Understanding Password Choices: How Frequently Entered Passwords Are Re-used across Websites

Rick Wash and Emilee Rader, ***Michigan State University;*** **Ruthie Berman,** ***Macalester College;***
Zac Wellmer, ***Michigan State University***

https://www.usenix.org/conference/soups2016/technical-sessions/presentation/wash

This paper is included in the Proceedings of the
Twelfth Symposium on Usable Privacy and Security (SOUPS 2016).

June 22–24, 2016 • Denver, CO, USA

ISBN 978-1-931971-31-7

Open access to the Proceedings of the
Twelfth Symposium on Usable Privacy
and Security (SOUPS 2016)
is sponsored by USENIX.

Understanding Password Choices: How Frequently Entered Passwords are Re-used Across Websites

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

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

Ruthie Berman
Macalester College
rberman@macalester.edu

Zac Wellmer
Michigan State University
wellmerz@msu.edu

ABSTRACT

From email to online banking, passwords are an essential component of modern internet use. Yet, users do not always have good password security practices, leaving their accounts vulnerable to attack. We conducted a study which combines self-report survey responses with measures of actual online behavior gathered from 134 participants over the course of six weeks. We find that people do tend to re-use each password on 1.7–3.4 different websites, they reuse passwords that are more complex, and mostly they tend to re-use passwords that they have to enter frequently. We also investigated whether self-report measures are accurate indicators of actual behavior, finding that though people understand password security, their self-reported intentions have only a weak correlation with reality. These findings suggest that users manage the challenge of having many passwords by choosing a complex password on a website where they have to enter it frequently in order to memorize that password, and then re-using that strong password across other websites.

1. INTRODUCTION

Passwords are a key part of many security technologies; they are the most commonly used authentication method. For a password system to be secure, users must make good choices about what password to use, and where to re-use passwords. Advice from security experts directs people to create, remember, and use passwords that are long, random, and unique to each account [21]. However, evidence from prior research suggests that people struggle to comply with this advice. For example, Das et al. [7] estimated that 43-51% of users re-use passwords across accounts, and Ur et al. [36] found that people feel like re-using passwords is not a problem, because they have never personally experienced negative consequences stemming from re-use. In reality, password re-use can introduce a serious security vulnerability which is difficult for any individual service operator to protect against [7].

People self-report that they re-use passwords to cope with the difficulty of remembering too many passwords, and that they believe they are not at risk because they re-use mainly passwords they believe are strong [36]. It isn't clear whether these self-reports represent wishful thinking by the users or whether they accurately reflect actual behavior. Few studies have been able to connect users' password-related attitudes and intentions with their own real-world password behavior, across accounts and over time.

It is especially important to be able to draw these connections between self-report and actual behaviors regarding password re-use, because re-use is a coping mechanism that occurs as a result of the demands and constraints users face when authenticating. Re-use is a user response to the burden of allocating limited memory capacity across the accounts and systems people use on a daily basis [15]. Despite many attempts to design more secure and usable systems, passwords remain one of the most widely deployed security systems in use today. The majority of people who use computers enter a password at least once a day; prior estimates [12, 30] suggest that computer users undertake between 8 and 23 password entry events every day![1]

We analyze a dataset that measures actual use and re-use of real-world passwords for web accounts. We captured password entry events that occurred in 134 subjects' web browsers over approximately six weeks. We also surveyed those same subjects immediately before and after the study period to collect self-reported demographics, attitudes, and intentions related to passwords. This allows us to examine not only how people think about passwords, but how that thinking translates into real-world password creation and re-use.

We found that people are re-using passwords across multiple websites. Our subjects primarily re-used passwords that are more complex, and re-used passwords that they entered frequently, such as the password for their university's website. We suspect that frequently entering a password is a way to memorize strong passwords, which are then re-used because

Symposium on Usable Privacy and Security (SOUPS) 2016, June 22–24, 2016, Denver, Colorado.

[1]Our users entered an average of 3.8 passwords per day that they were active on their computer, or 3.2 passwords per day overall.

they are already very familiar. We also found that when asked about password use, subjects' responses were correlated with their actual password behaviors, but the correlation is relatively weak. This suggests that password choices are intentionally made, but that there are influences on password behavior other than password intentions.

Our results illustrate an important constraint on users' behavior that impacts password choice: how often a user is forced to authenticate with a particular password is related to how much they re-use that password on other accounts. This presents an opportunity for organizations to encourage the memorization of objectively strong passwords. However, it also results in greater potential for cross-site vulnerabilities as users prioritize using that stronger password in more places.

2. RELATED WORK

2.1 Password Creation and Management

People use passwords to authenticate on many different systems and servers on a daily basis. Estimates of the number of accounts that users maintain range from an average of 7-8 per person reported in a 2006 paper using data from a self-report user study with 58 subjects [16], to around 25 per user measured in a large-scale data collection including data from 544,960 browsers that was conducted around the same time [12]. In 2013, an online survey of 583 subjects found an average of about 18 accounts per user (median 14). And in a recent interview study, Stobert and Biddle [33] found that subjects reported having between 9 and 51 accounts, with a median of 27. People log in to a significant fraction of these accounts daily; Florêncio and Herley [12] found that people enter 8.11 passwords per day, and Hayashi and Hong [19] estimated that people use around 12 accounts per day.

Common password advice directs users to create passwords that are unique to each account, and random. However, for most people, authentication is a secondary task that presents a hurdle they must overcome in order to accomplish their primary task [2, 8]. So when people create passwords, their main goal is to make them easy to remember so entering a password does not impede their progress. When creating passwords, people often use information that is meaningful and important to them [8, 33], or has some connection to the service for which they are creating the password. For example, Inglesant and Sasse [20] reported that a subject described creating a password based on an item on his or her desk. People often use common names, words, and phrases in their passwords [29]; Shay et al. [31] found that about 80% of subjects reported they based their passwords on a word or a name. People also use rules, or an 'algorithm' [36], to compose new passwords. These strategies allow people to more easily recall their passwords when they are needed [6].

Creating easy-to-remember passwords is especially important for people as the number of passwords they must remember increases. The more passwords one has, the harder it is to remember all of them [38]. Infrequently used passwords are also harder to remember, as are passwords that people are forced to change on a regular basis [30]. Despite these difficulties, memorization is still a common strategy for managing passwords. Several studies have found that relying on one's memory is more common than other mechanisms of storing passwords such as saving passwords in one's browser, using password manager software, or writing down passwords in an electronic file or on paper [16, 19]. Only two out of 49 subjects in a recent think-aloud lab study conducted by Ur et al. [36] reported that they use a password manager; 17 said that they "simply memorize their passwords without writing them down or storing them anywhere". Another common strategy is relying on automatic software mechanisms to store passwords. For example, 81% of subjects in an interview study conducted by Stobert and Biddle [33] said that their passwords are stored in their browsers or in the Apple Keychain. People write down or store hard-to-remember passwords even when they recognize that this is a "bad" password management strategy [34].

2.2 Password Strength and Guess Resistance

Because people report that they rely on their memories for password management and feel like they're protecting against attacks by other human beings [33, 35], even when they create passwords that meet or exceed forced constraints imposed by password composition policies [36], their passwords are still not very complicated. In a lab experiment that asked subjects to create passwords for 8 different kinds of websites, passwords for sites that subjects rated as less important were shorter, and for the least important sites the passwords tended to be lowercase only [18]. However, people do self-report that they try to use stronger passwords for more sensitive accounts [34, 17, 8]. Ur et al. [36] found that subjects believed adding a digit or a symbol to a password they were already using elsewhere would make it stronger and more secure.

In a survey of people affiliated with Carnegie Mellon University who updated their password as a result of a changed password composition policy, only 24% of respondents reported creating a password of length 8 (the minimum length to meet the new requirements); the rest of the passwords were longer [31]. The average length of the passwords of the subset of subjects who answered questions about length and the types and positions of classes of characters in their passwords was 10.1 characters, with estimated entropy of 31 bits. In contrast, Bonneau [3] found in a dataset of 70 million passwords (69.3 million users) for Yahoo! sites collected in May 2011 that passwords were in the 10-20 bit range, and Florêncio and Herley [12] found in their dataset from 544,960 browsers that had the Windows Live Toolbar installed (between 7/24/06 and 10/1/06) that average entropy was 40.54 bits.

Traditionally, password strength has been measured using Hartley Entropy [4]: the log base two of the size of the set of possible passwords. This corresponds to Shannon's definition of entropy only when all passwords are equally likely. However, Hartley Entropy mostly measures complexity, and is not a good measure of objective password strength when it comes to offline guessing attacks. While there is a relationship between entropy and guess resistance [22], Bonneau found that entropy doesn't measure the same thing as guess resistance [3]. It does not take into account that there are non-random patterns in users' password creation choices that make guessing easier than if all passwords were random sequences of characters. For example, a longer password made up of a dictionary word is easier to guess but can have a higher entropy score than a shorter, random password [14]. People who engage in the common practice of adding num-

bers to the ends and capitalizing the beginnings of passwords expect that this makes their passwords stronger. However, this is not the case; passwords with these patterns are likely to be less guess resistant in an offline attack because they are non-random. Password composition policies may be able to increase entropy, but adhering to a composition policy is not a guarantee of guess resistance [36].

2.3 Password Re-Use

In addition to creating passwords that are easy to remember, people cope with the cognitive demands of authenticating on many different systems by re-using passwords. This is a very common practice; for example, 50% of subjects in an interview study conducted by von Zezschwitz, De Luca, and Hussman [37] reported that they re-used passwords, and explained that if they did not re-use passwords it would be too hard for them to remember them all. In that same study, 45% of subjects said they were still using the very first password they had ever created, and most of them were still using it to create new accounts! Florêncio and Herley [12] collected "re-use events" where a password was re-used across different websites, and found that in 2006 an average user had 6.5 passwords, and each was used on 3.9 different accounts. Komanduri et al. [23] found that even when subjects in their online experiment did not reuse exact passwords in their entirety, they created new passwords by modifying existing passwords. Less than 30% of subjects in Shay et al.'s survey [31] said they had created an entirely new password to meet the new password requirements—most said they modified a password they were already using. Only three subjects in Ur et al.'s 49-subject think aloud study [36] said they would never re-use passwords; most said they had not experienced any problems stemming from password re-use on any of their accounts.

Analyses of leaked password datasets also show that people re-use passwords on multiple different accounts. For example, Das et al. [7] identified 6077 usernames that appeared in two or more leaked password datasets; for 43% of these usernames the passwords on the different sites were identical, and for 19% they were similar. Bailey, Dürmuth and Paar [1] obtained access to a dataset containing usernames and the associated passwords that had been collected by a malware trojan, and calculated a metric they called the "re-use rate": for two randomly chosen accounts of a random user, how likely is it that the two passwords for the accounts are identical? In their dataset, the re-use rate for identical passwords was 14%, and for similar passwords it was 19%. Most of the password re-use in their dataset was "exact" reuse of an entire password on another site. One subject in Sasse et al.'s interview study [30] said that they have one "central" password that they use for everything, which they make as strong as possible.

The more accounts people have, the more they report that they re-use passwords across accounts [27]. One finding from Inglesant and Sasse's diary study [20] was that people use "good" passwords—ones that are memorable and conform to password composition policy—as a "resource" they return to again and again when creating passwords for new accounts. Subjects in Stobert and Biddle's interview study [33] spoke about re-using passwords on infrequently used accounts because those accounts had less "need for security". Many other self-report studies have found that people categorize accounts and re-use the same password for accounts that are similar to each other. People say that they re-use passwords more on low-importance accounts, and avoid password reuse for high-importance accounts that have a greater need for security [16, 27, 33]. However, in a lab study, Haque, Wright and Scielzo [18] found that it was possible to use a common password list and knowledge of a subject's password created in the "lower-level" account condition to successfully guess their "higher-level" account condition passwords 33% of the time. This indicates that people's beliefs and intentions may be inconsistent with their actual re-use of passwords across account categories. Because lower-level accounts may be easier to compromise [14, 1], such re-use is a risky security practice.

2.4 Research Questions

Password Reuse:

There are contradictory results in the literature regarding which passwords people re-use more often. Most password data that speaks to re-use is self report from user studies, in which people say that they tend to reuse weaker passwords more often than stronger passwords (e.g., Stobert and Biddle [33]). However, Egelman et al. [10] found that there was no difference in password strength between passwords created by subjects in their experiment who reported reusing existing passwords, and those who said they had not re-used passwords. And Ur et al. [36] found that subjects believed re-use would not be a problem for them, because they felt that the passwords they re-use are strong. In addition, when asked about why they re-use passwords, subjects in many studies self-report that it makes passwords easier to remember [16, 36]; this implies that due to memory constraints passwords that users have to enter more frequently should also be re-used on more different accounts.

In order to measure re-use directly, it is necessary to have access to repeated instances of password use over time by the same person, and a mechanism that makes it possible to compare passwords to find out whether a person has entered the same password on more than one account. Florêncio and Herley [12] had access to this kind of data, and found that strong passwords are re-used at fewer sites ($M = 4.48$); weak passwords are used at more sites ($M = 6.06$). However, Bailey, Dürmuth and Paar [1] found in a different dataset that password re-use is more common for the high-value accounts (e.g., financial accounts) which have stronger passwords, than for all accounts. In our study, we collected data from specific individuals over a period of weeks. This means that we can examine which passwords are reused more by specific individuals, and on how many different accounts the frequently-entered passwords are re-used. Therefore, we ask:

Do people reuse their strong(er) passwords more, or their weak(er) passwords more? Do people reuse frequently entered passwords more than infrequently entered passwords?

Password Intentions:

In some studies, people self-report that they do have some idea what strong versus weak passwords look like, and what they say mirrors common password advice. Generally speaking, people report that they know unique and random passwords are more secure [16]. Ur et al.'s [35] subjects knew

that, for example, it was better to put upper-case letters, digits, and symbols in the middle of passwords rather than at the beginning or end, and that randomly chosen digits are better than years or "obvious sequences". But, when people create passwords, analyses of leaked datasets and experiment passwords show that they do not behave consistently with this knowledge [7]. They choose passwords that are simpler and easier to remember [38]. There is evidence from previous research about software updates that users do not always enact their security intentions correctly [40], however, this has not been examined before with respect to passwords. In our study, we collected log data about individuals' behaviors and survey data asking about their intentions, so we can connect how users think about passwords with password strength and re-use more directly than was possible in previous work. Therefore, we ask:

Do peoples' intentions for the passwords they create correlate with the characteristics of their actual passwords, and with which passwords they reuse more?

3. METHOD

3.1 Methods for Studying Passwords

Researchers have used a number of different methods to study passwords. Each method has strengths and weaknesses. Interview studies like Stobert and Biddle [33] allow for in-depth questioning about a small number of users, but are hindered by the tendency to remember what one normally or typically does and not what one actually does. This is a problem for password research, because for most users passwords are a secondary task [30]. This can mean their memory for their past behavior is biased. Diary studies like Hayashi and Hong [19] help to get around that by asking users to record instances of password behaviors, but are only as accurate as subjects are able to adhere to the data recording protocol and routine, and can only be conducted with a small number of users. Surveys (e.g., Shay et al. [31], Ur et al. [35]) allow the researcher to gather data from many more people, on the order of hundreds to thousands, but are limited in that they are self-report which may be inaccurate, especially when it comes to security intentions which might not match actual behavior [40].

User studies conducted in the lab or online often ask subjects to create passwords under specific conditions, and typically take steps to create scenarios that closely approximate situations users are likely to encounter in the real world to increase external validity of the research (e.g., Egelman et al. [10]). Online user studies such as Komanduri et al. [23] using Amazon Mechanical Turk can potentially reach a large number of people. However, many people behave differently when creating passwords for a user study than they do normally [11].

Password datasets collected through partnerships with companies or organizations and leaked password datasets include users' actual passwords, and some of these datasets are quite large. The security community has used these datasets to learn more about the passwords users choose, and analyzed them for patterns of common password composition characteristics. However, these datasets typically include little information about the users who created the passwords. An exception is Mazurek et al. [24] which through a partnership with Carnegie Mellon University was able to analyze password data from every account holder. This study correlated demographic data about faculty, staff and students of the university with password characteristics, in addition to analyzing the guess resistance of the passwords. Two papers use data collected over days (in the case of Bonneau [3], 69.3 million users) or months (in the case of Florêncio and Herley [12], 544,960 users) to present findings at the user level as well as at the password level.

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 sample

The study by Florêncio and Herley [12] is the most similar study to ours. However, they only were able to collect "reuse events": instances when a password was reused across more than one website. We have more accurate data about how frequently a password is entered into each website, data about passwords that were only entered into a single website (69% of passwords in our study), and self-report data about user perceptions.

3.2 Data Collection and Participants

Our study combines survey methods asking subjects about beliefs, behaviors and behavioral intentions, with log data about actual behaviors over time. Subjects installed custom-written log data collection software on their personal computers and web browsers for a median duration of six full weeks, and also took a survey at the beginning and at the end of the data collection period. This allowed us to collect both self-reported beliefs, behaviors and behavioral intentions and log-based behavioral measures for the same subjects, which enabled us to correlate subjects' security beliefs and intentions with their actual password characteristics and re-use. In this way we can examine how knowledge, attitudes, and intentions match up with behaviors within a person.

Our data collection software consisted of a web browser plugin for both Google Chrome and Mozilla Firefox. This plugin collected web use data, and uploaded it to our server. The plugin recorded all URLs visited by the web browser, as well as any form submission on a web page. Additionally, the plugin recorded all security-related settings and recorded information about all add-ons (plugins, extensions) installed and/or running. The plugin did not record anything while the user was in Private Browsing mode (Firefox) or Incognito mode (Chrome); subjects were instructed to use these modes for activities they did not want recorded. All connections to our server were encrypted to protect user privacy.

When the plugin detected a password HTML element in a form submission, it recorded the password entry: when the user entered the password, what webpage the user entered a password into, which password was entered, and how strong each password was (entropy, following Florêncio and Herley [12]). We did not collect plain text passwords; instead our browser plugin measured password entropy on the client and then hashed the passwords with a per-user salt before the information was sent to our server. This enabled us to examine which passwords were re-used by each subject across different websites without knowing his or her actual passwords. Additionally, since we collected data for a number of weeks, we were able to identify which passwords were re-used by each subject, and on what websites. We were not able to compare plain text passwords across subjects.

We recruited subjects from a large midwestern university by asking the registrar to email a random sample of students (both undergraduate and graduate). Students in computer science and engineering were excluded from participating. We sent out a total of 15,000 emails in three waves, and had approximately 247 students respond to our recruiting mail (1.6% response rate). Of those 247, about 180 were eligible to participate in the study: they had a personal computer running Windows 7 or Windows 8 which they said they used regularly, used either Google Chrome or Mozilla Firefox as their main web browser, and responded to our instruction emails. They were also required to have the ability to install software on the computer, and be the only user of the computer. The first two constraints (Windows, web browsers) are a limitation of our data collection software—supporting other operating systems and web browsers was prohibitively complex, so we designed the software to support the most popular operating systems and web browsers.

Of those subjects that were eligible to participate in the study, we received usable data from 134 subjects (0.8% usable response rate). The remaining subjects mostly were excluded due to unforeseen bugs in the data collection software that prevented sending accurate data, or because subjects did not use their computer enough (e.g. had more than 7 consecutive days without using the computer, not counting spring break). Two subjects had hardware problems with their computer that caused them to withdraw, and two other subjects withdrew without explanation. Our sample is fairly representative of the population of the university. Almost all subjects were in the 18-29 age range. Close to the demographics of the student population, our sample was 52% female and 76% white. Approximately 76% of the subjects were undergraduates, while the remaining are graduate students. Only 3 of the 122 subjects had children. Table 1 has more details.

All subjects provided informed consent to the data collection. Subjects were compensated a total of $70 for their participation; those who withdrew early received partial compensation. Subjects had the ability to turn off the data collection software at any time using a control panel that we provided, and we also provided instructions as part of the sign-up procedure for how to use private browsing mode. Our study was approved by our institutions's IRB.

	Min	25%	50%	75%	Max
Password Entries per day	0.4	1.6	2.5	3.9	33
Unique passwords entered	2	8	12	17	58
Unique correct passwords	1	4	6	8	18
Average password entropy	35	46	49	57	83
Length of Passwords	6.0	8.0	8.7	9.8	15
Websites with password	5	12	16.5	22	67
Website-to-Password Ratio	1.0	2.3	3.0	4.1	18
Frequency of Password Entry	1.2	2.1	2.7	3.6	37
Uses Password Manager			Yes: 26 — No: 108		

Table 2: Summary statistics about per-subject password usage. The 50% column contains the value for the median user; the 25% and 75% columns contain the first quartile and the third quartile users, respectively. "Frequency of Password Entry" is the average number of times a password is entered into each website.

4. RESULTS

4.1 Description of Password Use

Our dataset allows us to have a fairly comprehensive view of how each subject uses passwords on the web on a daily basis, over a number of consecutive weeks. We were able to capture every time a subject entered a password into a web page, and associate that with a specific browser and the user of that web browser. Our subjects visited an average of 5,613 web pages during the study (SD = 5,002), which translates to an average of 118 web pages per day (SD = 104). The median user entered a password into a web page 128 times over their participation in our study, though often they entered passwords into the same web pages on different days, or multiple times in a single day. Subjects ranged from a minimum of 22 password entries to a maximum of 1,474 entries, though most fell between 78 and 158 entries. The median user entered a password on 70% of the days they participated in the study, for an average of 3.2 passwords entered per day (SD = 3.5).

Our subjects used a median of 12 distinct passwords, though the number of passwords per subject varied quite a bit. On the low end, one subject entered only 2 distinct passwords (into 11 different websites). On the high end, another subject entered 58 different passwords over the study period, though most subjects ranged between 8 and 17 distinct passwords. This is not very many different passwords, given how frequently subjects needed to enter a password into a web page.

We grouped web pages into websites by domain name. Subjects entered passwords into a median of 17.5 different websites. They entered passwords into as few as 5 different websites, and into as many as 69, though most ranged between 12 and 19 different websites. As these numbers are higher than the number of distinct passwords, it is clear that our subjects tend to re-use the same passwords across multiple websites. One hundred fourteen of our subjects (85%) had fewer unique passwords than they did websites that they entered passwords into.

4.1.1 Likely Correct Passwords

At times, some users will enter more than one password into a website. This may be because they entered a typo or they forgot their correct password and are guessing passwords

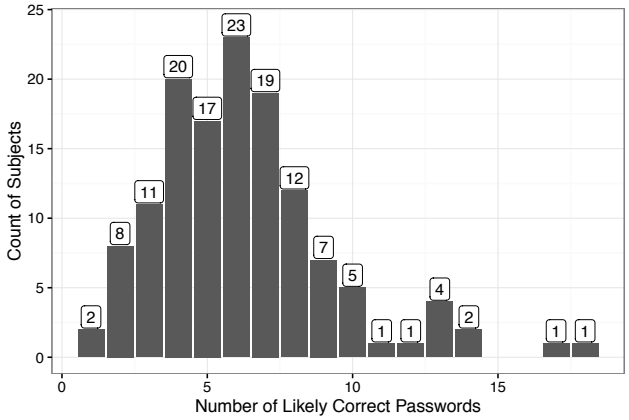

Figure 1: Histogram showing the number of passwords used by subjects in our sample.

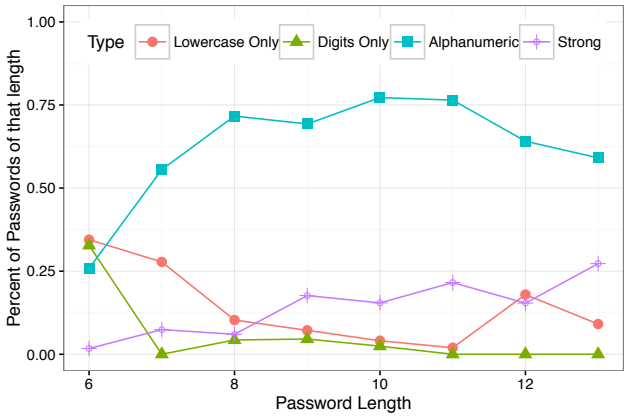

Figure 2: The type of password used, by length of password. Unlike in Florêncio and Herley [12], a clear majority of passwords we observed are alphanumeric passwords. We also observed significantly more strong passwords.

until they are able to successfully authenticate. It could be because they confused the password for one site with another site. It could be that the user has multiple accounts on the same website, or that they changed their password during the period of the study. Or it could be because they are guessing a friend's password for that website.

From our log data, we cannot tell which password was the correct password for a given website account. However, we can make an attempt to identify which password was *likely correct* based on usage patterns over time. We used a three step process for identifying which password is likely correct for a user on a given website:

1. The password that was entered most frequently into a given website is likely to be the correct one.
2. For websites where more than one password was frequently used, choose the password that was used on the larger number of days.
3. If there is still a tie (8% of websites), then choose the password that was used on the largest number of other websites by that user (the *Re-Use Assumption*).

This process successfully identified a likely correct password for 98% of websites. Most websites were fairly easy to choose a likely correct password; for example, one subject used 4 different passwords to log into his most frequently used website—3 were used once each, while the fourth was entered 96 times. Our subjects had a median of 6 likely correct passwords. Two subjects used only 1 likely correct password (which were correct on 10 and 18 different websites); one of our subjects correctly entered 18 different passwords over the study period, though most subjects ranged between 4 and 8 likely correct passwords.

4.1.2 Password Strength

For privacy reasons, we did not directly collect subjects' passwords. Instead, our data collection software calculated a standard *entropy* measure for each password before hashing the password and recording the hash. In our analysis, we use entropy not as a precise measure of how resistant a given password is to compromise, but to compare passwords created and used by the same person, across people, and across websites. Entropy allows us to roughly describe the complexity of a password, and as a result, how hard it might be for the user to remember. This is similar to how entropy has been used in several other studies of passwords. For example, Fahl et al. [11] used it to characterize relative differences between multiple passwords created by the same user. Florêncio, Herley, and van Oorshot [15] refer to entropy as a way to "represent user effort to remember a password". Egelman et al. [10] use entropy to quantify differences between groups of passwords created by participants in different conditions of their experiment.[2]

Averaging across all passwords that any of our subjects ever entered into a website, the average entropy is 49.2 bits (*SD* = 22.1). Passwords ranged in entropy from 4.322 bits (for a password consisting of a single symbol) up to 165.438 bits (for a 32 character alphanumeric password). These numbers, however, include all passwords that were entered into a website, including incorrect passwords and password guesses. If we only consider passwords that we identified as likely correct, then the average entropy across our sample is 49.5 (*SD* = 18.1). (The range is the same, as both the strongest and weakest passwords in our sample were likely correct on at least one website.) Subjects' strongest likely correct password had a median entropy of 65.5 bits, and the interquartile range was 53.6 bits to 82.7 bits.

4.1.3 Password Characteristics

Following Florêncio and Herley [12], we reverse engineered characters of passwords from the recorded entropy value. Given only a password's entropy and the knowledge of how it was calculated it is possible to reconstruct information about the password without ever knowing exactly what the

[2]Since we collected our data, Melicher et al. demonstrated a new technique based on deep learning to approximate the guessability of a single password in real-time [25]. We plan to use this in addition to entropy in future work. https://github.com/cupslab/neural_network_cracking

password was. A password's entropy was calculated by computing $entropy = \log_2 set\_size^{length}$. Rearranging, we see that $length = \frac{entropy}{\log_2 set\_size}$. For each possible size of character set, we can calculate an estimated password length. The correct set size and length is the one where the estimated password length is a whole number. This method does leave us with some possible limitations; it does not provide a way to differentiate between using lower case or upper case letters because they share the same character set size.

The average subject used passwords of length 8.98 (SD = 1.43) that used 2.29 (SD = 0.376) different character sets (from the set {Lowercase letters, Uppercase letter, Numbers, Basic Symbols, Extended Symbols}). Approximately 87% of passwords included a letter, 80% of passwords included a digit (number), and 14% included a symbol. Florêncio and Herley [12] found that the vast majority of the passwords that they observed were solely lowercase letters, with PINs (passwords consisting solely of numbers) the second most common; this is summarized in their Figure 9. Reproducing their Figure 9 using our dataset (Figure 2), we find that our subjects frequently used more complex passwords; the majority of our subjects use alphanumeric passwords, with "strong" passwords the second most common.

4.2 What passwords do people re-use?

Our median subject entered their passwords into 16.5 websites, and entered 12 distinct passwords into those websites. Overall, 31% of all passwords were entered into more than one website, and 20% of passwords were likely correct on more than one website.

Since the number of websites is larger than the number of passwords, this indicates that subjects re-used their passwords. We use our data to quantify password re-use: for each subject we calculate a website-to-password ratio — how many different websites on average each password is entered into by each user. A website-to-password ratio of 1.0 means that each website gets a different password. A website-to-password ratio greater than 1.0 suggests password re-use: passwords are entered into more than one website. A website-to-password ratio less than 1.0 happens when people change their passwords or enter incorrect passwords, thus entering multiple different passwords into a single website.

If we calculate a website-to-password ratio for each subject and then average them across subjects, we find that each password is entered into a median of 1.6 different websites. This number is likely a lower bound estimate for the true median website-to-password ratio, because it includes a number of incorrect passwords. If instead we only count passwords that we deemed to be *likely correct* for at least one website, then we find that the median subject in our sample entered a likely correct password into 3.0 different websites. This is because once incorrect passwords are removed, there remain a median of 6 correct passwords per subject.

In identifying which password was likely to be correct, we made a *re-use* assumption: among passwords used equally often, the one that was used on the most other websites was most likely correct. This assumption affected about 6% of user/website pairs, and it biases our re-use estimates toward higher levels of re-use. Thus, this second re-use estimate (each password is used on 3 websites) is likely an upper bound on the estimate for the true website-to-password ratio for the sample.

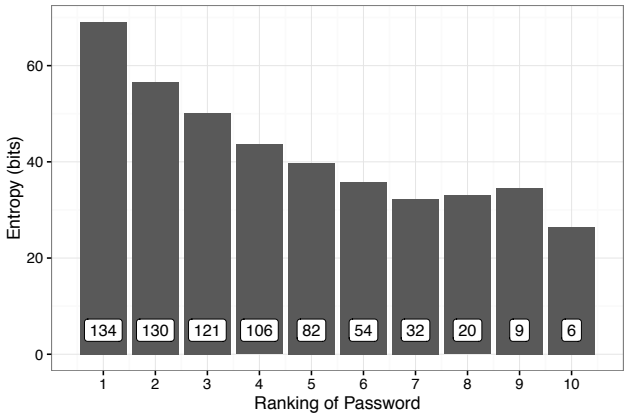

Figure 3: Average entropy for passwords at a given rank. The number near the bottom of each bar is the number of subjects with passwords at that rank.

4.2.1 People re-use strong passwords

Weaker passwords are easier to remember and to type, and most websites that require a password are not high-importance websites with complex password composition policies. People might therefore re-use their weaker passwords. On the other hand, they might re-use stronger passwords; memorizing a strong password takes more work so users might want to get the most out of that effort by re-using it wherever possible.

Among our subjects, people re-used their stronger passwords. There was a 0.063 correlation between the entropy of a password and the number of websites that password was entered into ($p = 0.007$). This positive, statistically significant correlation suggests that a subject's stronger passwords are the ones being re-used, though the small size of the correlation means that password strength isn't the whole story.

To better understand password re-use, we ranked all of the passwords that each subject used during our study by the strength (entropy) of that password. The password ranked #1 is the strongest password that subject ever entered, #2 is the second strongest, and so on down to password #N, the sybject's weakest password. This ranking is an individual ranking per subject, and allows us to examine whether it is the absolute strength of a password that influences re-use, or whether it is the strength relative to the subject's other passwords that matters. Figure 3 shows the average entropy of a password by individual ranking.

Putting both password entropy and password ranking into the same regression allows us to separately estimate the effects: for different passwords with the same entropy, does having a better ranking (lower number) lead to more re-use? And likewise, for password ranked the same, is the stronger one more likely to be re-used? However, password entropy and password ranking are highly correlated, (r=-0.628). Including both predictors in the same regression model can lead to collinearity issues. To address this, we ran three separate multi-level linear regression models (Table 3): one

	Entropy		Ranking		Both	
(Intercept)	1.82	***	2.70	***	2.61	***
Entropy	0.01	**			0.00	
Ranking			-0.06	***	-0.06	***
$R^2_{LMM_c}$	0.0040		0.0086		0.0086	

Table 3: Three multi-level linear models that analyze the effect of password strength on how many websites a password is re-used on. Each regression includes a random effect control for subject.

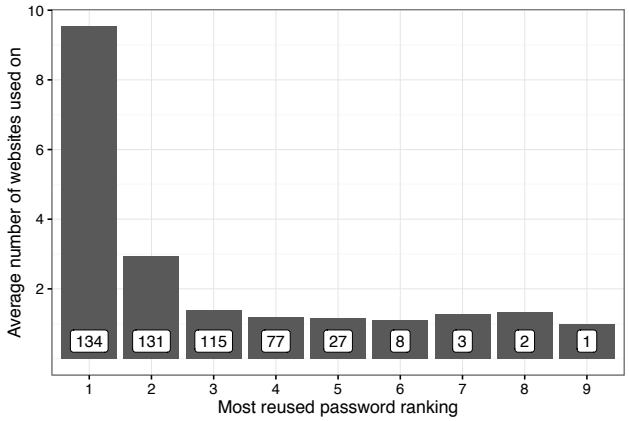

Figure 4: How often passwords are re-used. The leftmost bar shows the average for a subject's most-reused password; the second bar the second-most-reused password; and so on. The number near the bottom of each bar shows the number of subjects with passwords at that rank.

model with entropy, one model with ranking, and a model with both. By comparing the R^2 value for each model[3], we can see that the personal ranking is a better predictor of password re-use than absolute entropy, and indeed personal ranking explains almost all of the variance that both variables together explain.[4] In addition, Figure 4 shows that a subject's most re-used password is used far more often than any of that subject's other passwords. Thus, we conclude that it is not the absolute strength of a password that leads to re-use. People are re-using *their* strongest passwords, but not necessarily passwords that are objectively strong.

4.2.2 People re-use frequently entered passwords

Another possibility is that people are re-using passwords that they have to enter frequently. It is easier to remember a password if you have to enter it on a regular basis [5], and thus, passwords that need to be entered frequently might be easier to remember, and therefore easier to use. However, password re-use is correlated with the number of websites a password is entered into; passwords that are re-used on more websites will also naturally need to be entered more frequently. Instead of overall frequency, we looked at the number of times a password was entered, and divided it by the number of websites the password was used on to get a measure of the average number of times a password is entered into any single website. An average password from a median user was entered into each website it was used on 2.7 times.

[3]The R^2 measure for linear mixed models is the conditional R^2 for the whole model, $R^2_{LMM_c}$ from Nakagawa and Schielzeth [26].

[4]These R^2 numbers are are fairly low, which suggests that neither of these variables has much explanatory power. We just use these regressions to draw relative comparisons between the entropy and ranking variables to identify which predictor to include in future regressions. Our other regressions that we use to draw more substantive conclusions have more appropriate R^2 values.

	# Websites Correct	Password Re-used?	Non-univ Websites
(Intercept)	1.39 ***	-1.07 ***	0.87 ***
Ranking	-0.05 ***	-0.04 **	-0.01
Entry Frequency	0.11 ***	0.18 ***	-0.00
Uses Password Mgr.	0.03	0.00	0.01
University Password			3.20 ***
$R^2_{GLMM_c}$	0.069	0.320	0.124

Table 4: Multi-level regressions predicting password re-use. The left column is a linear regression where the DV is the number of different websites that a password will be re-used on. The center column is a logistic regression estimating the probability of a *likely correct* password being re-used. The right column is a linear regression where the DV is the number of *non-university* websites that a password will be re-used on. Each regression includes a random effect control for subject.

Using this measure of password *entry frequency*, we ran additional regressions to examine whether subjects re-used frequently-entered passwords (Table 4). We found that more frequently entered passwords are more likely to be re-used—much more likely. For every 9 times a password is entered into a website, that password is used on one additional website. Figure 5 shows a graphical representation of the probability that a password is reused that is generated by the logistic regression in the middle column of Table 4. The frequency that a password is entered (the x-axis) has a much larger effect on reuse than the password strength (the difference between lines). Also, the coefficient on Ranking is very similar in both in Table 4 and Table 3. This suggests that relative password strength within an individual and entry frequency are separate effects: people re-use their stronger passwords, and they also re-use passwords that they enter frequently into websites.

4.2.3 People re-use university passwords

One feature that all of our subjects have in common is that they all have accounts at the same university that they use on a regular basis for accessing email and other university services. This university has a password composition policy (passwords must be at least eight alphanumeric characters long) but does not require users to regularly change their password. It is possible that this commonality across subjects in our sample explains our results: our subjects use a strong password for their university accounts, and have to log into multiple university services frequently, so this is a natural password for them to re-use. However, we can

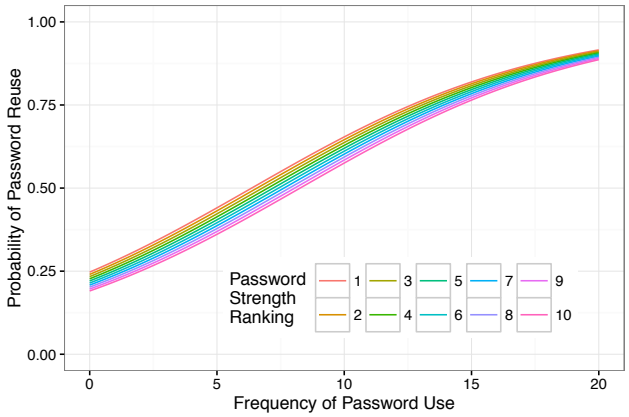

Figure 5: Predicted probability of re-using a password. There is a separate line for each strength of password: a password ranked #1 strongest for that subject, ranked #2 for that subject, etc.. How frequently a password is entered is a more important influence on password reuse than password strength.

test this. We identified all university websites and which password was the subject's likely correct password for those websites. Thirty-five subjects (23.1%) had their university password as their strongest password. Sixty-seven subjects (46.3%) had their university password as their strongest likely correct password. Also, 106 (79%) entered their university password more frequently than any other password.

To understand re-use across non-university accounts, we calculated a dependent variable consisting of the number of non-university websites where each likely correct password was entered. The rightmost column in Table 4 summarizes these results. The university password was heavily re-used across non-university websites; on average across all of our subjects, it was used on 3.2 additional non-university websites. Since for many of our subjects this was one of their strongest passwords, this means that they were using a relatively strong password (which is good), but were re-using a very high-value password (which is bad). The university strongly recommends against doing this; the first piece of advice on their password webpage says "Don't use your [university] ID and password for non-[university] accounts." [5]

4.2.4 *Password managers don't affect re-use*

In understanding re-use, one important consideration is password managers. Using a password manager makes it easier for people to use different passwords on every site because the passwords don't need to be remembered; they are stored by the computer instead. All of our subjects had the potential to use a password manager because both Google Chrome and Mozilla Firefox will save passwords for websites as they are used. Due to API restrictions, we were not able to identify when a password was filled in by the browser's built-in password saving feature.

However, our browser plugin recorded all add-ons such as browser plugins and browser extensions that were installed and/or enabled on each subject's web browser during the study. Manually looking through this list, we found that 26 of our subjects (19%) had a browser-based password manager enabled during the study. We saw six different password managers in use; the most popular password manager was Norton Identity Safe (9 users), followed by SimplePass (7 users).

[5] https://secureit.msu.edu/passwords/index.html, retrieved May 28, 2016.

	# Websites Incorrect
(Intercept)	0.59 ***
Ranking	-0.02 ***
Frequency	-0.01 **
# Websites where correct	0.03 ***
University Password	0.03
$R^2_{LMM_c}$	0.148

Table 5: Mutli-level linear regression predicting the number of websites a password will be *incorrectly* entered into. Each regression includes a random effect control for user.

The regressions in Table 4 include a subject-level variable indicating whether that subject used a third party password manager. A password used by a subject running a password manager was used on about 0.02 more websites than an equivalent password used by a subject without a password-manager, and this difference is not statistically significant. Third-party password managers do not significantly reduce password re-use across websites. However, we cannot tell if this is because many of our subjects are using the password saving features built-in to web browsers (everyone is storing passwords using a different mechanism), or if the subjects with password managers simply aren't using them or aren't using them effectively.

4.2.5 *People guess passwords from their other accounts*

When people forget their password for a website, they often guess passwords that they know they have used. We can learn a lot about what passwords people *think* are appropriate for a website by looking at the password that they incorrectly guessed. When we identified *likely correct* passwords, we separately identified password entries that we are fairly certain are incorrect guesses. We labeled as incorrect any password that was only entered once on a website where other passwords were used more often. We also labeled as incorrect any password entered into a website less than half as often or on less than half as many days as the password we identified as correct. Subjects entered incorrect passwords on 20% of websites.

Table 5 shows the results from a multi-level linear regression predicting how many times a password would be incorrectly guessed. A password that is correctly used on many other websites is more likely to be guessed incorrectly also. Subjects entered their commonly used passwords even into accounts where they were incorrect. Also, higher ranked, and thus stronger, passwords are slightly more likely to be guessed incorrectly. This is further evidence that re-use of stronger passwords on multiple accounts is an intentional strategy. Interestingly, the university password is not more

	Strength		Re-Use	
(Intercept)	40.54	***	2.95	***
"Use good passwords"	1.78	*		
"I use different passwords"			-0.30	*
Uses Password Manager	-1.13		0.29	
University Password	8.54	***	4.68	***
$R^2_{LMM_c}$	0.119		0.226	

Table 6: Muli-level linear regressions looking at the connection between intentions and behavior. The regression on the left examines whether self-reported password strength intentions predict password entropy. The regression on the right examines whether self-reported password re-use predicts the number of websites each of that subject's passwords is used on. Each regression includes a random effect control for subject.

likely to be guessed than any other password after controlling for how often it is used.

4.3 Do people self-report password use accurately?

Self-report questions are typically framed in one of two ways: self-reported intentions (future-oriented), or self-reported actions (past-oriented). However, it isn't clear whether people's self-reports regarding passwords accurately reflect their actual behavior. In a meta-meta-analysis by Sheeran [32], self-reported intentions in general have a 0.6 correlation with behavior across a number of domains. This is a high correlation, which is good; it suggests self-report can be fairly accurate. But Sheeran also found that the strength of the correlation can also vary widely by circumstance, which is why it is important to examine self-report accuracy in different areas to see what people can self-report accurately and what people do not self-report accurately.

In our survey, we included two intention (future-oriented) questions that are directly related to passwords and comparable with our log data:

- Password Strength: "Use good passwords (good passwords include uppercase and lowercase letters, numbers, and symbols)." [Scored Never (1) to Always (5), $M = 4.09$, $SD = 0.92$].
- Password Re-Use: "I use different passwords for different accounts that I have." [Scored Never (1) to Always (5), $M = 2.97$, $SD = 0.96$].

The first question is part of Wash and Rader's protection behaviors scale [39], and directly asks about subject intentions for password strength. The second question is part of the SeBIS behavioral intentions scale [9], and directly asks about subject intentions for password re-use. Twelve subjects did not provide an answer to the SeBIS question; those subjects have been removed from these analyses.

4.3.1 People understand password strength

Our subjects appear to be able to approximately self-report their intentions for using strong passwords. There is a 0.19 correlation between a subject's intentions to use strong passwords and the average entropy of the passwords that person entered during our study ($p = 0.027$). This statistically significant correlation is relatively small for an intention/behavior correlation, but it suggests that people do have some understanding of whether they are choosing stronger passwords.

	Entered Passwords	Correct Passwords
Strongest	0.12	0.11
Weakest	0.07	0.14
Avg by Password	0.19 *	0.19 *
Avg by Website	0.23 **	0.25 **
Avg by Use	0.16 .	0.15 .

Table 7: How well each measure of password strength correlates with a subject's self-reported intention to "Use good passwords". Each number is a Pearson correlation between the self-report measure "Use good passwords" and the indicated password or average of a set of passwords.

Table 6 contains more detailed regression results for the intention/behavior link. The left column uses a multi-level regression to predict a password's entropy using the subject's answer to the self-report survey question about password strength, while controlling for other differences across subjects. On average, a subject that chooses one higher answer on the scale from 'Never' (1) to 'Always' (5) will have passwords that are approximately 1.8 bits stronger. This is approximately equivalent to taking an all-letter password and replacing one letter with a number. This is not a large effect; there is a lot of variation in password strength that is not explained by self-reported intentions. But it is statistically significantly greater than no effect. When people self-report that they intend to use good passwords, their passwords are stronger—but only slightly.

4.3.2 What do people self-report?

When people self-report whether they "Use good passwords", which passwords are they thinking about? They could be thinking about their strongest password when answering this question; alternatively, they could think about their weakest password and evaluate whether they think it is strong. They could imagine all the different passwords they've created and mentally average their strength. They could look at the different websites they have entered passwords on recently and average the strength of the passwords on those websites. Or they could think about all the different times that they had to enter a password, and average the strength of the passwords that they entered. Each of these is a slightly different way of operationalizing which password(s) a person is thinking of when self-reporting.

Our dataset allows us to explore these different interpretations. We can look at different ways of aggregating a subject's passwords and see which aggregation most strongly correlates with a subject's self-reported intention to use strong passwords. Table 7 reports these correlations. The first row examines whether self-reported intention to use strong passwords is correlated with the strongest password each subject

has. The second row looks at the correlation with the weakest password, which is a measure of how strong *all* passwords are. The third row is the correlation with the average entropy if the subject thinks about each distinct password separately. The fourth row represents the correlation with the average strength of the passwords used on each different account. The last row is the correlation if the subject thinks about each time they enter a password and how strong that password is.

Comparing these correlations shows that subjects are not thinking about specific passwords as the strongest or weakest password; instead, when subjects answered this question they likely were thinking across all of the websites that they have accounts on, and looking at the average strength of those passwords. We suspect that when answering this survey question, subjects thought of each website as having a separate password (and thus, a separate choice for that website's password strength), even if he or she re-uses a password on multiple accounts.

4.3.3 People also understand password re-use

Our subjects also seemed to be able to self-report password re-use somewhat accurately. The correlation between a subject's self-reported intention to re-use passwords and their actual re-use (as measured by the ratio of websites-per-correct-password) is -0.12 ($p = 0.18$). This correlation is negative, which is the expected direction; subjects who self-report stronger intentions to use different passwords use each password on *fewer* websites.

Controlling for differences across subjects, we find similar results (Table 6, second column). Using a multi-level linear regression we find that on average, a subject who chooses one level higher on the scale from 'Never' (1) to 'Always' (5) for their intention to use different passwords will use each of their passwords on 0.3 fewer websites. This indicates that a greater intention to use unique passwords is related to less actual password re-use. Though, as with password strength, there is still a lot of unexplained variance.

4.4 Limitations

Our study has several limitations. Our subjects are undergraduate and graduate students at a large midwestern university, and all were from the same university; therefore, our results may not generalize to a wider population. For example, older users tend to select stronger passwords [3]. In addition, the specifics of the university's password composition policy and enforcement of frequent authentication are undoubtedly factors contributing to our results. However, our findings regarding the number of unique passwords and the amount of password re-use are in the same general range as other password studies.

We potentially do not capture all password entry events, either when subjects used private browsing mode, or because a website didn't use a recognized HTML form element. During development, we tested many websites and included special-case code to detect a variety of password forms. We capture password behavior for the majority of websites; for example, we have good data from at least 97 of the 100 most frequently visited websites in our dataset. For ethical reasons, we allowed users to disable data collection which may mean our results do not apply to sensitive online activity.

In addition, our data collection method does not allow us to differentiate between successful and unsuccessful authentication attempts. In other words, we do not know what the true correct password is for any of our subjects' website accounts. This also means that we can't tell if or when our subjects may have changed any of their passwords during the study period. Finally, approximately six weeks of data collection is not enough longitudinal data to make causal claims about these phenomena based on timing or sequence of events, and may have missed passwords entered less often than every six weeks.

5. DISCUSSION AND CONCLUSION

From prior literature, we know that people say they re-use passwords to reduce the difficulty of remembering too many passwords [27]. A median subject in our study used 6 unique passwords that we identified as *likely correct* for the websites they were entered on. While the median password was used on 3 websites, each subject's most re-used password was used on an average of 9 different websites (Figure 4). Subjects tend to re-use passwords that they have to enter frequently, and those passwords tend to be among the user's strongest passwords. In addition, likely correct passwords were also more likely to be entered incorrectly on other accounts, indicating that when subjects attempted to authenticate they naturally tried their "go-to" passwords.

Many studies that have examined password re-use have found that users have a similar number of distinct passwords that they re-use across their websites. Florêncio and Herley [12] found that users averaged 6.5 distinct passwords. Fahl et al. [11] found that people used between 2 and 5 passwords for most of their online accounts. Gaw and Felton's [16] subjects used an average of 3.31 distinct passwords. Stobert and Biddle's [33] subjects reported having between 2 and 20 unique passwords, with a median of 5 passwords. Rinn et al. [29] reported low-literacy subjects used between 1 and 9 unique passwords, with a median of 4. And our subjects mostly used between 4 and 8 passwords with a median of 6. This suggests that there may be a practical constraint that is a hard limit on the number of passwords that most people can remember.

Memorizing strong passwords is difficult for most users to do. Bonneau and Schechter [5] were able to influence 94% of their subjects to memorize a randomly generated 56-bit password by asking them to repeatedly log in 90 times over a period of two weeks with some clever interface manipulations. Logging in with a password frequently is an effective means of memorizing strong passwords. Florêncio and Herley [13] suggest that organizations for which there are no alternatives, such as one's bank, employer, or university, tend to have stronger password composition policies and require users to authenticate more often than websites where use is voluntary (e.g. social media, news websites). These organizations may be helping users memorize stronger passwords by forcing them to choose a long, complex password and enter it frequently. Once memorized, that password can then be re-used elsewhere. This may be what happened in our dataset: the university our subjects are associated with requires fairly strong passwords, and also requires users to enter them frequently.

Among our sample of non-technical users, how frequently a user had to enter a password was one of the strongest

predictors of password re-use. We suspect that once they had a strong password memorized, it was easier to use that password on other websites. This points to an unexpected interdependence between accounts: if users must memorize a strong password on a website where they have to enter it frequently, they then re-use it elsewhere. This results in stronger passwords on more websites. While this practice puts users at greater risk of cross-site password guessing attacks, it helps prevent within-site password guessing by spreading stronger passwords rather than weaker passwords. Since most non-expert users believe that password strength is more important than password re-use [21], it makes sense for them to adopt this strategy. This is evidence that users are trying to adapt their password practices to the security advice they are being told—"Use strong passwords as much as you can."

While people seem to have a mental model of what "stronger" passwords look like [35], our subjects' intentions for using strong passwords and choosing different passwords for each account were only weakly correlated with behavioral measures of password strength and re-use. Responses about password strength intentions were most correlated with the average entropy of the passwords used on each different account, indicating that when people think about what using strong passwords means, each account is considered separately and re-use is not considered.

Bonneau et al. [4] show that entropy is a poor measure of password strength. However, the weak correlation of password entropy with self-reported behavior suggests that when thinking about strong passwords, people think about something similar to complexity (which is what entropy measures).

Our results suggest that asking users about how well they adhere to common password advice from experts asks about password behavior in a way that does not approximate how people actually behave. The ideal situation for security experts would be no re-use: unique, random passwords for every account [7]. Expert advice tends to treat passwords as a black-and-white issue; anything less than the ideal introduces unacceptable vulnerabilities. When the ideal is used as the benchmark, it fails to reflect the reality behind users' choices, and our results speak to the size of the gap between the ideal for security and the realities users face.

Our results show that some amount of re-use might actually be good from a cost/benefit perspective, because if users have a few fairly strong passwords that they use on appropriate categories of sites (e.g., don't use the strong, high-value password on a weaker category of site [14]), they may be more secure than if they used weak passwords everywhere [15]. If the (stronger) university password is used appropriately, then this re-use pattern could lead to a positive effect on overall security. This presents an opportunity for organizations with the ability to force system use (e.g. large employers or universities) to help users memorize stronger passwords by requiring strong passwords and frequent re-authentication. Password composition policies [23] and feedback from password meters [10] can cause people to create stronger passwords than they would otherwise. This might help people use stronger passwords, and is often considered a good security practice by many organizations. However, this practice also puts the organization at greater risk; if the password is re-used on a site with lower security (which is an optimal strategy for some types of websites [28]), an attacker can learn the user's password and use it to compromise the organization. While forcing re-authentication solves one security problem, it creates another: it encourages re-use of the organization's password. Practically speaking, sites that are likely to be compromised need the strongest passwords so they can withstand an offline guessing attack, but users shouldn't have to spend their limited memory capacity and effort creating very strong passwords for sites that are unlikely to be compromised [14].

Unfortunately, defining appropriate categories of websites for re-use of passwords of varying strengths is an open area of research; should it be defined by how much the user values the information [14] or how much an attacker stands to gain [1], or by how much the website invests in security [28]? There isn't a consensus about this, and it seems to be an area of disagreement among researchers. Our study provides insight into how the human and technical constraints imposed on users shape their password choices and behaviors over time, which highlights additional constraints to consider: relative password strength within an individual, and how often the password must be used.

6. ACKNOWLEDGMENTS

We thank Kami Vaniea, Tyler Olson, Nick Saxton, Nathan Klein, Raymond Heldt, Ruchira Ramani, Jallal Elhazzat, Tim Hasselbeck, Shiwani Bisth, Robert Plant Pinto Santos, Meghan Huynh, and Simone Merendi for assistance in developing the software and analyzing the data. This material is based upon work supported by the National Science Foundation under Grant Nos. CNS-1116544 and CNS-1115926.

7. REFERENCES

[1] D. V. Bailey, M. Dürmuth, and C. Paar. Statistics on Password Re-use and Adaptive Strength for Financial Accounts. In *Proceedings of the 9th Conference on Security and Cryptography for Networks (SCN)*, pages 218–235, 2014.

[2] A. Beautement, M. A. Sasse, and M. Wonham. The compliance budget: managing security behaviour in organisations. In *Proceedings of the 2008 workshop on New Security Paradigms Workshop (NSPW)*, 2008.

[3] J. Bonneau. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *Proceedings of the IEEE Symposium on Security and Privacy (SP)*, pages 538–552, 2012.

[4] J. Bonneau, C. Herley, P. C. van Oorschot, and F. Stajano. Passwords and the evolution of imperfect authentication. *Communications of the ACM*, 58(7):78–87, June 2015.

[5] J. Bonneau and S. Schechter. Towards reliable storage of 56-bit secrets in human memory. In *Proceedings of the 23rd USENIX Security Symposium*, 2014.

[6] S. Chiasson, A. Forget, E. Stobert, P. C. van Oorschot, and R. Biddle. Multiple password interference in text passwords and click-based graphical passwords. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*, pages 500–511, 2009.

[7] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang. The Tangled Web of Password Reuse. In

Proceedings of the 2014 Network and Distributed System Security Symposium (NDSS), 2014.

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

[9] S. Egelman and E. Peer. Scaling the security wall: Developing a security behavior intentions scale. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 2873–2882, 2015.

[10] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley. Does my password go up to eleven? The impact of password meters on password selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 2379–2388, 2013.

[11] S. Fahl, M. Harbach, Y. Acar, and M. Smith. On the ecological validity of a password study. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2013.

[12] D. Florêncio and C. Herley. A large-scale study of web password habits. In *Proceedings of the 16th International Conference on World Wide Web (WWW)*, pages 657–666, 2007.

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

[14] D. Florêncio, C. Herley, and P. 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.

[15] D. Florêncio, C. Herley, and P. 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.

[16] S. Gaw and E. W. Felten. Password management strategies for online accounts. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, pages 44–55, 2006.

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

[18] S. M. T. Haque, M. Wright, and S. 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.

[19] E. Hayashi and J. Hong. A diary study of password usage in daily life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 2627–2630, 2011.

[20] P. G. Inglesant and M. A. Sasse. The true cost of unusable password policies: password use in the wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 383–392, 2010.

[21] I. Ion, R. Reeder, and S. Consolvo. "... no one can hack my mind": Comparing Expert and Non-Expert Security Practices. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2015.

[22] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez. Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms. In *Proceedings of the IEEE Symposium on Security and Privacy (SP)*, pages 523–537, 2012.

[23] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. 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.

[24] M. L. Mazurek, S. Komanduri, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, P. G. Kelley, R. Shay, and B. Ur. Measuring password guessability for an entire university. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*, pages 173–186, 2013.

[25] W. Melicher, B. Ur, S. M. Segreti, S. Komanduri, L. Bauer, N. Christin, and L. F. Cranor. Fast, lean and accurate: Modeling password guessability using neural networks. In *Proceedings of USENIX Security*, 2016.

[26] Nakagawa and Schielzeth. A general and simple method for obtaining r^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142, 2012.

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

[28] S. Preibusch and J. Bonneau. The Password Game: Negative Externalities from Weak Password Practices. In *Proceedings of the Conference on Decision and Game Theory for Security (GameSec)*, pages 192–207. 2010.

[29] C. Rinn, K. Summers, E. Rhodes, J. Virothaisakun, and D. Chisnell. Password Creation Strategies Across High- and Low- Literacy Web Users. In *Proceedings of the 78th ASIS&T Annual Meeting (ASIST)*, 2015.

[30] M. A. Sasse, M. Steves, K. Krol, and D. 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.

[31] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, and L. F. Cranor. Encountering stronger password requirements: user attitudes and behaviors. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2010.

[32] P. Sheeran. Intention-behaviour relations: A conceptual and empirical review. *European Review of Social Psychology*, 12:1–36, 2002.

[33] E. Stobert and R. Biddle. The Password Life Cycle: User Behaviour in Managing Passwords. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, pages 243–255, 2014.

[34] L. Tam, M. Glassman, and M. Vandenwauver. The psychology of password management: a tradeoff between security and convenience. *Behaviour & Information Technology*, 29(3):233–244, 2010.

[35] B. Ur, J. Bees, S. M. Segreti, L. Bauer, N. Christin, and L. F. Cranor. Do Users' Perceptions of Password Security Match Reality? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2016.

[36] B. Ur, F. Noma, J. Bees, S. M. Segreti, R. Shay, L. Bauer, N. Christin, and L. F. Cranor. "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, pages 123–140, 2015.

[37] E. von Zezschwitz, A. De Luca, and H. 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.

[38] K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. B. Tai, J. Cook, and E. Eugene Schultz. Improving password security and memorability to protect personal and organizational information. *International Journal of Human-Computer Studies*, 65(8):744–757, 2007.

[39] R. Wash and E. Rader. Too much knowledge? security beliefs and protective behaviors among us internet users. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2015.

[40] R. Wash, E. Rader, K. Vaniea, and M. Rizor. Out of the Loop: How Automated Software Updates Cause Unintended Security Consequences. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, pages 89–104, 2014.