypotheses, and directly compared these hypotheses as potential explanations against each other. While others have studied these hypotheses separately, they have never been quantitatively compared to see which one(s) best describe users' choices.

Users do seem to reuse passwords, as many others have noted [2, 10]. However, that research also shows that almost all users have more than one different passwords that they reuse. We found that most participants in our study don't even seem to have a dominant password — one that they use on most websites. Instead, they have multiple different passwords (or password variants) that they reuse on different websites. This suggests that reusing passwords is an incomplete answer to how users choose passwords; *which* password or password variant do they choose to reuse on which websites?

We examined the influence that password creation policies have on the passwords in this real-world situation. We found little evidence that participants in our study choose passwords that just barely meet this constraint. More than half of passwords used by participants in our study are both longer and use more character classes than they are required to, and on average passwords are significantly more complex than the policy requires. This suggests that for a large number of password decisions, users' choices are not constrained by password policies, and they are not choosing to create password variants to just barely meet some constraint in the password policy.

We found almost no evidence that participants in our study take into account concerns about the ease of entering passwords when choosing passwords. The frequency of use of a password has almost no relationship with the complexity of the chosen password.

However, we were unable to rule out the idea that users are choosing passwords by focusing on security concerns about the websites requesting the password. It appears that participants in our study chose more complex passwords on websites that might have important or sensitive information about them, and less complex passwords on websites that raise fewer security concerns. Our evidence suggests that the type of website is the primary driver of these concerns, though we do also find weak evidence for an overall website "importance" factor that may be playing a role also. Additionally, using university passwords as a "critical experiment" for strong inference found that password choices are most consistent with this idea that users are focusing on security concerns.

These findings are encouraging. They suggest that users consider passwords to be security-sensitive decisions, and are trying to make these decisions by focusing on security concerns more than other concerns or constraints. They also suggest that while security seems to be the major focus, users take usability seriously and choose less complex passwords on websites where there are fewer security concerns. The participants in our study appear to be following the logic recommended by security experts Florencio et al. about how to manage a portfolio of passwords efficiently [52]. These results also quantitatively confirm self-reports from users in previous qualitative studies [8, 25, 24, 26], but dispute some of the speculation about password choices by past quantitative studies [21, 22, 32, 37, 10].

This suggests that to really understand how users are choosing passwords, it is not enough to look just at passwords, or even at the set of passwords that a user has [8, 2, 10]. Passwords exist in an ecosystem, and an important part of that ecosystem is the set of websites (or apps/devices/etc.) that users need to authenticate to. Where the password is needed and what it is protecting is an important part of the way users make decisions about which password to use.

Our results suggest that there may be better ways to educate users about password security. Most current password advice focuses solely on the password, not on the importance or security needs of the website where the password is being used [53]. Authors of security advice should also consider including information about making usability tradeoffs and choosing different passwords for different websites. The participants in our study are taking this into account currently, and password advice might be able to better support users as

they make these tradeoffs if it included information about the security needs of the website the password was for.

It is not clear that users are actually achieving higher levels of security with their decisions. More complex passwords are not necessarily more secure [9, 17, 18]. Passwords protect against both offline brute force attacks and online guessing attacks, and there is a big difference in security needs between these attacks. It isn't clear that the increased complexity that results from these user decisions is enough to actually make users secure against offline attacks [54].

Kirlappos et al. found that when users in large organizations don't follow security policies, they often create "shadow" security policies that try to achieve reasonable security goals [55]. We found that a similar situation exists for passwords in everyday web use; participants in our study might not be exactly following security experts' guidance for passwords, but they do seem to be trying to come up with their own way of being secure. The participants in our study appear to be operating within what they have been told is a reasonable framework for thinking about the security of passwords (length + character classes).

This could help explain why password meters—real-time feedback about the strength of a password as it is being entered—are often acceptable to users and often lead to more complex passwords [56, 57]. Users are likely using the feedback to better accomplish their security goals, at least for websites they deem have high enough security needs.

From the perspective of end users, mandatory password policies effectively focus on the usability part of the security/usability tradeoff. While password policies try to force users into being more secure, password meters effectively enlist end users as partners in security. At least for passwords, our results suggest that this is a reasonable strategy. Understanding how users are trying to be more secure can hopefully help technologists design security systems that support users in achieving their security goals, rather than try to force them to meet goals set by other people.

Passwords represent a situation where end users have some amount of control over the usability of their technology. While usability is important, it is not necessarily the highest priority; our participants seemed to be willing to voluntarily trade off usability to gain security in situations where it was warranted.

8 Acknowledgements

We thank Kami Vaniea, Tyler Olson, Nick Saxton, Nathan Klein, Raymond Heldt, Ruchira Ramani, Jallal Elhazzat, Tim Hasselbeck, Shiwani Bisht, Robert Plant Pinto Santos, Meghan Huynh, Simone Merendi, and Cindy Ochoa for assistance in developing the software and analyzing the data.

9 Funding

This work was supported by the U.S. National Science Foundation [CNS-1116544 and CNS-1115926].

References

[1] Stephen Phillips. Passwords could be past tense by 2002. *Computer Weekly*, November 26:12, 1998.

[2] Rick Wash, Emilee J Rader, Ruthie Berman, and Zac Wellmer. Understanding Password Choices - How Frequently Entered Passwords Are Re-used across Websites. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2016.

[3] Martina Angela Sasse, Michelle Steves, Kat Krol, and Dana Chisnell. The Great Authentication Fatigue–And How to Overcome It. In *Proceedings of the Cross-Cultural Design 6th International Conference (CCD)*, pages 228–239, 2014.

[4] Rick Wash and Jeffrey K MacKie-Mason. Security when people matter: structuring incentives for user behavior. In *ICEC '07: Proceedings of the ninth international conference on Electronic commerce*, page 7, New York, New York, USA, August 2007. ACM.

[5] Joseph Bonneau, Cormac Herley, Paul C van Oorschot, and Frank Stajano. Passwords and the evolution of imperfect authentication. *Communcations of the ACM*, 58(7):78–87, 2015.

[6] Elizabeth Stobert and Robert Biddle. Expert password management. In Frank Stajano, Stig F. Mjølsnes, Graeme Jenkinson, and Per Thorsheim, editors, *Technology and Practice of Passwords*, pages 3–20, Cham, 2016. Springer International Publishing. ISBN 978-3-319-29938-9.

[7] Iulia Ion, Rob Reeder, and Sunny Consolvo. "... no one can hack my mind": Comparing Expert and Non-Expert Security Practices. *... on Usable Privacy and Security (SOUPS)*, 2015.

[8] Elizabeth Stobert and Robert Biddle. The Password Life Cycle. *ACM Transactions on Privacy and Security (TOPS)*, 21(3):13–32, June 2018.

[9] Joseph Bonneau. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *IEEE Symposium on Security and Privacy*, pages 538–552. IEEE, 2012.

[10] Sarah Pearman, Jeremy Thomas, Pardis Emami Naeini, Hana Habib, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, and Alain Forget. Let's Go in for a Closer Look: Observing Passwords in Their Natural Habitat. In *the 2017 ACM SIGSAC Conference*, pages 295–310, New York, New York, USA, October 2017. ACM.

[11] John R. Platt. Strong inference. *Science*, 146(3642):347–353, 1964. ISSN 0036-8075. doi: 10.1126/science.146.3642.347. URL `https://science.sciencemag.org/content/146/3642/347`.

[12] Karl Popper. *The Logic of Scientific Discovery*. Routledge, 2005.

[13] Abe Singer and Warren Anderson. Rethinking password policies. *;login: The Usenix Journal*, 38(4):14–18, August 2013.

[14] William Burr, Donna Dodson, Elaine Newton, Ray Perlner, W. Timothy Polk, Sarbari Gupta, and Emad Nabbus. Electronic authentication guideline. NIST Special Publication 800-63-2, US National Institute of Standards and Technology, August 2013.

[15] Karen Scarfone and Murugiah Souppaya. Guide to enterprice password management (draft). NIST Special Publication 800-118, US National Institute of Standards and Technology, April 2009.

[16] Richard Shay, Saranga Komanduri, Patrick Gage Kelley, Pedro Giovanni Leon, Michelle L Mazurek, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Encountering stronger password requirements: user attitudes and behaviors. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2010.

[17] 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 *the 2013 ACM SIGSAC conference on Computer and Communications Security (CCS)*, pages 173–186, Berlin, Germany, November 2013. ACM.

[18] Patrick Gage Kelley, Saranga Komanduri, Michelle L Mazurek, Richard Shay, Timothy 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 *IEEE Symposium on Security and Privacy (SP)*, pages 523–537. IEEE, 2012.

[19] 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 *USENIX Security Symposium*, 2015.

[20] William Melicher, Blaze Ur, Sean M Segreti, Saranga Komanduri, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Fast, Lean, and Accurate: Modeling Password Guessability Using Neural Networks. In *USENIX Security Symposium*, August 2016.

[21] Blase Ur, Jonathan Bees, Sean M Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Do Users' Perceptions of Password Security Match Reality? In *ACM Conference on Human Factors in Computing (CHI)*, pages 3748–3760, New York, New York, USA, May 2016. ACM.

[22] Blaze Ur, Fumiko Noma, Jonathan Bees, Sean M Segreti, Rich 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 *USENIX Security Symposium*, Ottawa, Canada, 2015.

[23] Beate Grawemeyer and Hilary Johnson. Using and managing multiple passwords: A week to a view. *Interacting with Computers*, 23(3):256–267, 2011.

[24] Gilbert Notoatmodjo and Clark Thomborson. Passwords and Perceptions. In *Proceedings of the Seventh Australasian Conference on Information Security (AISC)*, pages 71–78, 2009.

[25] S M Taiabul Haque, Matthew Wright, and Shannon Scielzo. A study of user password strategy for multiple accounts. In *Proceedings of the third ACM conference on Data and Application Security and Privacy (CODASPY)*, pages 173–176, 2013.

[26] Geoffrey B Duggan, Hilary Johnson, and Beate Grawemeyer. Rational security: Modelling everyday password use. *Journal of Human Computer Studies*, 70(6):415–431, 2012.

[27] Michelle Steves, Dana Chisnell, Angela Sasse, Kat Krol, Mary Theofanos, and Hannah Wald. Report: Authentication Diary Study. Technical report, National Institute of Standards and Technology, Information Technology Laboratory, Gaithersburg, MD, February 2014.

[28] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.

[29] Ariel Rubinstein. Comments on the interpretation of game theory. *Econometrica*, 59 (4):909–924, 1991.

[30] P.-A. Chiappori, S. Levitt, and T. Groseclose. Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, 92(4):1138–1151, September 2002.

[31] Robert Axelrod. *The Evolution of Cooperation*. Basic Book, revised edition, 2006.

[32] Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussman. Survival of the Shortest: A Retrospective Analysis of Influencing Factors on Password Composition. In *Proceedings of Human-Computer Interaction–INTERACT*, pages 460–467, 2013.

[33] Mat Honan. How apple and amazon security flaws led to my epic hacking. *Wired*, June 2012.

[34] Rick Wash and Emilee Rader. Human interdependencies in security systems. In *CCC Visioning Workshop on Grand Challenges in Sociotechnical Cybersecurity*, 2016.

[35] Philip G Inglesant and M Angela Sasse. The true cost of unusable password policies: password use in the wild. In *ACM Conference on Human Factors in Computing (CHI)*, pages 383–392, New York, New York, USA, April 2010. ACM.

[36] Dinei Florêncio and Cormac Herley. Where Do Security Policies Come From? In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2010.

[37] 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 *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 2595–2604, 2011.

[38] Ameya Hanamsagar, Simon S. Woo, Chris Kanich, and Jelena Mirkovic. Leveraging semantic transformation to investigate password habits and their causes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 570:1–570:12, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3174144. URL `http://doi.acm.org/10.1145/3173574.3174144`.

[39] The internet archive wayback machine, 2018. URL `https://archive.org/web/`.

[40] Webshrinker, 2018. URL `https://www.webshrinker.com/`.

[41] Blakeley B. McShane, David Gal, Andrew Gelman, Christian Robert, and Jennifer Tackett. Abandon statistical significance. *The American Statistician*, 73(S1):235–245, 2019.

[42] Franklin P. Tamborello and Kristen Greene. Memory and motor processes of password entry error. In *Proceedings of the Human Factors and Ergonomics Society*, 2015.

[43] Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proceedings of the 17th ACM Conference on Computer and Communications Security*,

CCS '10, pages 162–175, New York, NY, USA, 2010. Association for Computing Machinery. doi: 10.1145/1866307.1866327.

[44] Kristen K. Greene and Yee-Yin Choong. Must i, can i? i don't understand your ambiguous password rules. *Information and Computer Security*, 2017.

[45] Don Dillman, Jolene Smith, and Leah Melani Christian. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Wiley, 4th edition, August 2014.

[46] Robert Devellis. *Scale Development: Theory and Applications*. Number 26 in Applied Social Research Methods. SAGE Publications, Inc., fourth edition, April 2016.

[47] Joseph Henrich, Steven J Heine, and Ara Norenzayan. Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, 33(2-3):111, 2010.

[48] Anthony Vance, Mikko Siponen, and Seppo Pahnila. Motivating is security compliance: Insights from habit and protection motivation theory. *Information and Management*, 49(3):190 – 198, 2012. ISSN 0378-7206. doi: https://doi.org/10.1016/j.im.2012.04.002. URL http://www.sciencedirect.com/science/article/pii/S0378720612000328.

[49] CSID and Research Now. Consumer survey: Password habits. White Paper, September 2012. URL https://www.csid.com/wp-content/uploads/2012/09/CS_PasswordSurvey_FullReport_FINAL.pdf.

[50] Nate Lord. Uncovering password habits: Are users' password security habits improving? Published on the Digital Guardian Blog, December 2018. URL https://digitalguardian.com/blog/uncovering-password-habits-are-users-password-security-habits-improvi

[51] Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Why People (Don't) Use Password Managers Effectively. In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security*, SOUPS'19, pages 319–338, Berkeley, CA, USA, 2019. USENIX Association. ISBN 978-1-939133-05-2. URL http://dl.acm.org/citation.cfm?id=3361476.3361500.

[52] Dinei Florêncio, Cormac Herley, and Paul C van Oorschot. Password Portfolios and the Finite-Effort User: Sustainably Managing Large Numbers of Accounts. In *Proceedings of the 23rd USENIX Security Symposium*, pages 575–590, 2014.

[53] Emilee J Rader and Rick Wash. Identifying patterns in informal sources of security information. *Journal of Cybersecurity*, 1, 2015.

[54] Dinei Florêncio, Cormac Herley, and Paul C van Oorschot. An administrator's guide to internet password research. In *Proceedings of the 28th USENIX conference on Large Installation System Administration (LISA)*, pages 44–61, 2014.

[55] Iacovos Kirlappos, Simon Parkin, and M Angela Sasse. Learning from "Shadow Security". In *NDSS Workshop on Usable Security*, 2014.

[56] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L 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. *USENIX Security Symposium*, 2012.

[57] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, Noah Johnson, and William Melicher. Design and Evaluation of a Data-Driven Password Meter. In *ACM Conference on Human Factors in Computing (CHI)*, pages 3775–3786, New York, New York, USA, May 2017. ACM.