Do Users' Perceptions of Password Security Match Reality?

Blase Ur, Jonathan Bees[†], Sean M. Segreti, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor
Carnegie Mellon University, [†]The Pennsylvania State University
{bur, ssegreti, lbauer, nicolasc, lorrie}@cmu.edu, [†]jfb5406@psu.edu

ABSTRACT

Although many users create predictable passwords, the extent to which users realize these passwords are predictable is not well understood. We investigate the relationship between users' perceptions of the strength of specific passwords and their actual strength. In this 165-participant online study, we ask participants to rate the comparative security of carefully juxtaposed pairs of passwords, as well as the security and memorability of both existing passwords and common password-creation strategies. Participants had serious misconceptions about the impact of basing passwords on common phrases and including digits and keyboard patterns in passwords. However, in most other cases, participants' perceptions of what characteristics make a password secure were consistent with the performance of current password-cracking tools. We find large variance in participants' understanding of how passwords may be attacked, potentially explaining why users nonetheless make predictable passwords. We conclude with design directions for helping users make better passwords.

Author Keywords

User behavior; perceptions of security; passwords; authentication; users' folk models; usable security

ACM Classification Keywords

H.5.m Information Interfaces and Presentation (e.g., HCI): Miscellaneous; K.6.5 Security and Protection: Authentication

INTRODUCTION

For better or worse, passwords remain today's dominant form of user authentication [11]. While the predictability of user-chosen passwords has been widely documented [9, 37, 51, 72, 74, 77, 80], very little research has investigated users' perceptions of password security. That is, do users realize they are selecting terrible passwords and choose to do so intentionally, or are they unwittingly creating weak passwords when they believe they are making secure ones?

In this paper, we report on a 165-participant study of users' perceptions of password security. Participants provided their perceptions about the security and memorability of passwords chosen to exhibit particular characteristics, as well as common strategies for password creation and management. We compare participants' perceptions to the passwords' actual resilience to a variety of large-scale password-guessing attacks.

In the first of four tasks, we showed participants 25 pairs of passwords differing in specific characteristics (e.g., appending a digit, as opposed to a letter, to the end of the password). We asked participants to rate which password was more secure, if any, and to justify their rating in free text. In the second and third tasks, we showed participants a selection of passwords from the well-studied breach of the website RockYou [72], as well as descriptions of common password-creation strategies. We asked participants to rate both the security and the memorability of each password or strategy. In the fourth task, we had participants articulate their model of password attackers and their expectations for how attackers try to guess passwords.

We observed some serious misconceptions about password security. Many participants overestimated the benefits of including digits, as opposed to other characters, in a password. Many participants also underestimated the poor security properties of building a password around common keyboard patterns and common phrases. In most other cases, however, participants' perceptions of what characteristics make a password more secure matched the performance of today's password-cracking tools. This result calls into question why users often fail to follow their (correct) understanding when crafting passwords. However, most participants displayed an unrealistic mental model of attackers, which may prevent them from fully accounting for the actual spectrum of threats to their passwords.

Although much has been written about text passwords in recent years, our study is the first to focus specifically on users' perceptions of security. The main outcome of our work is to inform design directions for helping users both make stronger passwords and better understand the implications of their password-creation decisions.

BACKGROUND AND RELATED WORK

We summarize related work examining users' perceptions of security and discuss the most closely related studies on passwords. We then discuss the actual threats to password security and approaches to measuring password strength.

Users' Perceptions of Security

Hundreds of research studies have been conducted at the general intersection of usability and security [28], but few have specifically investigated users' *perceptions* of security. One stream of qualitative work has examined users' mental models [3, 76] and "folk models" [57, 78] of security, finding that

CHI'16, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-3362-7/16/05.
http://dx.doi.org/10.1145/2858036.2858546

non-expert users' mental models often differ from those of experts. For instance, non-experts perceive losing a password as similar to losing a key, whereas experts perceive the same event as more akin to losing a credit-card number [3]. Likewise, a study that examined users' perceptions of the most important computer-security practices [39] again showed a disconnect between non-technical users and experts. Notably, users' perceptions of the efficacy of different security practices impact the adoption of security technologies [7,20,38,65].

In the password domain, a recent interview study of password creation using a think-aloud protocol implicitly rested on users' perceptions of password security [70]. Security perceptions, however, were not the focus of that study. Furthermore, that study was qualitative, whereas ours is mostly quantitative.

Most closely related to our work, a 384-participant study examined users' perceptions of the security and usability of Android graphical unlock patterns [4]. Participants compared two unlock patterns and assessed their relative security and memorability (e.g., whether the first pattern was more secure, but less memorable, than the second). Unlike our study, that study did not compare the actual security of these patterns.

Prior Studies of Passwords

The rich literature on passwords has documented many password characteristics. Many users make passwords that are quite predictable [50, 77] even for relatively important accounts [24,51]. Passwords are generally too short to provide much resistance to attacks [9,47,51,72], and users tend to put digits and symbols at the end of the password [9,69,70] and capital letters at the beginning [69,70].

Users tend to base passwords around predictable words and phrases, including names [37], dates [75], song lyrics [46,69], and other concepts or objects they like [70]. Furthermore, when a password contains multiple words, those words tend to be semantically related [13,74]. Keyboard patterns (e.g., "1qaz2wsx") are common [70, 74], and passwords sometimes contain character substitutions (e.g., replacing "a" with "@") [42]. These characteristics vary somewhat for passwords created on touchscreen devices [52]. We use this prior work to inform our hypotheses and selection of passwords to test.

Many users view passwords as a burden [21] and exhibit potentially insecure behaviors when managing passwords [29,35, 54,62]. However, many of these behaviors are likely rational coping strategies for users who are asked to make far more distinct, complex passwords than they could possibly remember [2,27,62]. A key coping mechanism is password reuse. Users often reuse passwords across accounts [18,25,41,77]. Even when they do not reuse a password verbatim, they frequently make only small, predictable modifications [18,70,82].

Although a typical user has created hundreds of passwords [25, 27], most of the feedback he or she has received about password strength comes from password-strength meters [23,68], whose estimates are often inaccurate [1,19,81]. While popular media highlights the most predictable passwords [37,72], comparatively less attention is paid to helping users create strong passwords. In fact, one of the most widely discussed examples of password-creation advice is actually a web cartoon [53]. While visualizations can help users understand password-guessing attacks [83] and some password-strength meters can give detailed feedback about predictability [45], neither approach has yet been widely adopted.

The Password Attack Ecosystem

The authentication ecosystem is vulnerable to a number of attacks. For each type, the security of a password has a different impact. In this section, we describe the most important threats.

Sometimes, the security of a password does not matter [12]. If a user is phished, the attacker gets the password in plaintext. In some other cases, it is most important that a password not be trivially predictable. In what is known as an online attack, the attacker attempts to authenticate to a running system using guesses of what the user's password might be. After a few incorrect guesses, often 3–10, best practices dictate the system rate-limit subsequent attempts or lock the account and require alternate authentication [26]. To be protected against an online attack, a password should not be among the million most common passwords [26], nor should it include the user's personal information (e.g., family member's name, birthdate). This chain of reasoning assumes that rate-limiting is properly implemented. As the recent Apple iCloud hack regrettably demonstrated [16,48], this is not always the case. In other words, even for online attacks, large-scale guessing may be possible if the system's implementation is flawed.

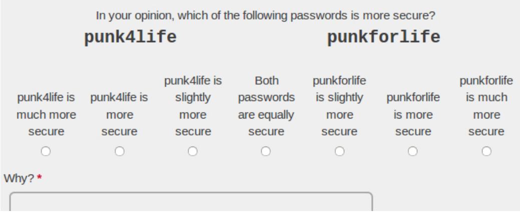
Another threat, an offline attack, usually involves large-scale guessing. Best practices dictate that systems store passwords hashed using a cryptographically secure (irreversible) one-way function. When a user authenticates, the system hashes the password submitted and verifies that it matches the value in its database. Sadly, numerous hashed password databases have been compromised in recent years [8,14,33,67].

A password's resistance to an offline attack depends on both the type of hash function used to store the password and the attacker's resources. Hash functions like MD5 were designed for efficiency, which makes them poor for storing passwords due to their speed. Modern hardware can try billions of MD5 guesses per second [61,64]. Unfortunately, numerous services [8, 14, 67] have hashed passwords with MD5. Best practices dictate the use of intentionally slow hash functions like bcrypt [56], for which attackers can only try hundreds of guesses per second [34,61]. The 2015 Ashley Madison breach [5] was the first major compromise involving bcrypt. Unfortunately, they retained a legacy database using MD5 [34], demonstrating that best practices are not always followed.

The main security threat of an offline attack derives from password reuse [12]. Once attackers have learned a particular user's password in an offline attack, they will try the same username and same password, or close variants [82], on other sites. Users often reuse passwords [18,41,70,77], which can cause serious harm. For instance, attackers recently infiltrated Mozilla's Bugzilla database because one Mozilla administrator had reused his password on another, compromised site [30].

Measuring Password Strength

In general, gauging the strength of an individual password is a complex and nuanced problem [10,31,71]. Users generally only encounter estimates of password strength provided by



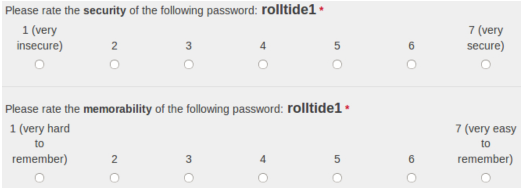
In your opinion, which of the following passwords is more secure?

punk4life **punkforlife**

punk4life is much more secure	punk4life is more secure	punk4life is slightly more secure	Both passwords are equally secure	punkforlife is slightly more secure	punkforlife is more secure	punkforlife is much more secure
○	○	○	○	○	○	○

Why? *

Figure 1. Example *password pair* comparison testing the hypothesis that substituting a digit for multiple letters will be perceived as more secure.



Please rate the **security** of the following password: **rolltide1** *

1 (very insecure)	2	3	4	5	6	7 (very secure)
○	○	○	○	○	○	○

Please rate the **memorability** of the following password: **rolltide1** *

1 (very hard to remember)	2	3	4	5	6	7 (very easy to remember)
○	○	○	○	○	○	○

Figure 2. An example task for rating the security and memorability in our *selected-password analysis*.

password meters. These meters are usually based on heuristics, such as length or the number of character classes used [19, 68], that frequently do not reflect the actual strength of a password [1, 81]. While some meters use more advanced heuristics [81], the correlation between these advanced heuristics and actual strength has not been studied scientifically.

More accurate password strength measurements can be obtained either through statistical methods [9, 10], which are primarily suitable for very large sets of passwords, or by parameterized password *guessability* [10, 43, 71]. Measuring guessability entails computing a *guess number* for each password indicating how many guesses a particular password-cracking approach configured and trained in a particular way would take to guess that password. Because guessability estimators can only run for finite time, there is necessarily a guess cutoff at which remaining passwords are labeled "unguessed."

In this work, we use the guessability metric because it provides strength estimates on a per-password basis for even small sets of passwords. In addition, it models adversarial password cracking, and our study focuses on users' perceptions of security against an adversary.

METHODOLOGY

We conducted an online study to gauge users' perceptions of password strength and memorability, as well as their understanding of attacker models. We then compared these perceptions to passwords' resistance to current large-scale attacks.

Study Structure

We structured the study in five parts designed to take 30 minutes total. The first part of the study asked about demographics, including age and gender. Because participants' perceptions would likely be influenced by their technical understanding of the password ecosystem, we also asked whether they were a "security professional or a student studying computer security," and whether they had a job or degree in a technology field.

In the second part of the study, which we term *password pairs*, we investigated 25 hypotheses about how different password characteristics impact perceptions of security. As shown in Figure 1, a participant saw two similar passwords that varied in a way dictated by the hypothesis. The participant rated the passwords on a 7-point, labeled scale denoting which password is more secure. In addition, we required a free-response justification for the rating.

We chose the 25 hypotheses (see Table 4 in the results section), to investigate eight broad categories of password characteristics inspired by prior work [13, 42, 70, 74, 75]: capitalization; the location of digits and symbols; the strength of letters vs. digits vs. symbols; the choice of words and phrases; the choice of digits; keyboard patterns; the use of personal information; and character substitutions. As an attention check, a 26th pair compared a password to itself. We randomized the order of the 26 pairs and left-right orientations of each pair per participant.

To reduce potential selection biases, we created three pairs of passwords for each of the 25 hypotheses. Each participant saw one of the three pairs, randomly selected. To create each pair, we first chose a password from the widely studied [47, 51, 79] dataset of 32 million passwords leaked in plaintext from the gaming website RockYou in 2009 [72]. In particular, we randomly permuted this set and selected the first password that could plausibly be tested as part of each hypothesis. Thus, at least one password in each pair is from the RockYou breach. For the second password in each pair, we either created the minimally different password to test the hypothesis (e.g., we created "astley123" to correspond to RockYou's "astleyabc") or selected a second RockYou password, as appropriate.

In the third part of the study, *selected-password analysis*, we investigated broader perceptions by asking participants to rate their opinion of the security and memorability of 20 passwords selected from the RockYou set [72]. As detailed below, we selected new passwords for each participant without replacement. As shown in Figure 2, participants used a 7-point scale to rate the security ("very insecure" to "very secure") and memorability ("very hard to remember" to "very easy to remember"); we labeled only the endpoints of the scale. We biased the selection of the passwords shown to each participant to include diverse characteristics. Ten passwords were selected randomly from among RockYou passwords matching each of the ten most common character-class structures. In addition, we selected one password containing at least three character classes, one password containing a symbol, two long passwords (12+ characters), and six additional passwords that do not fit any of the previous categories. We randomized the order in which we showed the passwords.

The fourth part of the study was similar to the third, except we instead asked about 11 strategies for password creation and password management. We chose common strategies from prior work on password creation [13, 46, 70, 74] and password management [18, 29, 62]. For example, one strategy we presented was creating a password "using a phrase taken from the lyrics to a song (e.g., somewhere over the rainbow)." We provide the full list of 11 strategies in the results section.

The fifth and final part of the study focused on participants' impressions and understanding of attackers who might try to guess their password. We intentionally presented this part of the study last to avoid priming participants as they evaluated password security in the rest of the study.

We asked seven questions about attackers. Participants wrote free-text responses to separate questions about "what characteristics make a password $\{$easy, hard$\}$ for an attacker to guess." Participants "describe[d] the type of attacker (or multiple types of attackers), if any, whom you worry might try to guess your password," and explained to the best of their knowledge *why* an attacker would try to guess their password, as well as *how* attackers try to do so. Finally, participants provided a numerical estimate of "how many guesses (by an attacker) would a password need to be able to withstand for you to consider it secure," as well as a free-text justification.

Recruitment

We recruited participants on Amazon's Mechanical Turk (MTurk) platform for a "research study about password security." While imperfect, MTurk can provide data of at least the same quality as methods traditionally used in research as long as the experiment is designed carefully [6, 15]. We limited participation in this study to MTurk users age 18 and older who live in the United States. We compensated participants $5 U.S. for the study, which took approximately 30 minutes.

To ensure quality MTurk data [40], we inserted the attention check described above. We only accepted data from participants who rated that pair as equal in strength and wrote a variant of "the passwords are the same" in their justification.

Measuring Real-World Attacks on Passwords

To understand how users' perceptions of password security correspond to actual threats, we calculate each password's guessability [9,10,43] by simulating attacks using modern password-cracking techniques. We use the Password Guessability Service (PGS) [17,71], a product of our group's prior evaluations of metrics for calculating password strength [43, 71].

In prior work, we showed that considering only one of the numerous approaches to password cracking can vastly underestimate the guessability of passwords with particular characteristics, while using a number of well-configured approaches in parallel can conservatively estimate password guessability against an expert attacker [71]. Thus, PGS simulates password-guessing attacks using Markov models [49], a probabilistic context-free grammar [43, 44, 80], and the software tools oclHashcat [61] and John the Ripper [55]. For each password, PGS conservatively outputs the smallest guess number across these four major password-cracking approaches. Evaluating guessability using several password-cracking approaches in parallel helps account for passwords that are modeled particularly well by some approaches, but not by others.

These approaches order their guesses based on training data, comprising sets of leaked passwords and natural-language dictionaries [17]. Furthermore, we configured the software tools to reflect behaviors common in the password-cracking community [71]. Thus, within the limitations of the training data and theoretical models of how humans craft passwords, the ordering of guesses is grounded in data. If the guess numbers are within an order of magnitude of each other, we judge the passwords to be of similar security. When we judge passwords to be of different security, their guess numbers differ by over an order of magnitude. These differences occur when some words or characteristics are far more common than others in the sets of real passwords used to train the tools.

In the results section, we frequently compare participants' perceptions of the relative security of passwords to the relative difference in guess numbers. Because the PGS guess numbers reflect the performance of current password-cracking approaches, we either state that participants' perceptions were *consistent* or *inconsistent* with current approaches.

Quantitative Analysis

We used different statistical tests for our quantitative analyses investigating, respectively, participants' strength ratings, the relationship between security and memorability, and the relationship between independent variables. For all tests, we set $\alpha = .05$. We corrected for multiple testing using the conservative Bonferroni method, which we applied per type of test (e.g., we multiplied p values by 75 for the 75 password pairs).

We treated participants' rating for each password pair $\{\mathrm{PW}_1$, $\mathrm{PW}_2\}$ as an ordinal rating from -3 to 3, where -3 indicates the perception that PW_1 is much stronger and 0 indicates that the passwords are equally strong. To test whether participants tended to rate one password in the pair as stronger than the other, we used the one-sample, two-sided Wilcoxon Signed-Rank test. This non-parametric test evaluates the null hypothesis that the true password rating is 0 (equally secure) and the alternative hypothesis that the true rating is non-zero (one password is perceived more secure than the other).

To investigate the relationship between security and memorability for the selected-password analysis and password-creation strategies, we calculated Spearman's rank correlation coefficient (Spearman's ρ), which is a nonparametric evaluation of the correlation between variables. The value for ρ varies between 1 (perfect correlation) and -1 (perfect inverse correlation), where 0 indicates no correlation.

For our selected-password analysis, we also used regression models to evaluate the relationship between numerous independent variables (e.g., password length, number of digits) and participants' ratings of password security and memorability. In particular, because participants' ratings were ordinal on a 7-point scale and because each participant rated 20 different passwords, we use a mixed-model ordinal regression.

Qualitative Analysis

We also used qualitative methods to better understand participants' free-text responses. In particular, we coded responses to the seven questions about attacker models, as well as all password pairs where participants' perceptions differed statistically significantly from the guess numbers we calculated. One member of the research team first read through all responses to a question and proposed codes that would capture common themes. This researcher then coded all responses and updated the codebook when necessary. A second coder used the annotated codebook to independently code the data. Intercoder agreement ranged from 85.0% to 91.4% per question, while Krippendorff's α ranged from 0.80 to 0.88. The coders met, discussed discrepancies, and agreed on the final codes.

In presenting our results, we report counts of how many participants wrote responses exhibiting particular themes to comprehensively summarize our data, not to suggest statistical significance or generalizability of proportions.

Limitations

The generalizability of our study is limited due to our use of an online convenience sample that is not representative of any larger population. Password practices are impacted by an individual's technical skills [63], and the MTurk population is younger and more technical than the overall U.S. population [58]. This skew may be exacerbated by the self-selection biases of workers who would select a study on password security. However, very few of our participants displayed any sophisticated understanding of password threats.

The security of a password, both in actuality and in perception, depends on far more factors [21, 31] than one could test in a single study. These factors include expectations about potential attackers, how the user values the account [27, 54], the user's demographics [9, 47, 51], and how well the training data used to guess passwords matches the target population and the individual [47, 49]. Some of these factors require a very large set of user-chosen passwords to analyze accurately [10]. Furthermore, the types and number of guesses an attacker might make against a particular password are influenced by the value of the information the password protects and either the hash function used or the rate-limiting employed.

While the Password Guessability Service we use reflects the performance of current password-cracking approaches and has been shown to model a skilled attacker [71], no model is perfect. A new algorithm or unexplored source of training data could vastly improve cracking and impact the ordering of guesses, changing what features make a password secure.

RESULTS

We first briefly describe our participants. To contextualize their other answers, we then report on participants' impressions of attackers and password threats. Note that we asked these questions about attackers last in the actual study to avoid priming participants. We then present participants' perceptions of the password pairs, followed by perceptions of both security and memorability for selected passwords and strategies.

Participants

A total of 165 individuals participated in our study. Our sample was nearly gender-balanced; 49% of participants identified as female, and 51% as male. Participants hailed from 33 U.S. states. They ranged in age from 18–66, with a mean age of 34.2 years and median of 33. All participants correctly answered the attention-check question.

Few participants had special familiarity with computer security. Only six participants (4%) responded that they were a professional or student in computer security. In addition, only 26 participants (16%) said they had held a job or received a degree in "computer science or any related technology field."

Attacker Model

Because any analysis of perceived or actual password security depends on the threat, we investigated *whom*, if anyone, participants expected might try to guess their passwords and *why* such people might do so. We also investigated participants' understanding of *how* attackers guess passwords and expectation for how many guesses an attacker might make.

Type of Attacker	#	%
Stranger	**135**	**82%**
Financially motivated	88	53%
Hackers	66	40%
Other strangers	14	8%
Government	3	2%
Familiar person	**38**	**23%**
People I know (generic)	23	14%
Family	9	5%
Friend	9	5%
Coworker	3	2%
No one	**8**	**5%**

Table 1. Themes describing "the type of attacker (or multiple types of attackers), if any, whom you worry might try to guess your password." # is the number of participants who mentioned the theme. The bolded categories represent participants who mentioned at least one sub-theme.

Who Tries to Guess Passwords

Actual threats to passwords include both familiar people, who might attempt to access the account of a friend or family member, and strangers conducting large-scale attacks on passwords. Most participants' expectations for *who* might try to guess their password centered on some combination of these two types (Table 1). Overall, 135 participants (82%) mentioned a stranger of any sort as a possible attacker, and 38 participants (23%) mentioned someone they know as a possible attacker.

Participants generally expected strangers to be both unfamiliar and geographically far away. For instance, P62 feared "someone on the other side of world who compromises all my accounts." Hackers were specifically mentioned by 66 participants (40%). Most of these participants discussed hackers abstractly; only one (P30, who has a technical background) expressed detailed knowledge of attacks. He wrote, "I mainly worry about large scale attacks....If my password used personal information like my telephone number...it might not be that detrimental because the attackers aren't going to do a search for personal information on each individual."

In contrast, many other participants anticipated that attackers who were strangers would have access to their personal information. For instance, P126 worried about "a stranger that has gotten hold of the names and birthdays of my family and pets," while P164 worried about people who have "hacked into businesses and gotten personal information, like my name, account numbers, my birth date."

Other expected attackers were familiar; 38 participants (23%) mentioned worrying about attacks from someone they know, such as "an angry ex or friend" (P98). Only 23 participants (14%), however, listed both strangers and familiar people. P111 was one of these participants, listing both "cyber-thieves and nosy friends or family members."

Eight participants (5%) did not expect anyone would try to guess their password, most frequently because they did not think they had anything an attacker would want. For example,

Motivation	#	%
Financial payoff	109	66%
Gather personal information	67	41%
Identity theft	33	20%
Fun / prove they can	10	6%
Spamming	6	4%
Spying	4	2%

Table 2. Themes' frequency of occurrence in participants' responses to "*why* would an attacker try to guess your password, if at all?"

P20 did not "worry too much about people trying to guess my password as I am an insignificant person."

Why Attackers Guess Passwords

There are a litany of reasons attackers might try to guess passwords, ranging from the hope of selling credentials on the black market to pride within the hacker community. When we explicitly asked *why* someone might try to guess their password, participants most frequently mentioned financial motivations and the theft of personal information (Table 2).

Of the 165 participants, 109 (66%) specifically mentioned financial motivations why attackers would try to guess a user's password. For instance, P3 and P30 mentioned "credit card" and "banking information" as objectives. Thirty-three participants (20%) specifically mentioned identity theft.

Next most commonly, participants listed the theft of personal information (67 participants, 41%). For instance, P146 worried attackers "might try to hack into my email account so they can find more personal information about me." P19 articulated both financial and personal reasons, expecting attackers would try "to find something embarrassing" or use personal information to "impersonate me, or get my money."

Ten participants mentioned motivations related to attackers having fun or proving their skills. P105 articulated this motivation as, "To cause chaos, to say they did, because they can," while P128 described such attacks as "for their sick chuckles." Six participants (4%) mentioned either email spam or social media spam, while another four participants (2%) mentioned spying, including for "state intelligence purposes" (P11).

How Attackers Try to Guess Passwords

We also asked, "As far as you know, *how* do attackers try to guess your password?" As detailed in Table 3, participants most commonly mentioned large-scale, automated guessing attacks (121 participants, 73%) or attacks targeted to the particular user (72 participants, 44%). In reality, both types of attacks occur. While 146 participants (88%) mentioned at least one of these types, only 47 participants (28%) mentioned both.

Most, but not all, participants (121 participants, 73%) anticipated that passwords might be subjected to large-scale guessing attacks. Nearly half of participants (79, 48%) specifically mentioned that they expected attackers to use software or other algorithmic techniques for large-scale guessing. P3 explained, "They use software designed to hack passwords. I've read about it." Similarly, P48 anticipated "some kind of script that runs down through password combinations automatically."

Attackers often use lists of leaked passwords and dictionaries of words and phrases as a starting point [33, 71]. Participants expected that large-scale guessing would first prioritize, in P31's words, "common things. People are fairly uncreative." For instance, P120 thought attackers would try "common names and numbers first and then work from there like maybe what people like." P157 expected attackers would "first [try] a dictionary of common words" before proceeding to try all "combinations of letters & numbers." While 42 participants (26%) used the phrase "brute force," some meant trying every possible combination, while others meant trying many possibilities (e.g., P10's "brute forcing the dictionary").

Guessing Method	#	%
Automated, large-scale	**121**	**73%**
Software / algorithms	79	48%
"Brute force"	42	26%
Dictionaries / words	27	16%
Common passwords	26	16%
Common names	8	5%
Try guessing dates	7	4%
Targeted to user	**72**	**44%**
Use personalized information	62	38%
Social engineering	7	4%
Manual guessing	4	2%
Other means	**22**	**13%**
Hacking into system / database	12	7%
Keyloggers	10	6%
Phishing	5	3%

Table 3. Themes' frequency of occurrence in participants' responses to "*how* do attackers try to guess your password?"

In contrast to large-scale guessing, 72 participants (44%) expected that password-guessing attacks could be targeted specifically to them by "using information that they already know about me" (P29), or if an attacker were to "scrape [my] personal details from social media" (P32). Participants expected that attackers might use information including "my likes, hobbies, music" (P58), "important dates" (P87), "favorite places" (P119) and "family members' names or birthdates" (P61).

Some of the 47 participants who mentioned both large-scale guessing and targeted attacks spoke of them as separate attacks, while others expected the techniques to be used in tandem. P162 exemplified those who discussed them separately, describing that attackers would guess passwords "if they know personal information about you or use hacker software to decipher passwords." In contrast, P80 wrote, "I think they look for weak passwords that are commonly used and narrow their guesses with any personal information that they have."

Estimating Numbers of Adversarial Guesses

To understand participants' security calculus, we asked, "How many guesses (by an attacker) would a password need to be able to withstand for you to consider it secure?" We required a numerical estimate (neither words nor exponential notation were permitted) and free-text justification. If stored following best practices, a password that can withstand 10^6 and 10^{14} guesses would likely be safe from online and offline attacks, respectively [26]. For passwords stored unsalted using a fast hash function (e.g., MD5), 10^{20} guesses is plausible [32, 64].

Participants' responses ranged very widely, from considering a password secure if it can withstand 2 guesses to estimating a secure password should be able to withstand 10^{59} guesses. We observed three main categories of estimates; 34% of participants wrote a number of guesses that was 50 or smaller, 67% of participants wrote a number of guesses that was 50,000 or smaller, and only seven participants (4%) wrote at least 10^{14} guesses, which required that they type 14 or more zeros.

These three categories map to three streams of reasoning. The first stream focused on online attacks. In total, 27 participants (16%) specifically noted lock-out policies, in which a server blocks guessing after a few incorrect attempts. P12 explained, "Most secure sites cut you off after 3 or 4 guesses," while P67 chose 20 guesses because "if the authentication mechanism hasn't shut down attempts by this point, I'm more worried about the platform than my password."

The second stream of reasoning centered on an attacker "giving up." In total, 42 participants (25%) explicitly mentioned that an attacker would give up, yet the number of guesses they estimated it would take varied widely. Some participants expected an attacker to get frustrated after dozens of guesses. For example, P150 wrote, "I feel like by the 10th [guess] they'd give up." Other participants chose far larger numbers. For instance, P104 chose 150,000 guesses because "hackers have short attention spans...Hopefully if by that many guesses they haven't gotten it they are on to something else." Ten other participants (6%) wrote that attackers would move on to other users with even weaker passwords than them, implicitly giving up. P4 chose 1,000 guesses, explaining, "I feel as though it wouldn't be efficient to continue attempting beyond that point, even if the process were automated. There are so many more potential victims whose passwords might be more obvious."

The third stream of reasoning involved participants estimating a strong attacker's computational resources. The magnitude that constituted a very large number varied widely. For example, P3 chose 1 million guesses, explaining, "I've read that hackers use sophisticated software that can bombard a computer or website with thousands of 'guesses' a minute." P78 also chose 1 million guesses because passwords "should be able to withstand a pretty extensive 'brute force' attack." Other participants chose far larger numbers. P103 chose 10^{14} because it "seemed like a high enough number to make it impossible or take more than 50 years to crack."

As evidenced by the wide variance in estimates, many participants were uncertain of the scale of guessing attacks. P38 wrote, "I really wanted to write 'infinite.' I didn't know how to quantify this because I don't know how many guesses hackers typically take." She settled on 1,000 guesses as a proxy for "infinite." Many others made very low estimates. For example, P88 wrote, "A dozen guesses would mean they tried every obvious password and hopefully move on after that point." In contrast, P127 chose 10^{12} because "that's basically the highest number I can think of short of infinity," yet represents under three seconds of guessing in an offline attack against MD5-hashed passwords [61, 64]. Expertise did not reduce this variance. Of the seven participants who wrote 10^{14} guesses or higher, corresponding to an offline attack [26], only one held a degree or job in computer science. Similarly, guess estimates from the six participants who reported computer security expertise ranged from 3 guesses to 500 million.

Table 4. The 25 hypotheses we investigated among password pairs. The # column indicates for how many of the three pairs per hypothesis participants' perceptions matched the hypothesized perception.

Hypothesized user perception	**#**
Capitalization	
Non-standard capitalization more secure than capitalizing first letters	3
The use of letters vs. digits vs. symbols	
Appending a lowercase letter more secure than appending a digit	2
Appending lowercase letters more secure than appending digits	0
Symbol more secure than corresponding digit (e.g., "!" vs. "1")	3
All-symbol password more secure than all-digit password	3
Adding an exclamation point makes a password more secure	3
Appending 1 makes a password more secure	3
Appending 1! makes a password more secure	3
Location of digits and symbols	
Digit in middle of password more secure than at beginning or end	3
Symbol in middle of password more secure than at beginning or end	2
Choice of digits and symbols	
Appending random digits more secure than appending a recent year	3
Random digits more secure than common sequence (e.g., "123")	3
Choice of words and phrases	
Dictionary word more secure than a person's name	3
Word that is hard to spell more secure than easy-to-spell word/phrase	1
Uncommon phrase more secure than common phrase	0
Phrase more secure than single word	2
Targeted and personal information	
Unrelated word more secure than word related to the account	3
Unrelated name more secure than name of friend/family	3
Unrelated date more secure than birthdate of friend/family	3
Keyboard patterns	
Common keyboard pattern more secure than common phrase	1
Password without obvious pattern more secure than keyboard pattern	2
Character substitutions	
Lowercase letter less secure than number/symbol (e.g., "3" for "e")	3
Uppercase letter less secure than number/symbol (e.g., "3" for "E")	1
Relevant digit (e.g., "punk4life") less secure than unrelated digit	3
Relevant digit (e.g., "4") less secure than full word (e.g., "for")	0

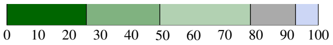
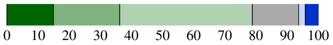
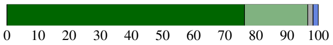
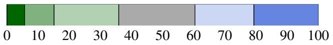
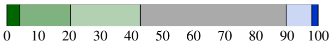
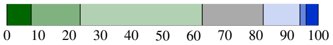
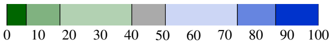
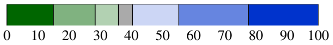
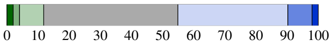
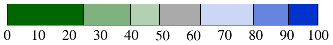
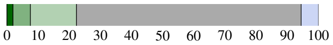
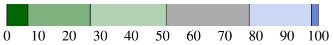
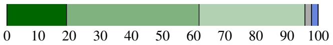
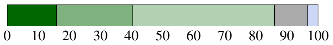
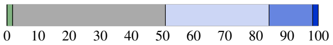
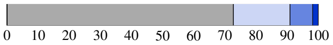
Password Pairs

In the password pairs portion of the study, participants rated the relative security of careful juxtapositions of two passwords. As shown in Table 4, participants' perceptions matched many, but not all, of our 25 hypotheses of their perceptions.

Beyond matching our hypothesized perceptions, participants' perceptions were frequently consistent with the passwords' relative guessability. Of the 75 pairs of passwords (25 hypotheses × 3 pairs each), participants' perceptions of the relative security of 59 pairs (79%) were consistent with the performance of current password-cracking approaches. In short, participants realized the following behaviors are beneficial to security:

- capitalizing the middle of words, rather than the beginning
- putting digits and symbols in the middle of the password, as opposed to the end
- using random-seeming digit sequences, rather than years or obvious sequences
- using symbols in place of digits
- preferring dictionary words over common first names
- avoiding personal content (e.g., a relative's name)
- avoiding terms related to the account (e.g., "survey" for an MTurk password)

Their free-text responses supported their numerical ratings. Participants preferred "random capitization of letters rather

PW_1	PW_2	Actually Stronger	Perceived Stronger	p	Perceptions
p@ssw0rd	pAsswOrd	PW_2 (4×10^3)	PW_1	<.001	
punk4life	punkforlife	PW_2 (1×10^3)	PW_1	<.001	
1qaz2wsx3edc	thefirstkiss	PW_2 (3×10^2)	PW_1	<.001	
iloveyou88	ieatkale88	PW_2 (4×10^9)	Neither	–	
astley123	astleyabc	PW_2 (9×10^5)	Neither	–	
jonny1421	jonnyrtxe	PW_2 (7×10^5)	Neither	–	
brooklyn16	brooklynqy	PW_2 (3×10^5)	Neither	–	
abc123def789	293070844005	PW_2 (8×10^2)	Neither	–	
puppydog3	puppydogv	PW_2 (7×10^2)	Neither	–	
qwertyuiop	bradybunch	PW_2 (4×10^2)	Neither	–	
bluewater	nightgown	PW_2 (3×10^1)	Neither	–	
iloveliverpool	questionnaires	PW_2 (2×10^1)	Neither	–	
L0vemetal	Lovemetal	Neither	PW_1	<.001	
sk8erboy	skaterboy	Neither	PW_1	<.001	
badboys234	badboys833	Neither	PW_2	.001	
jackie1234	soccer1234	Neither	PW_2	.034	

■ PW_1 much more secure ■ PW_1 more secure ■ PW_1 slightly more secure ■ Equally secure ■ PW_2 slightly more secure ■ PW_2 more secure ■ PW_2 much more secure

Table 5. Pairs of passwords for which participants' perceptions of the relative security of the passwords differed from actual security. The number in parentheses indicates how many times stronger PW_2 was than PW_1 (ratio of guess numbers).

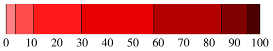
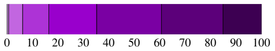
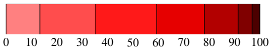
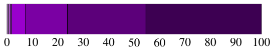
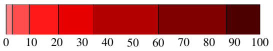
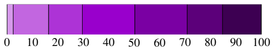
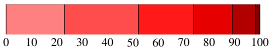
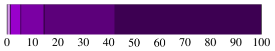
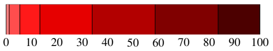
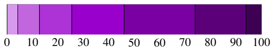
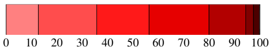
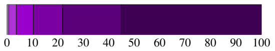
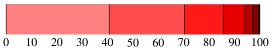
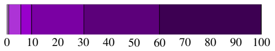
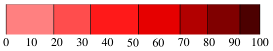
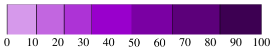
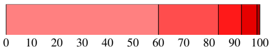
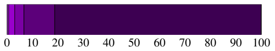
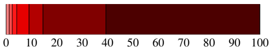
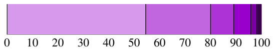
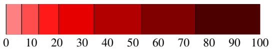
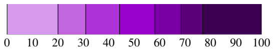
Strategy	ρ	Perceived Security	Perceived Memorability
S1: Starting with a word that comes to mind, and then adding digits or symbols to the end (e.g., bubblegum1!).	-0.22		
S2: Using a phrase taken from the lyrics to a song (e.g., somewhere over the rainbow).	-0.26		
S3: Using a phrase that you make up exclusively for this account and that has nothing to do with the account (e.g., skyscraper cornstalks).	-0.24		
S4: Using the name of one of your family members and their birth year (e.g., Zachary1976), assuming that information is not on Facebook or other social media.	-0.30		
S5: Combining words from two different languages (e.g., desaparecido rainbow).	-0.23		
S6: For your password, using a date that is meaningful to you (e.g., 12151976 because your sibling was born on 12/15/1976), assuming that information is not on Facebook or other social media.	-0.27		
S7: Basing a password on a phrase that describes your relationship to the account (e.g., iloveshoppingonamazon for your Amazon.com password).	-0.47		
S8: Building the password by following a pattern on the keyboard (e.g., 1qaz2wsx3edc).	-0.47		
S9: Using the same password that you use for other accounts.	-0.40		
S10: Using a tool that can randomly generate a complex password for you.	-0.45		
S11: Picking a complex password and writing that password down on a piece of paper that only you know about.	n.s.		

Table 6. Perceptions of the security and memorability of strategies. Participants rated both on a 1–7 scale, where 7 (darker colors on the graphs) indicates "very secure" and "very easy to remember," respectively. Spearman's ρ indicates the correlation between security and memorability ratings.

than capitalizing each word" (P8). They knew "the use of people's names is more common" (P12) and that "everyone puts the numbers at the end, moving them to a different spot helps" (P66). They knew they should not use words associated with their account; P165 correctly noted, "Surveys are popular on mturk and one password is associated with that." In addition, they knew "people use years in passwords (birthdays, anniversaries, etc.) often, so they are easier to guess" (P163).

Participants also correctly recognized that users rarely include symbols in their passwords, rating passwords with symbols higher than those with digits even though we always replaced a digit with the symbol that shares its key on the keyboard. As P16 explained, "The ˆ symbol is slightly more obscure than the 6, although it's the same key on the keyboard." They also realized that "obscure words" (P4) would be less likely to be guessed, particularly when the words were uncommon enough for P106 to incorrectly assert, "Moldovan is more secure because it's a made up word."

In contrast, the 16 pairs for which users' perceptions were inconsistent with current password cracking reveal four main misconceptions. In Table 5, we list these pairs and the actual ratio between the passwords' guess numbers. We consider two passwords to be equivalent in strength if their guess numbers are within an order of magnitude of each other (i.e., the ratio is between 0.1 and 10). We also graph the distribution of users' perceptions and give the (Bonferroni-corrected) p-value from the one-sample Wilcoxon Signed-Rank Test. Significant p-values indicate that participants tended to rate one password as more secure than the other.

The first common misconception was that adding digits inherently makes a password more secure than using only letters. Participants expected passwords like *brooklyn16* and *astley123* to be more secure than *brooklynqy* and *astleyabc*, respectively. Participants felt that "a mix of numbers and letters is always more secure and harder to guess" (P23) and that using "both numbers and letters...makes it more secure (unless the numbers were a birthday, address, etc.)" (P101). Because users frequently append numbers, however, the opposite is true in current password cracking. While, as P126 wrote, "Adding numbers makes the password more complex (more potential combinations when 26 letters and 10 numbers are possible)," arguments based on combinatorics fail because attackers exploit users' tendency to append digits to passwords [33, 55].

Participants' misconceptions about the security of digits also influenced how they perceived passwords that subsitute digits or symbols for letters. Inconsistent with password-cracking tools, which exploit users' tendency to make predictable substitutions, participants expected passwords like *punk4life* to be more secure than *punkforlife* and *p@ssw0rd* to be more secure than *pAsswOrd*. Participants incorrectly expected that "adding a number helps a lot" (P81). Similarly, P8 underestimated the rarity of unexpected capitalization, writing that "symbols and numbers are used instead of just capitalization."

Third, participants overestimated the security of keyboard patterns. Inconsistent with current password cracking [36], participants believed that *1qaz2wsx3edc* would be more secure than *thefirstkiss*, and that *qwertyuiop* would be more secure than *bradybunch*. The fact that *1qaz2wsx3edc* "contains [both] numbers and letters" (P54) outweighed its status as a keyboard pattern. The significance of the Brady Bunch in popular culture led participants to think it was more obvious than a keyboard pattern. P14 wrote, "Bradybunch is a dictionary type of guess which makes it more vulnerable." These participants, and many others, failed to realize that attackers' "dictionaries" include common strings like keyboard patterns, not just words.

Characteristic	Coefficient	Pr(>|z|)
Length	0.144	<.001
Contains uppercase letter	0.584	.020
Contains digit	0.971	<.001
Contains symbol	1.220	.006
Interaction: Upper*Digit	0.441	.010
Interaction: Digit*Symbol	-0.613	.008

Table 7. Significant terms in our mixed-model, ordinal regression of how password characteristics correlate with participants' *security* ratings.

Characteristic	Coefficient	Pr(>|z|)
Length	-0.125	<.001
Contains digit	-0.753	<.001

Table 8. The only two significant terms in our mixed-model, ordinal regression of how characteristics correlate with *memorability* ratings.

Finally, participants misjudged the popularity of particular words and phrases. In our security analysis, *ieatkale88* required over a billion times as many guesses as *iloveyou88* because the string "iloveyou" is one of the most common in passwords [37]. While some participants realized that "eating kale is a lot more rare than love" (P122), most did not; participants on the whole did not perceive one as more secure than the other. For instance, P50 wrote, "I think both are the same. Both are a combination of dictionary words and are appended by numbers." Even beyond "iloveyou," passwords often contain professions of love [74]. Participants did not realize that the dictionary word *questionnaires* would thus be a more secure password than *iloveliverpool*. Many participants thought the latter would be "more secure because it is a phrase, whereas the other password is just one word" (P146).

Selected-Password Analysis

For the passwords in our selected-password analysis, we ran two mixed-model, ordinal regressions where the security and memorability ratings (1–7) were each dependent variables, and the password's length and inclusion of {0,1,2+} uppercase letters, digits, and symbols were the independent variables. We included terms for interactions among character classes.

In our model of *security* ratings (Table 7), participants tended to rate a password as more secure if it was longer and if it included uppercase letters, digits, or symbols. More precisely, security ratings for paswords selected from the RockYou set were signficantly correlated with all four main independent variables. We also observed a significant positive interaction between the inclusion of uppercase letters and digits, and a significant negative interaction between digits and symbols.

Participants' *memorability* ratings were less clear-cut (Table 8). Participants perceived a password as significantly less memorable if it was longer or contained digits. Note, however, that many RockYou passwords contain long, random-seeming

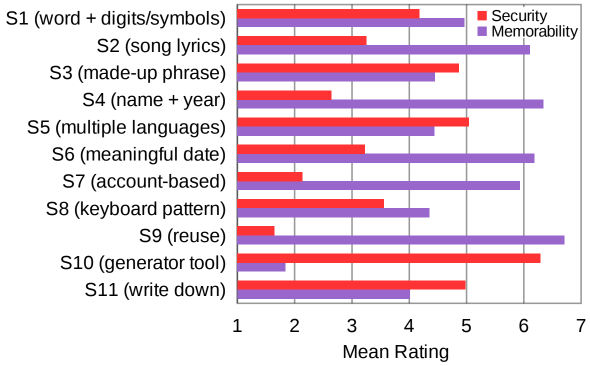


Figure 3. Mean ratings for the security and memorability of the 11 password-creation strategies.

strings of digits that contain subtle patterns or are semantically significant for speakers of other languages [47], which we hypothesize caused participants to perceive digits as particularly hard to remember. No other regression terms were significant, suggesting that factors other than length and character-class usage primarily impact perceived memorability. Unsurprisingly, participants' memorability ratings were inversely correlated with strength ratings (Spearman's $\rho = -0.678$, $p<.001$).

Password-Creation Strategies

Participants' perceptions of the 11 common strategies for password creation and management that we showed were generally consistent with current attacks on passwords. As shown in Table 6, participants realized that password reuse is wholly insecure, yet memorable. While participants believed passwords based on song lyrics or relevant dates would be memorable, they also mostly realized such passwords are insecure. In contrast, participants had divergent perceptions of the security of writing a password down. Writing passwords down was traditionally discouraged, yet has more recently been argued as a sensible coping mechanism [27, 59].

As one might expect, participants perceived a tradeoff between security and memorability; the more secure a participant rated a strategy, the less memorable he or she tended to rate it. As shown in Table 6, for each strategy we calculated Spearman's ρ to find the correlation between security and memorability ratings. For all ten strategies other than writing a password down, we found a negative correlation between security and memorability (ρ ranging from -0.22 to -0.47).

Some strategies balanced security and memorability more successfully. To ease comparison, we plot the mean ratings for each strategy in Figure 3. Creating a made-up phrase (S3) and combining languages (S5) had both security and memorability ratings with means above 4.4 on the 7-point scale. In contrast, automatically generated passwords (S10) were perceived as secure, but not memorable, whereas basing passwords on the account was perceived as very memorable, yet very insecure.

DISCUSSION AND CONCLUSIONS

We have presented the first study comparing users' perceptions of the security of text passwords with those passwords' ability to withstand state-of-the-art password cracking. Because predictable passwords are ubiquitous [9, 33, 37, 72] even for important accounts [51], we were surprised to find that participants' perceptions of what characteristics make a password more secure are, more often than not, consistent with the performance of current password-cracking approaches.

Participants did have some critical misunderstandings, however. They severely overestimated the benefit of adding digits to passwords and underestimated the predictability of keyboard patterns and common phrases (e.g., "iloveyou"). In essence, participants did not realize how common these behaviors are, which is not surprising since users never see other users' passwords. A promising direction to help users better evaluate their passwords relative to common practices is through targeted, data-driven feedback during password creation. Current password-strength meters only tell users if a password is weak or strong, not why [19, 68, 81]. Future work in this area could build on a recent study that showed users likely "autocompletions" of the partial password they had typed [45]. In large part, our results suggest that users are already aware of ways to make their passwords stronger, but they do not do so. Thus, such future work could build on research using motivational statements [22, 73] or peer pressure [23, 60] to "nudge" users [66] to create stronger passwords.

Our finding that participants mostly knew whether particular characteristics would make passwords easier or harder for attackers to guess may seem at odds with the pervasiveness of poor passwords. This gap, however, may be the result of neglecting to help users understand the spectrum of attacks against passwords. As in other studies [39, 70, 78], our participants knew passwords were important, yet their models of attackers were often incomplete. Whereas one-third of our participants considered a password secure if it can withstand as little as several dozen guesses, others believed a password must withstand quadrillions of guesses or more.

Users' incomplete understanding of the scale of potential attacks thus seems to be a root cause of bad passwords. As we surveyed in the background section, the spectrum of threats to passwords is complex and nuanced. For instance, a password's resistance to large-scale guessing matters mostly if the user reuses that password for other accounts [26] or if the service provider fails to follow security best practices [16, 34, 48]. Following the principle of defense in depth, users should protect themselves against all likely attackers, which is why security experts often recommend using password managers to store unique passwords for each account [39]. Unfortunately, users receive scant holistic advice on the overall password ecosystem [27, 39, 57]. No system administrators are incentivized to encourage users to make weak passwords for unimportant accounts or to write their passwords down [59]. Thus, users derive oversimplified folk models [78] and misconceptions [70]. In this paper, we showed that users understand quite a bit about the characteristics of strong and weak passwords, which should be leveraged to help users create stronger passwords.

ACKNOWLEDGMENTS

We thank Adam Aviv and Amrith Deepak for their help. This research was supported in part by NSF grants DGE-0903659 and CNS-1116776, as well as gifts from the PNC Center for Financial Services Innovation and Microsoft Research.

REFERENCES

1. Steven Van Acker, Daniel Hausknecht, Wouter Joosen, and Andrei Sabelfeld. 2015. Password Meters and Generators on the Web: From Large-Scale Empirical Study to Getting It Right. In *Proc. CODASPY*.
2. Anne Adams and Martina Angela Sasse. 1999. Users are not the enemy. *Commun. ACM* 42, 12 (1999), 40–46.
3. Farzaneh Asgharpour, Debin Lu, and L. Jean Camp. 2007. Mental Models of Computer Security Risks. In *Proc. WEIS*.
4. Adam J. Aviv and Dane Fichter. 2014. Understanding Visual Perceptions of Usability and Security of Android's Graphical Password Pattern. In *Proc. ACSAC*. 286–295.
5. Chris Baraniuk. 2015. Ashley Madison: Two women explain how hack changed their lives. *BBC* `http://www.bbc.co.uk/news/technology-34072762`. (August 27, 2015).
6. Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis* 20 (2012), 351–368.
7. Chandrasekhar Bhagavatula, Blase Ur, Kevin Iacovino, Su Mon Kywe, Lorrie Faith Cranor, and Marios Savvides. 2015. Biometric Authentication on iPhone and Android: Usability, Perceptions, and Influences on Adoption. In *Proc. USEC*.
8. Joseph Bonneau. 2010. The Gawker hack: How a million passwords were lost. *Light Blue Touchpaper* Blog. (December 2010). `http://www.lightbluetouchpaper.org/2010/12/15/the-gawker-hack-how-a-million-passwords-were-lost/`.
9. Joseph Bonneau. 2012a. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *Proc. IEEE Symposium on Security and Privacy*.
10. Joseph Bonneau. 2012b. Statistical metrics for individual password strength. In *Proc. Workshop on Security Protocols*.
11. Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. 2012. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *Proc. IEEE Symposium on Security and Privacy*.
12. Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. 2015. Passwords and the Evolution of Imperfect Authentication. *CACM* 58, 7 (June 2015), 78–87.
13. Joseph Bonneau and Ekaterina Shutova. 2012. Linguistic properties of multi-word passphrases. In *Proc. USEC*.
14. Jon Brodkin. 2012. 10 (or so) of the worst passwords exposed by the LinkedIn hack. *Ars Technica*. (June 2012).
15. Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6, 1 (2011), 3–5.
16. Dell Cameron. 2014. Apple knew of iCloud security hole 6 months before Celebgate. *The Daily Dot*. (September 24 2014). `http://www.dailydot.com/technology/apple-icloud-brute-force-attack-march/`.
17. Carnegie Mellon University. 2015. Password Guessability Service. `https://pgs.ece.cmu.edu`. (2015).
18. Anupam Das, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. 2014. The Tangled Web of Password Reuse. In *Proc. NDSS*.
19. Xavier de Carné de Carnavalet and Mohammad Mannan. 2014. From Very Weak to Very Strong: Analyzing Password-Strength Meters. In *Proc. NDSS*.
20. Alexander De Luca, Alina Hang, Emanuel von Zezschwitz, and Heinrich Hussmann. 2015. I Feel Like I'm Taking Selfies All Day! Towards Understanding Biometric Authentication on Smartphones. In *Proc. CHI*.
21. Geoffrey B. Duggan, Hilary Johnson, and Beate Grawemeyer. 2012. Rational security: Modelling everyday password use. *International Journal of Human-Computer Studies* 70, 6 (2012), 415 – 431.
22. David Eargle, John Godfrey, Hsin Miao, Scott Stevenson, Richard Shay, Blase Ur, and Lorrie Cranor. 2015. You Can Do Better — Motivational Statements in Password-Meter Feedback. *SOUPS Poster* (2015).
23. Serge Egelman, Andreas Sotirakopoulos, Ilar Muslukhov, Konstantin Beznosov, and Cormac Herley. 2013. Does My Password Go up to Eleven?. In *Proc. CHI*.
24. Sascha Fahl, Marian Harbach, Yasemin Acar, and Matthew Smith. 2013. On The Ecological Validity of a Password Study. In *Proc. SOUPS*.
25. Dinei Florêncio and Cormac Herley. 2007. A large-scale study of web password habits. In *Proc. WWW*.
26. Dinei Florêncio, Cormac Herley, and Paul C. van Oorschot. 2014. An Administrator's Guide to Internet Password Research. In *USENIX LISA*.
27. Dinei Florêncio, Cormac Herley, and Paul C. van Oorschot. 2014. Password Portfolios and the Finite-Effort User: Sustainably Managing Large Numbers of Accounts. In *Proc. USENIX Security*.
28. Simson Garfinkel and Heather Richter Lipford. 2014. Usable Security: History, Themes, and Challenges. *Synthesis Lectures on Information Security, Privacy, and Trust* (2014).
29. Shirley Gaw and Edward W. Felten. 2006. Password management strategies for online accounts. In *Proc. SOUPS*.
30. Megan Geuss. 2015. Mozilla: data stolen from hacked bug database was used to attack Firefox. *Ars Technica* `http://arstechnica.com/security/2015/09/mozilla-data-stolen-from-hacked-bug-database-was-used-to-attack-firefox/`. (September 4, 2015).

31. Jeffrey Goldberg. 2013. Defining Password Strength. In *Passwords*.

32. Dan Goodin. 2012. Why passwords have never been weaker—and crackers have never been stronger. *Ars Technica*. (August 2012). `http://arstechnica.com/security/2012/08/passwords-under-assault/`.

33. Dan Goodin. 2013. Anatomy of a hack: How crackers ransack passwords like "qeadzcwrsfxv1331". *Ars Technica*. (May 2013). `http://arstechnica.com/security/2013/05/how-crackers-make-minced-meat-out-of-your-passwords/`.

34. Dan Goodin. 2015. Once seen as bulletproof, 11 million+ Ashley Madison passwords already cracked. *Ars Technica* `http://arstechnica.com/security/2015/09/once-seen-as-bulletproof-11-million-ashley-madison-passwords-already-cracked/`. (September 10, 2015).

35. S.M. Taiabul Haque, Matthew Wright, and Shannon Scielzo. 2013. A Study of User Password Strategy for Multiple Accounts. In *CODASPY*.

36. Shiva Houshmand, Sudhir Aggarwal, and Randy Flood. 2015. Next Gen PCFG Password Cracking. *IEEE TIFS* 10, 8 (Aug 2015), 1776–1791.

37. Imperva. 2010. Consumer Password Worst Practices. (2010). `http://www.imperva.com/docs/WP_Consumer_Password_Worst_Practices.pdf`.

38. Iulia Ion, Marc Langheinrich, Ponnurangam Kumaraguru, and Srdjan Čapkun. 2010. Influence of user perception, security needs, and social factors on device pairing method choices. In *Proc. SOUPS*.

39. Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. "…no one can hack my mind": Comparing Expert and Non-Expert Security Practices. In *Proc. SOUPS*.

40. Panagiotis G. Ipeirotis, Foster Provost, and Jing Wang. 2010. Quality Management on Amazon Mechanical Turk. In *Proc. HCOMP*. ACM, New York, NY, USA, 64–67.

41. Blake Ives, Kenneth R. Walsh, and Helmut Schneider. 2004. The Domino Effect of Password Reuse. *Commun. ACM* 47, 4 (April 2004), 75–78.

42. Markus Jakobsson and Mayank Dhiman. 2012. The Benefits of Understanding Passwords. In *Proc. HotSec*.

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

44. Saranga Komanduri. 2015. *Modeling the adversary to evaluate password strengh with limited samples*. Ph.D. Dissertation. Carnegie Mellon University.

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

46. Cynthia Kuo, Sasha Romanosky, and Lorrie Faith Cranor. 2006. Human selection of mnemonic phrase-based passwords. In *Proc. SOUPS*.

47. Zhigong Li, Weili Han, and Wenyuan Xu. 2014. A Large-Scale Empirical Analysis of Chinese Web Passwords. In *Proc. USENIX Security*.

48. Dylan Love. 2014. Apple On iCloud Breach: It's Not Our Fault Hackers Guessed Celebrity Passwords. *International Business Times*. (September 2 2014). `http://www.ibtimes.com/apple-icloud-breach-its-not-our-fault-hackers-guessed-celebrity-passwords-1676268`.

49. Jerry Ma, Weining Yang, Min Luo, and Ninghui Li. 2014. A Study of Probabilistic Password Models. In *Proc. IEEE Symp. Security & Privacy*.

50. David Malone and Kevin Maher. 2012. Investigating the distribution of password choices. In *Proc. WWW*.

51. Michelle L. Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Richard Shay, and Blase Ur. 2013. Measuring Password Guessability for an Entire University. In *Proc. CCS*.

52. William Melicher, Darya Kurilova, Sean M. Segreti, Pranshu Kalvani, Richard Shay, Blase Ur, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Michelle L. Mazurek. 2016. Usability and Security of Text Passwords on Mobile Devices. In *Proc. CHI*.

53. Randall Munroe. 2012. xkcd: Password strength. `https://www.xkcd.com/936/`. (2012).

54. Gilbert Notoatmodjo and Clark Thomborson. 2009. Passwords and perceptions. In *Proc. AISC*.

55. Alexander Peslyak. 1996-. John the Ripper. `http://www.openwall.com/john/`. (1996-).

56. Niels Provos and David Mazieres. 1999. A Future-Adaptable Password Scheme. In *Proc. USENIX*.

57. Emilee Rader, Rick Wash, and Brandon Brooks. 2012. Stories as informal lessons about security. In *Proc. SOUPS*.

58. Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who are the crowdworkers?: Shifting demographics in Mechanical Turk. In *CHI Extended Abstracts*.

59. Bruce Schneier. 2009. Password Advice. `http://www.schneier.com/blog/archives/2009/08/password_advice.html`. (August 2009).

60. Andreas Sotirakopoulos, Ildar Muslukov, Konstantin Beznosov, Cormac Herley, and Serge Egelman. 2011. Motivating Users to Choose Better Passwords Through Peer Pressure. *SOUPS Poster* (2011).

61. Jens Steubbe. 2009. Hashcat. `http://hashcat.net/oclhashcat-plus/`. (2009).

62. Elizabeth Stobert and Robert Biddle. 2014. The Password Life Cycle: User Behaviour in Managing Passwords. In *Proc. SOUPS*.

63. Elizabeth Stobert and Robert Biddle. 2015. Expert Password Management. In *Proc. Passwords*.

64. Stricture Consulting Group. 2015. Password Audits. `http://stricture-group.com/services/password-audits.htm`. (2015).

65. San-Tsai Sun, Eric Pospisil, Ildar Muslukhov, Nuray Dindar, Kirstie Hawkey, and Konstantin Beznosov. 2011. What makes users refuse web single sign-on?: An empirical investigation of OpenID. In *Proc. SOUPS*.

66. Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

67. Trustwave Spiderlabs. 2012. eHarmony Password Dump Analysis. (June 2012). `http://blog.spiderlabs.com/2012/06/eharmony-password-dump-analysis.html`.

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

69. Blase Ur, Saranga Komanduri, Richard Shay, Stephanos Matsumoto, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Michelle L. Mazurek, and Timothy Vidas. 2013. Poster: The Art of Password Creation. In *IEEE Symposium on Security and Privacy (Posters)*.

70. Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2015a. “I Added ‘!’ at the End to Make It Secure”: Observing Password Creation in the Lab. In *Proc. SOUPS*.

71. Blase Ur, Sean M. Segreti, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Saranga Komanduri, Darya Kurilova, Michelle L. Mazurek, William Melicher, and Richard Shay. 2015b. Measuring Real-World Accuracies and Biases in Modeling Password Guessability. In *Proc. USENIX Security*.

72. Ashlee Vance. 2010. If Your Password Is 123456, Just Make It HackMe. New York Times, `http://www.nytimes.com/2010/01/21/technology/21password.html`. (2010).

73. Anthony Vance, David Eargle, Kirk Ouimet, and Detmar Straub. 2013. Enhancing Password Security through Interactive Fear Appeals: A Web-Based Field Experiment. In *Proc. HICSS*.

74. Rafael Veras, Christopher Collins, and Julie Thorpe. 2014. On the Semantic Patterns of Passwords and their Security Impact. In *Proc. NDSS*.

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

76. Melanie Volkamer and Karen Renaud. 2013. Mental Models – General Introduction and Review of Their Application to Human-Centred Security. In *Number Theory and Cryptography*. Lecture Notes in Computer Science, Vol. 8260. 255–280.

77. Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussmann. 2013. Survival of the Shortest: A Retrospective Analysis of Influencing Factors on Password Composition. In *INTERACT*.

78. Rick Wash. 2010. Folk models of home computer security. In *Proc. SOUPS*.

79. Matt Weir. 2009. The RockYou 32 Million Password List Top 100. `http://reusablesec.blogspot.com/2009/12/rockyou-32-million-password-list-top.html`. (December 2009).

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

81. Dan Wheeler. 2012. zxcvbn: realistic password strength estimation. `https://blogs.dropbox.com/tech/2012/04/zxcvbn-realistic-password-strength-estimation/`. (2012).

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

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