Better managed than memorized? Studying the Impact of Managers on Password Strength and Reuse

Sanam Ghorbani Lyastani, *CISPA, Saarland University;*
Michael Schilling, *Saarland University;* Sascha Fahl, *Ruhr-University Bochum;*
Michael Backes and Sven Bugiel, *CISPA Helmholtz Center i.G.*

https://www.usenix.org/conference/usenixsecurity18/presentation/lyastani

This paper is included in the Proceedings of the 27th USENIX Security Symposium.

August 15–17, 2018 • Baltimore, MD, USA

ISBN 978-1-939133-04-5

Open access to the Proceedings of the 27th USENIX Security Symposium is sponsored by USENIX.

Better managed than memorized? Studying the Impact of Managers on Password Strength and Reuse

Sanam Ghorbani Lyastani
CISPA, Saarland University

Michael Schilling
Saarland University

Sascha Fahl
Ruhr-University Bochum

Michael Backes
CISPA Helmholtz Center i.G.

Sven Bugiel
CISPA Helmholtz Center i.G.

Abstract

Despite their well-known security problems, passwords are still the incumbent authentication method for virtually all online services. To remedy the situation, users are very often referred to password managers as a solution to the password reuse and weakness problems. However, to date the actual impact of password managers on password strength and reuse has not been studied systematically.

We provide the first large-scale study of the password managers' influence on users' real-life passwords. By combining qualitative data on users' password creation and management strategies, collected from 476 participants of an online survey, with quantitative data (incl. password metrics and entry methods) collected in situ with a browser plugin from 170 users, we were able to gain a more complete picture of the factors that influence our participants' password strength and reuse. Our approach allows us to quantify for the first time that password managers indeed influence the password security, however, whether this influence is beneficial or aggravating existing problems depends on the users' strategies and how well the manager supports the users' password management right from the time of password creation. Given our results, we think research should further investigate how managers can better support users' password strategies in order to improve password security as well as stop aggravating the existing problems.

1 Introduction

For several decades passwords prevail as the default authentication scheme for virtually all online services [44, 11, 30]. At the same time, research has again and again demonstrated that passwords perform extremely poor in terms of security [48]. For instance, various attacks exploit that humans fail to create strong passwords themselves [10, 19, 45, 31, 34]. Even worse, there is an observable trend towards an increasing number of online services that users register to. This increasing number of required passwords in combination with the limited human capacity to remember passwords leads to the bad practice of re-using passwords across accounts [26, 51, 16, 66].

In the past, different solutions have been implemented to help users creating stronger passwords, such as password meters and policies, which are also still subject of active research [41, 54, 17, 45, 68]. Among the most often recommended solutions [28, 59, 53, 62, 56] to these problems for end-users is technical support in the form of password management software. Those password managers come built-in to our browsers, as a browser plugin, or as separate applications. Password managers are being recommended as a solution because they fulfill important usability and security aspects at the same time: They store all the users' passwords so the users do not have to memorize them; they can also help users entering their passwords by automatically filling them into log-in forms; and they can also offer help in creating unique, random passwords. By today, there are several examples of third party password managers that fit this description, such as Lastpass [5], 1Password [1], and even seemingly unrelated security software, such as anti-virus [4] solutions.

Unfortunately, it has not been sufficiently studied in the past whether password managers fulfill their promise and indeed have a positive influence on password security or not? To break this question down, we are interested in *1) whether password managers actually store strong passwords that are likely auto-generated by, for instance, password generators, or if they really are just storage where users save their self-made, likely weak passwords?* Further, we are interested whether *2) users, despite using password managers, still reuse passwords across different websites or if do they use the managers' support to maintain a large set of unique passwords for every distinct service?* Prior works [66, 51] that studied password reuse and strength in situ have also considered password managers as factors, but did not find an influence by managers and could not conclusively answer those questions.

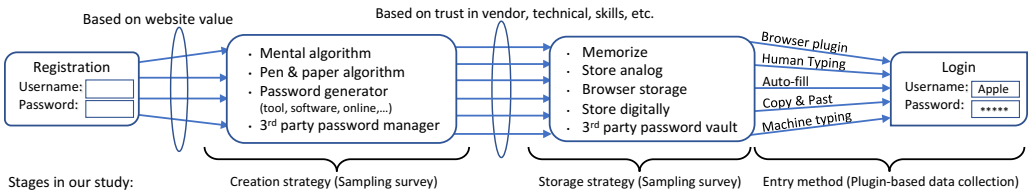

Figure 1: Users' strategies for password creation and storage plus the stages of our study to investigate managers' influence.

Our contributions: We argue that to specifically study the impact of password managers, important aspects were missing in prior work, and this paper's most tangible contribution is an extension of prior methodologies to be able to study password managers' impact in the wild. First, previous works considered only the presence of password management software on the user device and whether a password was auto-filled or not. However, to better distinguish the storage option of a password (i.e., memorized and manually entered, auto-filled by the browser, copy&pasted, or filled by a browser plugin) a more fine-grained entry method detection is required. Second, users do not axiomatically follow strict workflows for password creation, storage, and entry [27, 29, 62, 56, 58] (see Figure 1). For instance, the effort users are willing to invest in creating a unique and strong password often depends on the privacy-sensitivity of the associated account. For creating a new password, the approaches range from mental algorithms (e.g., leetifying a known word) over pen&paper algorithms and password generator tools (e.g., websites like `https://www.random.org/passwords/`) to 3rd party password managers (e.g., LastPass, KeePass, etc.). Based on different factors, such as technical skills, trust in software vendors, financial expenditure, multi-device support, or others, users resort to different password storage options from where the password finds its way via various entry methods into the login forms. To better study password managers' influence, one has to take the users' creation and storage strategies into consideration as well. In particular, one has to understand if the user pursues primarily a creation strategy based on password manager support and whether there then exists an observable effect of this strategy on the password strength and reuse.

In this paper, we present a study that reflects those considerations (see the bottom of Figure 1). We first recruited 476 participants on Amazon MTurk to conduct a survey sampling to better understand users' strategies for creating and storing passwords, their attitudes towards passwords, and past experiences with password leaks or password managers. From those insights, we identified two distinct groups in our participant pool: users of password managers and users abstaining from technical help in password creation. We were further able to recruit 170 of our participants, 49 of which reported using password managers, for a follow-up study in which our participants allowed us to monitor their passwords through a Google Chrome browser plugin that collected password metrics as well as answers to in situ questionnaires upon password entry. This gave us detailed information about real-life passwords, including their strength, their reuse, and, for the first time, their entry method (e.g., manually typed, auto-filled, pasted, or entered by a browser plugin) as well as the passwords' context, including user reported value of the password (e.g., loss of social repudiation or financial harm when the password would be leaked).

Based on the combined data from our survey sampling and plugin-based data collection, we are able to study the factors that influence password strength and reuse from a new perspective. Using exploratory data analysis and statistical testing, including regression models, we are first to actually show that password managers indeed influence password strength and reuse. In particular, the relation between different entry methods and the password strength depends on the users' entire process of password handling. Using a workflow that includes technical support from password creation through storage to entry leads to stronger passwords, while this positive effect on password strength cannot be detected when considering the input method individually. A similar picture emerges for password reuse. Passwords entered manually or by Chrome auto-fill were unique in only 20–25% of all cases. For LastPass or Copy&Paste password entry, the proportion of non-reused passwords increases to 53–78%. This is still far from ideal—that not even a single password is reused—but still a significant improvement through such dedicated password management tools. Similar to the results for password strength, we find that password reuse improves further if the password generation is technically supported. In contrast to password strength, however, this positive effect is similar for all input methods. Looking at managers that do not offer support for password creation, such as Chrome's auto-fill, we even found a negative influence in that those managers even contribute to the password reuse problem. In summary, our results support the fact that technical tools can have a very positive effect on password security. However, it is important that the entire password management process is supported—from generation, over storage, to entry—and not only the old and weak passwords of the users are stored.

2 Related Work

Textual passwords are for decades [44] the incumbent authentication scheme for online services [29, 30], and will very likely remain in that position for the foreseeable future. They distinguish themselves from alternative schemes through their very intuitive usage, however, as well as through a pathological inability of users to create passwords that withstand guessing attacks [11]. Given the permanence of passwords, users are commonly referred to technical help in form of password management software [28, 59, 53, 56] to create strong, unique passwords.

In this paper, we aim to better understand how password managers help users in this task and try to measure the impact password managers actually have on the current status quo. We do this through a comprehensive study that includes both self-reported user strategies and factors for password creation and storage as well as in situ collected password metrics and questionnaire answers. To put our approach into the larger context and to provide necessary background information, we give here an overview of prior research on how users select and (re)use passwords, how password strength can be measured, and on dedicated studies of password manager software.

2.1 Password creation

Different works have studied the strategies of users and the factors that influence the selection of new passwords. For instance, users create passwords based on something that has relevance or meaning to them [56], and very often passwords are based on a dictionary word [38, 52].

The effort the user is willing to invest into creating a stronger passwords can depend on different factors. For example, password policies that enforce a certain password composition (i.e., length and character classes) can influence the user [70, 26, 38]. Similarly, many websites use password strength meters to provide real-time feedback on new password's strength and nudge users into creating stronger passwords [23, 61]. However, often those policies and meters have inconsistent metrics across different websites [12, 65, 17], potentially confusing users about what constitutes a strong password [62]. Also the value of the password protected account can influence the user. Prior studies [8, 49, 56, 51] concluded that people try to create strong passwords for accounts that they consider more important, e.g., banking websites. In particular, users employed password managers for specific matters [56], such as just using at a work PC but not at home, or not using them for banking websites. Despite their apparent benefits, it is unclear how users *actually* use password managers and what the exact impact of password managers is on password reuse and strength.

2.2 Password strength

Password strength has been studied for several years and different mechanisms have been used to measure a password's strength. Shannon entropy [21] provides a way to estimate the strength based on the passwords composition. It was formerly used by the NIST guidelines [28] to estimate the password strength. However, more recent research [67, 10, 18, 40] argued that *guessability* metrics are a more realistic metric than the commonly used entropy metrics, and recommendations, such as NIST [28], recently picked up the results of this line of research and have been updated accordingly. One of the vital insights from this and other research [34] was that passwords are not chosen randomly but exhibit common patterns and are derived from a limited set of dictionary words.

Measuring a password's guessability has been realized in different ways. Those include Markov models [13, 19], pattern matching plus word mangling rules [68], or neural networks [45]. Since prior password strength meters were based on the password composition and the resulting entropy, those new approaches also found their way into contending password strength meters [68, 45, 60]. However, varying cracking algorithms or techniques can cause varying password strength results based on configuration, methods, or training data [63]. Also in our study we measure the password strength based on guessability, using the openly available *zxcvbn* [68] tool.

2.3 Password reuse

Prior work [56] has shown that users have an increasing number of online accounts that require creation of a new password. To cope with the task of remembering a large number of passwords, users resort to reusing passwords across different accounts [16, 37], creating a situation in which one password leak might affect multiple accounts at once. A large-scale data collection through an instrumented browser [26] was first to highlight this problem. Since then, newer studies further illustrated the issue of password reuse. For instance, in a combination of measurement study of real leaked passwords and user survey [16], 43% of the participants reused passwords and often a new password was merely a small modification of an existing one. As with password creation, different factors can influence the password reuse. For example, it was shown that the rate of reused passwords increased with the number of accounts [27], which is troublesome considering that users accumulate an increasing number of accounts. As with password strength, also the value of the website can affect whether a user creates a unique new password or reuses an existing one [8, 51].

Closest to our methodology are two recent studies [66, 51] based on data collected with browser plugins

from users. Both studies monitored websites for password entries and recorded the password characteristics, such as length and composition, a participant-specific password hash, the web domain (or domain category), as well as meta-information including installed browser plugins or installed software (e.g., anti-virus software). In case of the newer study [51], also hashes of sub-strings of the password were collected as well as a strength estimate using a neural network based password meter [45] and whether the password was auto-filled or not. Through this data, both studies had an unprecedented insight into user's real password behavior, the factors influencing password reuse, and could show that password reuse, even partial reuse of passwords, is a rampant problem. Further relating to our work, both prior studies also considered the potential influence of password managers, however, could not find any significant effect of password managers on password reuse or strength. However, their studies were not specifically targeted at investigating the impact of password managers, and with our methodology we extend those prior works in two important aspects. First, prior work only considered the presence of password managers and whether auto-fill was used. For our work, we derived a more fine-grained detection of the password entry method, which allows us to distinguish human, plugin-based, auto-fill, or copy&pasted input to password fields and thus better detection of managed passwords. Second, merely the entry method of a password does not reveal its origin (e.g., passwords from a password manager might also be copy&pasted or saved in the browser's auto-fill). To study the impact of password managers, a broader view is essential that includes the users' password creation strategies in addition to their in situ behavior.

2.4 Security of password managers

Password manager software has also been the subject of research. Human-subject studies [39, 14] have shown that they might suffer from usability problems and that ordinary users might abstain from using them due to trust issues or not seeing a necessity. Like any other software, password managers might also contain vulnerabilities [43, 71] that can compromise user information. Also the integration of password managers, in particular the password auto-filling, was scrutinized [55, 57] and flaws found that can help an adversary to sniff passwords.

3 Methodology

For our study of password managers' impact on password strength and reuse, we use data collected from paid workers of Amazon's crowd-sourcing service *Mechanical Turk*. We collected the data in two different stages: 1) a survey sampling, and 2) collection of *in situ* password metrics.

Ethical concerns: The protocols implemented in those two stages were approved by the ethical review board[1] of our university. Further, we followed the guidelines for academic requesters outlined by MTurk workers [20]. All server-side software (i.e., a LimeSurvey installation and a self-written server application) was self-hosted on a maintained and hardened university server. Web access to the server was secured with an SSL certificate issued by the university's computing center and all further access was restricted to the department's intranet and only made available to maintainers and collaborating researchers. Participants could leave the study at any time.

3.1 Password survey

In our survey sampling, we asked participants about their general privacy attitude, their attitude towards passwords, their skills and strategies for creating and managing passwords, as well as basic demographic questions. Those information enable us, on the one hand, to gain a general overview of common password creation and storage strategies. On the other hand, those information help us in detecting and avoiding any potential biases in the later stages of our study. The full survey contained 31–34 questions, categorized in 6 different groups (see Appendix A).

We first asked for their privacy attitude using the standard Westin index [42]. However, since the Westin index has been shown to be an unreliable measure of the actual privacy-related actions of users [69], we also asked about the participants' attitude towards passwords (e.g., whether they consider passwords to be futile in protecting their privacy).[2] This should help in better understanding if participants are actually motivated to put an effort into creating stronger and unique passwords. We further asked about the participants' strategies for password creation and management in order to get a more complete picture of the possible origins of passwords in our dataset.

All qualitative answers (e.g., *Q9* or *Q22* in Appendix A) were independently coded in a bottom-up fashion by two researchers. The researchers achieved an initial agreement between 95.6% (*Q9*) and 97.1% (*Q22*) and all differences could be resolved in agreement.

Participation in the survey was open to any MTurk worker that fulfilled the following criteria: the worker was located in the US and the number of previously approved tasks was at least 100 or at least 70% all of the tasks. The estimated time for answering the survey was 10–15 minutes and we paid $4 for participation. In total, 505 MTurk workers participated in our survey between August

[1] `https://erb.cs.uni-saarland.de/`

[2] Other instruments, which meet the latest requirements of scale constructions and which are often used in recent research, do not reflect the actual privacy/security attitude construct, but refer more strongly to security behavior (e.g., SeBIS [22]) or are strongly tailored to the corporate context (e.g., HAIS-Q [50]).

2017 and October 2017. After discarding responses that failed attention test questions [33], were answered too fast to be done thoughtfully, or that were duplicates, we ended up with 476 valid responses.

Lastly, we also asked whether the participant would be willing to participate in a follow-up study, in which we measure in an anonymized, privacy-protecting fashion the strength and reuse of their passwords. Only participants that indicated interest in the follow-up study were considered potential candidates for our Chrome plugin-based data collection. Only 21 workers were not interested.

3.2 Chrome plugin-based data collection

To collect in situ data about passwords, including strength, reuse, entry method, and domain, we created a Chrome browser plugin that monitors the input to password fields of loaded websites and then sends all collected metrics back to our server once the user logs in to the website. We distributed our plugin via the Google Web Store to invited participants. The plugin was unlisted in the Store, so that only participants to which we sent the link to the plugin store website were able to install it. Our primary selection criterion for participant selection was that they use Chrome as their primary browser and are not using exclusively mobile devices (smartphones and tablets) to browse the web; besides that we aimed for an unbiased sampling from the participants pool with respect to the participants' privacy attitude, attitude towards passwords, demographics, and usage of password managers. Between September and October 2017, we invited 364 participants from the survey sampling to the study, of which 174 started and 170 finished participation. We asked participants to keep our plugin installed for at least four days. Participants that finished the task were compensated with $20.

Our plugin collects the following metrics:

Composition: The length of the entered password as well as the frequency of each character class.

Strength: The password strength measured in Shannon and NIST entropy as well as zxcvbn score. Shannon and NIST entropy have been used in prior works [24, 66, 23] as a measure of password strength and complexity and are collected primarily to be backward compatible in our analysis with prior research. However, since entropy has been shown to be a poor measurement of the actual "crackability" of the password [67], we use the zxcvbn [68] score as the more realistic estimator of the password strength in our analysis.[3] Zxcvbn estimates every password's strength on a scale from 0 (weakest) to 4 (strongest) using pattern matching (e.g., repeats, sequences, keyboard patterns), common password dictionaries (including leaked passwords, names, English dictionary words), and mangling rules (e.g., leetify). Appendix B explains the meaning of this score in more detail. In our plugin we used the zxcvbn library [3] with its default settings. From a statistical point of view, a metrically scaled strength measurement instead of the ordinal zxcvbn score would have helped in finding possible effects on password strength easier (see Section 4), however, it does not affect the presence of possible effects per se.

Website category: The category of the website domain according to the *Alexa Web Information Service* [2]. Our plugin contains the category for the top 28,651 web domains at the time the study was conducted.[4]

Entry method: The method through which the password was entered, such as *human*, *Chrome auto-fill*, *copy&paste*, *3rd party password manager plugin*, or *external password manager program*. The detection of the entry method is described separately in Section 3.2.1.

In situ questionnaire: Participant's answers to a short questionnaire about the entered password and website (see Section 3.2.2). In particular, we ask about the website's *value* for their privacy. Other studies used the website category as a proxy for this value [51] and in our study we wanted first-hand knowledge (see also Appendix C).

Hashes: Adapting the methodology of [51, 66], we collect the hash of the entered password as well as the hash of every 4-character sub-string of the password. We use a keyed hash (i.e., PBKDF2 with SHA-256), where the key is generated and stored at the client side and never revealed to us. This allows identification of (partially) reused passwords *per participant*. We use the notions introduced in [51]: *Exactly reused* passwords are identical with another password, *partially reused* passwords share a sub-string with another password, and *partially-and-exactly reused* passwords have both of those characteristics. Like related work [51, 66], we cannot compare passwords across participants.

3.2.1 Detecting the entry method

Detection of the password entry method follows the decision tree depicted in Figure 2. If our plugin detects any kind of typing inside the password field (**(A)=Y**) and the typing speed is too fast to come from a human typist (**(B)=N**), we conclude that an external password manager program (such as KeePass) mimics a human typist by "replaying" the keyboard inputs of the password. Otherwise (**(B)=Y**), we assume a manually entered password. As threshold between human and external program, we set an average key press time of 30 ms. This is based on the observation that external programs usually do not consider mimicking the key press time, while some of them enter the password character-wise with varying speeds.

[3] Unfortunately, the fully trained neural network based strength estimator of [51, 45] was not publicly available.

[4] This is the number of web domains in the top 100K list, for which a category was assigned by Alexa.

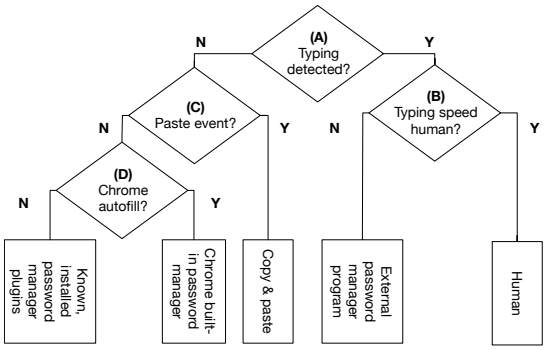

Figure 2: Decision tree to detect password entry methods

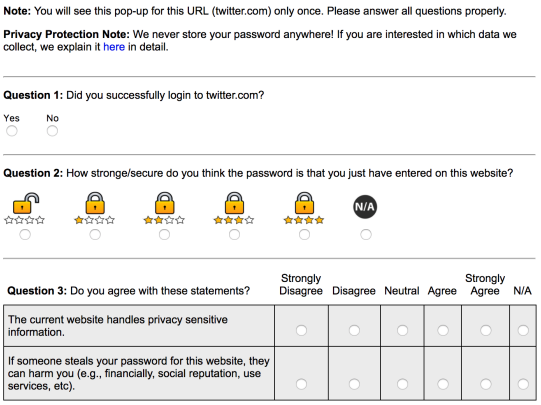

Figure 3: In situ questionnaire upon login to a new website.

In case there was no typing detected ((**A**)=**N**) and a paste event was observed ((**C**)=**Y**), we consider the password to be pasted by either a human or an external program. In either case, the password is managed externally to the browser in digital form. If no paste event was detected ((**C**)=**N**) and the Chrome auto-fill event was observed, this indicates that Chrome filled the password field from its built-in password manager. If Chrome auto-fill has not filled the password field ((**D**)=**N**), our plugin checks the list of installed plugins for eight well-known password manager plugins (see Appendix D) and reports the ones installed in the participant's browser, or an "unknown" value in case none of those eight was found.

We make the assumption that the user does not enter the password with a mixture of the different entry methods (e.g., pasting a word and complementing it with typing). Such mixture of entry methods would result in misclassification of the detected method. However, we assume that such behavior is too rare to affect our results significantly.

3.2.2 Participant instructions

We provided our participants with a project website that gave a step-by-step introduction on how to install our plugin, set it up, use it, and remove it post-participation. Google Web Store provided our participants with a very comfortable way of adding the plugin to their browser. To set the plugin up, participants had to simply enter their MTurk worker ID into the plugin. The worker ID was used as a pseudonym throughout this study to identify data of the same participant. After setup, the plugin starts monitoring the users' password entries. For every newly detected domain to which a password was submitted, our plugin asked the participant to answer a short three question questionnaire about the participants' estimate of the website's value, the participants' strength estimate of the just entered password, and whether the login was successful (see Figure 3). Every participant was instructed to use the plugin for four days, after which the plugin released a completion code to be entered into the task on MTurk to finish participation and collect the payment. Through our server logs and the Google Web Store Developer Dashboard we confirmed that all participants removed our plugin shortly after finishing participation. We also instructed participants to act naturally and not change their usual behavior during those four days in order to maximize the ecological validity of our study. The only exceptions from the usual behavior were the installation of our plugin and a request to re-login to all websites where they have an account in order to ensure a sufficient enough quantity of collected data.

3.2.3 Addressing privacy concerns

A particular consideration of our study design was the potential privacy concerns of our participants. Since we essentially ask our participants to install a key-logger that monitors some of the most privacy-sensitive data, this might repel participants from participating. Due to the lack of in-person interviews or consultation between the researchers and the participants, we tried to address those concerns through a high level of transparency, support, and collecting only the minimal amount of data in a privacy-protecting fashion, which also follows the guidelines for academic requesters [20].

First, we explained on our project website the motivation behind our study and why acting naturally is important for our results. In this context, we provided a complete list of all data that our plugin collects, for which purpose, and why this data collection does not enable us to steal the participants' passwords. We also answered all participants' questions in this regard that were sent to us via email or posted in known MTurk review/discussion forums. We received feedback from workers that this level of openness has convinced them to participate in the study. Second, we distributed our plugin in an authenticated way via the Google Web Store and did not obfuscate the plugin's code. Third, we limited the extent of the collected

data to the necessary minimum while still being able to study password managers' impact. For instance, we only collect the first successful login to any website, thus abstaining from monitoring participants' browsing behavior. Fourth, every participant could inspect the collected data per domain prior to sending them to us and chose to skip the data collection for highly sensitive websites.

4 Studying Password Managers' Impact

In this section, we analyze our collected data, but leave the discussion of our results for Section 5. After presenting our participants' demographics and an overview of their password reuse and strength, we group our participants based on their creation strategy and study the impact of different password management and creation strategies.

4.1 Demographics

Table 1 provides an overview of the demographics of our participants that answered our survey, that we invited to the plugin-based study, and that participated in the plugin-based data collection. We invited participants in equal parts from every demographic group and every demographic group participated in almost equal parts in the plugin-based data collection. We use a Mann-Whitney rank test [25] to test for significant differences between the demographic distributions of the 476 participants in the survey sampling and the 170 participants in the plugin-based study, and could not find any statistically significant ($p < .05$) differences between those two groups. In general, our participants' demographics are closer to the commonly observed demographics of qualitative studies in university settings than to the demographics of the 2010 US census [64]. Our participant number is skewed towards male participants (57.6% identified themselves as male). Also, our participants covered an age range from 18 to more than 70 years, where our sample skews to younger participants (75.2% of our study participants are younger than 40) as can be commonly observed in behavioral research, including password studies and usable security. The majority of our participants had no computer science background (80.88%) and was English speaking (98.3%). Most of the participants identified themselves as of white/Caucasian ethnicity (74.6%). The participants also covered a range of educational levels, where a Bachelor's degree was the most common degree (36.6% of all participants). Further, 80.9% of our participants reported using Chrome as their primary browser (see Table 2).

Since our study effectively asks participants to install a password-logger, we were concerned with a potential opt-in bias towards people that have low privacy concerns or consider passwords as ineffective security measures. To this end, we included the three questions of Westin's Privacy Segmentation Index [42] (*Q1* in Appendix A) to capture our participants' general privacy attitudes (i.e., fundamentalists, pragmatists, unconcerned). We further added two questions specifically about our participants' attitude about passwords (see *Q4* in Appendix A), e.g., if passwords are considered a futile protection mechanism or important for privacy protection. Table 3 summarizes the results of those questions. Only a minority of 86 of our survey participants are privacy unconcerned and the majority of 365 participants believe in the importance of passwords as a security measure. Almost a third of our survey participants experienced a password leak in the past. For our study we sampled in almost equal parts from those different groups. Using a Mann-Whitney rank test, we could not find any statistically significant differences between the survey and study participants' distribution of privacy and password attitudes/experiences. Thus, we argue that the risk of an opt-in bias towards either end of the spectrum for privacy and password attitude is unlikely.

	Survey	**Invited to study**	**Participated**
Number of participants	476	364	170
Gender			
Female	200	156 (78.0%)	73 (36.5%)
Male	274	208 (75.9%)	97 (35.4%)
Other	1	0	0
No answer	1	0	0
Age group			
18–30	180	139 (77.2%)	64 (35.6%)
31–40	178	135 (75.8%)	63 (35.4%)
41–50	71	58 (81.7%)	32 (45.1%)
51–60	35	24 (68.6%)	8 (22.9%)
61–70	11	7 (63.6%)	2 (18.2%)
≥71	1	1 (100%)	1 (100%)
Computer science background			
Yes	91	64 (70.3%)	27 (29.7%)
No	385	300 (77.9%)	143 (37.1%)
Native language			
English	468	358 (76.5%)	167 (35.7%)
Other	8	6 (75.0%)	3 (37.5%)
Education level			
Less than high school	3	3 (100%)	1 (33.3%)
High school graduate	68	53 (77.9%)	26 (38.2%)
Some college, no degree	117	85 (72.6%)	34 (29.1%)
Associate's degree	79	64 (81.0%)	34 (43.0%)
Bachelor degree	174	133 (76.4%)	62 (35.6%)
Ph.D	2	1 (50.0%)	1 (50.0%)
Graduate/prof. degree	32	25 (78.1%)	12 (37.5%)
Other	1	0	0
Ethnicity			
White/Caucasian	355	274 (77.2%)	123 (34.6%)
Black/African American	50	38 (76.0%)	25 (50.0%)
Asian	31	23 (74.2%)	9 (29.0%)
Hispanic/Latino	27	21 (77.8%)	12 (44.4%)
Native American/Alaska	1	0	0
Multiracial	7	5 (71.4%)	1 (14.3%)
Other	5	3 (60.0%)	0

Table 1: Demographics of our participants. Percentages indicate the fraction w.r.t. initial size in the survey sampling.

Browser	Chrome	Firefox	Safari	Opera	IE/Edge	Other
Share	385 (80.9%)	71 (14.9%)	7 (1.5%)	6 (1.3%)	1 (0.2%)	6 (1.3%)

Table 2: Primary browsers of our 476 survey participants.

	Survey	Invited to study	Participated
Privacy concern (Westin index)			
Fanatic	217	167 (77.0%)	66 (30.4%)
Unconcerned	86	56 (65.1%)	31 (36.0%)
Pragmatist	173	141 (81.5%)	73 (42.2%)
Attitude about passwords			
Pessimist	9	8 (88.9%)	3 (33.3%)
Optimist	365	279 (76.4%)	132 (36.2%)
Conflicted	102	77 (75.5%)	35 (34.3%)
Prior password leak experienced			
No	190	151 (79.5%)	72 (37.9%)
Yes	148	111 (75.0%)	58 (39.2%)
Not aware of	138	102 (73.9%)	40 (29.0%)

Table 3: Privacy attitude, attitude about passwords, and prior experience with password leakage among our participants.

4.2 General password statistics

Tables 4 and 5 provide summary statistics of all passwords collected by our plugin. We collected from our 170 participants 1,045 unique passwords and 1,767 password entries in total. That means, that our average participant entered passwords to 10.39 distinct domains with a standard deviation of 5.52 and median of 9. Our participants reported using on average 29.95 password-secured accounts (*Q2* in Appendix A) and we collected on average 61% of each participant's self-estimated number[5] of passwords. The lowest number of domains per participant is 1 and the highest is 27, where the 1^{st} quartile is 6 and the 3^{rd} quartile is 14. Those numbers are hence slightly lower than those reported in related studies [51]. When considering only unique passwords, our average participant has 6.15 passwords, indicating that passwords are reused frequently. Our participants entered their passwords on average with 2.24 different methods. Looking at all passwords, our participants reused on average 70.56% of their passwords, where exact-and-partial reuse is most common with 36.46% of all passwords. Interestingly the minimum and maximum in all reuse categories is 0% and 100%, respectively, meaning that we have participants that did not reuse any of their passwords as well as participants that reused all of their passwords. The average password in our dataset had a length of 9.61 and was composed of 2.52 character classes. The average zxcvbn score was 2.20, where the participant with the weakest passwords had an average of 0.67 and the participant with the strongest an average of 4.00. Like prior work [66], we observe a significant correlation between password strength and reuse (chi-square test: $\chi^2 = 75.48$, $p < .001$).

As shown in Table 5, the majority of the 1,767 logged passwords was entered with Chrome auto-fill (53.71%) followed by manual entry (33.39%). Although in our pilot study various password manager plugins, e.g., KeePass and 1Password, had been correctly detected, in our actual study only LastPass was used by our participants.

[5]Some participants underestimated this number

Statistic	Mean	Median	SD	Min	Max
No. of passwords	10.39	9.00	5.52	1.00	27.00
Entry methods	2.24	2.00	0.75	1.00	4.00
Percentage reused passwords					
Non-reused	29.44%	21.58%	28.25%	0.00%	100%
Only-exact	15.72%	0.00%	24.43%	0.00%	100%
Only-partially	18.38%	11.11%	19.88%	0.00%	100%
Exact-and-partial	36.46%	38.75%	30.88%	0.00%	100%
Password composition					
Length	9.61	9.29	1.72	6.33	16.86
Character classes	2.52	2.50	0.58	1.00	3.94
Digits	2.54	2.38	1.24	0.25	6.73
Uppercase letters	0.85	0.67	0.81	0.00	4.62
Lowercase letters	5.92	5.72	1.96	1.67	15.50
Special chars	0.30	0.10	0.54	0.00	5.19
Password strength					
Zxcvbn score	2.20	2.14	0.75	0.67	4.00
Shannon entropy	29.31	28.37	7.93	16.00	68.00
NIST entropy	23.50	23.00	2.98	17.17	35.69

Table 4: Summary statistics for all 170 participants in our plugin-based data collection. We first computed means for each participant and then computed the mean, median, standard deviation, and min/max values of those means.

Entry method	All passwords	Unique passwords
Chrome auto-fill	949 (53.71%)	540 (51.67%)
Human	590 (33.39%)	331 (31.67%)
LastPass plugin	128 (7.24%)	100 (9.57%)
Copy&paste	55 (3.11%)	51 (4.88%)
Unknown plugin	41 (2.32%)	23 (2.20%)
External manager	4 (0.23%)	0 (0.00%)
Σ	1,767	1,045

Table 5: No. of password entries with each entry method.

Of all passwords, 128 (7.24%) were entered with LastPass, which is a similar share of managers as in recent reports [46]. Copy&paste and unknown plugins formed the smallest, relevant-sized shares and only four passwords were entered programmatically by an external program.

With respect to general password reuse (see Figure 4), partial-and-exact reuse is by far the most common reuse across all entry methods, except for LastPass' plugin and Copy&paste, which have a noticeably high fraction of non-reused passwords (e.g., 68 or 53% of all passwords entered with LastPass were not reused) and have noticeably less password reuse than the overall average. Looking at the password strength for all *unique* passwords (see Figure 5), one can see that 65% or 44 of all passwords entered with LastPass are stronger than the overall average of 2.20, while the other entry methods show a more balanced distribution across the zxcvbn scores (except for score 0). In summary, this indicates that LastPass shows an improved password strength (mean of 2.80 with SD=1.07) and password uniqueness in comparison to the other entry methods. Copy&paste exhibits the strongest password uniqueness, however, at the same time the weakest password strength (1.98 on average with a SD=1.33).

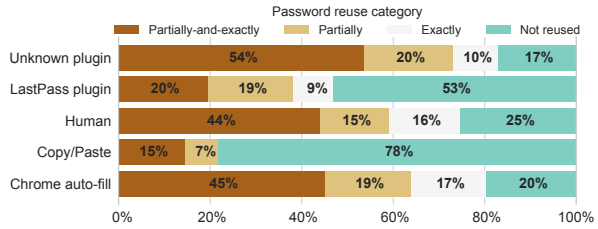

Figure 4: Password reuse by entry method for all passwords.

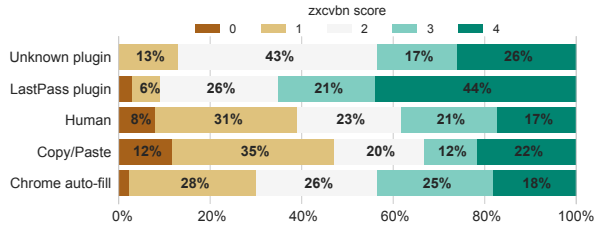

Figure 5: Zxcvbn score per entry method for unique passwords.

4.3 Grouping based on creation strategy

We grouped our participants based on their self-reported strategies for creating new passwords (see *Q9*, *Q13*, and *Q15* in Appendix A). Based on their answers, we discovered a dichotomous grouping:

Group 1: Password managers/generators ("PWM"): First, we identified participants that reported using a password generator, either as integrated part of a password manager program (e.g., *"I use lastpass.com, which automatically creates and saves very strong passwords."*) or as an extra service (*"I use a service to generate/create passwords that I put the parameters in that I would like."*). Many also implied the usage of a manager for password storage (e.g., *"I use a password creation and storage-related browser extension that also is related to an installed password manager application on my personal computer."*), however, some participants explicitly noted a separate storage solution (*"I use an app that creates random character strings to pick new passwords for me. I then memorize it so I don't have to keep it written down"* or *"I will use a random password generator. [...] I will save the new password in a secure location such as a password protected flash drive."*). In total, 45 (or 26.47%) out of 170 participants fell into this category.

Group 2: Human-generated ("Human"): We discovered that all 121 remaining participants described a strategy that abstains from using technical means. Almost all of the participants in this group reported that they *"try to come up with a (random) combination of numbers, letters, and characters."* For instance, one participant symptomatically reported: *"I think of a word I want to use and will remember like. mouse. I then decide to capitalize a letter in it like mOuse. I then add a special character to the word like mOuse@. I then decided a few numbers to add like mOuse@84."* Only a very small subgroup of seven participants reported using analog tools to create passwords, such as dice or books (*"I have a book on my desk I pick a random page number and I use the first letter of the first ten words and put the page number at the end and a period after."*), or using passphrases.

Many of the participants in this group also hinted in their answers to their password storage strategies. For instance, various participants emphasized ease of remembering as a criteria for new passwords (e.g., *"something easy to remember, replace some letters with numbers."*), others use analog or digital storage (e.g., *"I try to remember something easy or I right[sic] it down on my computer and copy&paste it when needed."*). Many participants also admitted re-using passwords as their strategy (e.g., *"I use the same password I always use because it has served me well all these years"* and *"I have several go to words i use and add numbers and symbols that i can remember"*).

4.3.1 Group demographics

We provide an overview of the groups' demographics in Appendix E. We again used a Mann-Whitney test to detect any significant differences in the distributions of those two demographic groups. We find that they have statistically significant different distribution for gender ($U = 2,366$, $p = .016$), computer science background ($U = 2,181$, $p < .001$), and attitude towards passwords ($U = 3,440$, $p = .024$). More participants in $\text{Group}_{\text{PWM}}$ identified themselves as male in comparison to $\text{Group}_{\text{Human}}$. The fractions of participants that have a computer science background and that are optimistic about passwords are higher in the group of password manager users. Gender and computer science background are significantly correlated for our participants (Fisher's exact test: $OR = 3.99$, $p = .005$) as are computer science background and password attitude (chi-square test: $\chi^2 = 9.24$, $p < .01$). One hypothesis for this distribution could be that computer science studies had historically more male students and that their technical background may have induced awareness of the importance of passwords as a security measure and the promised benefits of password managers.

4.3.2 Comparison of password strength and reuse

Figures 6 and 7 provide a comparison of the password strength and reuse between the two groups. The hatched bars indicate the overall number of passwords per zxcvbn score and reuse category. The plain bars break the number of passwords down by entry method. Participants

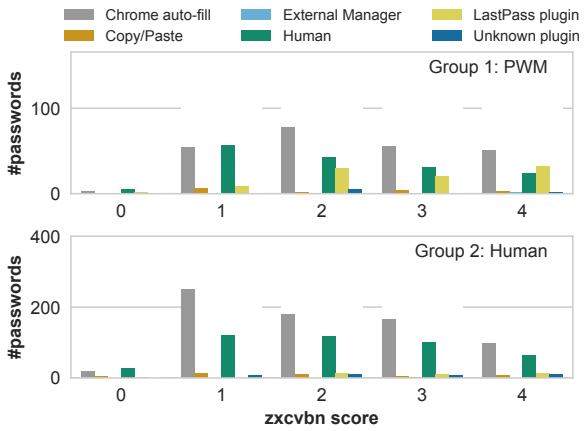

Figure 6: Password strength distribution by participant group and broken down by entry method. Hatched bars show total number of passwords per score. (Note the different y-axis limits)

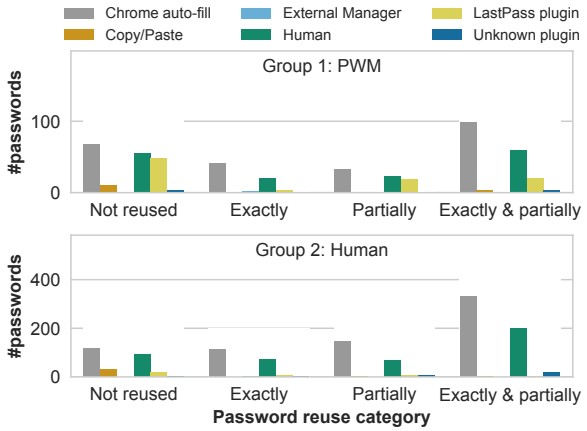

Figure 7: Distribution of reuse categories by participant group, broken down by entry method. Hatched bars show total number of passwords per category. (Note the different y-axis limits)

in $\text{Group}_{\text{PWM}}$ entered in total 522 passwords and participants in $\text{Group}_{\text{Human}}$ entered in total 1245 passwords (both numbers include reused passwords, see Table 6).

For password strength (see Figure 6), neither group contained a noticeable fraction of the weakest passwords (score 0). However, $\text{Group}_{\text{Human}}$ shows a clear tendency towards weaker passwords. For instance, there are almost twice as many score 1 passwords ($n = 390$) than score 4 passwords ($n = 191$). In contrast, the most frequent score for $\text{Group}_{\text{PWM}}$ is 2 ($n = 158$), but the distribution shows a lower kurtosis (e.g., scores 1, 3, and 4 have the frequencies 126, 113, and 114). When breaking the number of passwords down by their entry method, Chrome auto-fill is the dominating entry method for all zxcvbn scores 1–4 in both groups except for score 1 in $\text{Group}_{\text{PWM}}$ where manually entered passwords are most frequent. However,

Entry method	Group 1 (PWM)	Group 2 (Human)
All passwords		
Chrome auto-fill	242 (46.36%)	707 (56.79%)
Human	160 (30.65%)	430 (34.54%)
LastPass plugin	93 (17.82%)	35 (2.81%)
Copy&paste	16 (3.07%)	39 (3.13%)
Unknown plugin	8 (1.53%)	33 (2.65%)
External manager	3 (0.57%)	1 (0.08%)
Σ	522	1245
Unique passwords		
Chrome auto-fill	144 (42.99%)	396 (55.77%)
Human	101 (30.15%)	230 (32.39%)
LastPass plugin	72 (21.49%)	28 (3.94%)
Copy&paste	14 (4.18%)	37 (5.21%)
Unknown plugin	4 (1.19%)	19 (2.68%)
Σ	335	710

Table 6: Distribution of entry methods per participant group.

for $\text{Group}_{\text{PWM}}$ the fraction of passwords entered with LastPass' plugin ($n = 93$ or 17.82% of the passwords) is considerably larger than for $\text{Group}_{\text{Human}}$ ($n = 35$ or 2.81%). In particular, for $\text{Group}_{\text{PWM}}$, passwords entered with LastPass have mostly scores higher than 2 ($n = 82$), where score 4 is the most frequent ($n = 32$).

Regarding password reuse (see Figure 7), the most frequent category is exactly-and-partially reused ($n = 189$ or 36.21% for $\text{Group}_{\text{PWM}}$; $n = 555$ or 44.58% for $\text{Group}_{\text{Human}}$). However, $\text{Group}_{\text{PWM}}$ shows a bimodal distribution in which not-reused passwords are almost as frequent ($n = 187$) as exactly-and-partially reused ones. Further, Chrome auto-fill is the dominating entry method across all reuse categories in both groups. However, when breaking the passwords down by entry method, more than half ($n = 49$ or 52.69%) of the passwords entered with LastPass in $\text{Group}_{\text{PWM}}$ have not been reused in any way. The vast majority of reused passwords can be attributed to manual entry and Chrome auto-fill. In $\text{Group}_{\text{PWM}}$, 335 (64.18%) of the passwords have been reused and 979 (78.63%) of the passwords in $\text{Group}_{\text{Human}}$. Of the 335 reused passwords in $\text{Group}_{\text{PWM}}$, 278 (82.99%) have been entered manually or with Chrome auto-fill. In $\text{Group}_{\text{Human}}$, 926 (74.38%) of the reused passwords were entered manually or with auto-fill.

4.4 Modeling password strength and reuse

In the next step of our analysis we looked at factors influencing the password strength or password reuse among our participants. Our analyses showed that our participants significantly differ from each other in their average password strength (Kruskal-Wallis one-way analysis of variance, $\chi^2 = 779.19, df = 169, p < .001$) as well as in their average probability of password reuse ($\chi^2 = 692.70, df = 169, p < .001$). The underlying reasons for these differences may be factors that we were able to measure, like the password entry methods of the users, as well as latent characteristics of the users, like their personality

or their security awareness. The goal of our further analyses was to show that the effect of the password managers can be shown even beyond these individual differences in password behavior among participants.

One possible way to analyze such a question is a multi-level (aka hierarchical) analysis. This type of regression analysis takes into account the hierarchical structure of our data, where individual password entries are grouped under the corresponding user. Latent, individual differences between users are taken into account in the form of different intercept and/or slope for each user. To get a better understanding of the influencing factors for password strength and reuse, we tested step-wise several regression models. The multi-level models with the studied factors (e.g. entry method) showed a significantly better fit to our data than models that take into account the individual differences between users but do not include the influencing factors we studied. A better fit of the multi-level models was also found in comparison to models that contained the influencing factors but not the individual differences. In the following, we describe our approach to verify the prerequisites for multi-level analysis and our approach to construct the models. Afterwards we report the models for password strength and reuse that fit best to our data.

4.4.1 Correlation analysis

Before constructing the models, we started out with a correlation analysis of the available factors (e.g., password composition, participant group, self-reported website value, etc.). As multi-level models are highly vulnerable to multi-collinearity, detecting and potentially removing strongly correlated variables is essential to prevent inaccurate model estimations, which could lead to false positive results. In our dataset, we detected a very high, significant correlation between zxcvbn scores and password composition, in particular password length, as well as with the NIST and Shannon entropies. Since we consider zxcvbn a more realistic measurement of crackability, we omitted NIST and Shannon entropies from our model. Investigation of zxcvbn showed that zxcvbn rewards lengthy passwords with better scores and that its pattern and l33t speak detection can penalize passwords with digits and special characters. Since zxcvbn is the more interesting factor for us and since it partially contains the effect of the password composition on the prediction, we excluded password composition parameters from our models. Moreover, we noticed that password reuse was strongly correlated with the presence of a lowercase character in the password. A closer inspection of our dataset showed, that our data contained a number of PINs, which were all unique, and that every non-PIN password contains at least one lowercase character. In this situation, including the presence/absence of lowercase characters would result in our model just distinguishing between PINs and non-PINs when predicting password reuse.

	Estimate	Std. Error	z value	Pr(>|z|)
em:chrome	0.07	0.12	0.59	0.56
em:copy/paste	-0.13	0.35	-0.89	0.37
em:lastpass	0.24	0.35	0.69	0.49
em:unknownplugin	1.02	0.34	2.97	<0.01
in-situ:value	0.02	0.05	0.48	0.63
in-situ:strength	0.89	0.07	12.68	<0.001
user:entries	0.02	0.02	0.69	0.49
q9:generator	-0.45	0.67	-0.68	0.50
q14:memorize	-0.24	0.30	-0.79	0.43
q14:analog	0.05	0.29	0.16	0.88
q14:digital	0.09	0.31	0.29	0.77
q14:pwm	-0.16	0.28	-0.57	0.57
em:chrome * q9:gen.	2.30	0.60	3.84	<0.001
em:copy/paste * q9:gen.	3.40	1.22	2.79	<0.01
em:lastpass * q9:gen.	1.83	0.82	2.24	<0.05
em:unknownplugin * q9:gen.	0.22	1.34	0.16	0.87

em: Entry method; q9: Creation strategy; q14: Storage strategy; in-situ: Plugin questionnaire

Table 7: Logistic multi-level regression model predicting zxcvbn score. Estimates are in relation to manually entered passwords by a human. Statistically significant predictors are shaded. Interactions are marked with *.

4.4.2 Constructing the models

For both password reuse and strength prediction, we started with a base model without any explanatory variables, which we iteratively extended with additional predictors. In three steps we included a) entry methods, self-reported value, and strength; b) the number of individually submitted passwords per participant, the creation and storage strategy of the user; in a final step c) the interaction between creation strategy and detected entry method. This approach not only allows us to evaluate the effects of the individual explanatory variables, but also to investigate the interplay between different storage strategies and the password creation strategy. In each iteration we computed the model fit and used log likelihood model fit comparison to check whether the new, more complex model fit the data significantly better than the previous one (see Appendix F). As our final model we picked the one with the best fit that was significantly better in explaining the empirical data than the previous models. This is a well established procedure for model building, e.g., in social sciences and psychological research [32, 25, 9, 15], and allows the creation of models that have the best trade-off of complexity, stability, and fitness.

4.4.3 Zxcvbn score

For the zxcvbn score an ordinal model with all predictors and also the mentioned interaction described our data best. The model is presented in Table 7.

The interactions between the self-reported creation strategy (*q9:generator*; see *Q9* in Appendix A) and the de-

tected entry methods Chrome auto-fill, copy&paste, and LastPass were significant predictors in our model. Those entry methods and also the creation strategy are not significant predictors of password strength on their own. This means that using such a password management tool only leads to significant improvement in the password strength when users also employ some supporting techniques (password generator) for the creation of their passwords. The model might suggest that a general password entry with a plugin (other than LastPass in our dataset) increased the likelihood of a strong password. However, this could be attributed to the high standard error resulting from the minimal data for this entry method.

Moreover, the self-reported password strength was a significant predictor of the measured password strength. This suggests that the users have a very clear view on the strength of the passwords they have entered.

4.4.4 Password reuse

For password reuse a logistical model with all predictors but without interactions described our data best. Table 8 presents our regression model to predict password reuse.

Reuse was significantly influenced by the entry method of the password. In contrast to human entry the odds for reuse were 2.85 time *lower* if the password was entered with LastPass (odds ratio 0.35, predicted probability of reuse with Lastpass = 48.35%) and even 14.29 times *lower* if entered via copy&paste (odds ratio 0.07, predicted probability of reuse with copy&paste = 19.81%). Interestingly, the input via Google Chrome auto-fill even had a negative effect on the uniqueness of the passwords. In contrast to human entry the odds for reuse were 1.65 times *higher* if the password was entered with Chrome auto-fill (odds ratio 1.58, predicted probability of reuse with Chrome auto-fill = 83.72%). A further significant predictor of password reuse is the user's approach to creating passwords. For users who use technical tools to create their passwords (*q9:generator*), the chances that the passwords are *not* reused are 3.70 times higher (odds ratio 0.27, predicted probability of reuse if technical tools are used = 47.36%). In contrast to the models explaining the zxcvbn-score, our data does not indicate the presence of an interaction effect of the password creation strategy on the relation between entry method and password reuse.

In addition, we found a positive relation between the numbers of passwords entered by users and their reuse. In our model, each additional password of the user *increases* the chance that it will be reused by 6% (odds ratio 1.06). This suggests that with increasing numbers of passwords, it becomes more likely that some of them will be reused, which is in line with prior results [27].

We also found the self-reported website value and password strength a statistically significant predictor for reuse [8]. Passwords entered to a website with a higher value for the user were *less* likely to be reused (odds ratio of 0.87) and also passwords that the users considered stronger were *less* likely to be reused (odds ratio of 0.81).

Lastly, users that reported using an analog password storage (*q14:analog*; see *Q14* in Appendix A) were *less* likely to reuse their passwords (odds ratio of 0.62).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.62	0.45	5.80	<0.001
em:chrome	0.46	0.16	2.81	<0.01
em:copy/paste	-2.68	0.41	-6.54	<0.001
em:lastpass	-1.05	0.37	-2.86	<0.01
em:unknownplugin	0.76	0.51	1.51	0.13
in-situ:value	-0.13	0.06	-2.01	<0.05
in-situ:strength	-0.21	0.08	-2.50	<0.05
user:entries	0.06	0.02	2.67	<0.01
q9:generator	-1.31	0.40	-3.24	<0.01
q14:memorize	0.22	0.25	0.88	0.38
q14:analog	-0.48	0.24	-1.98	<0.05
q14:digital	-0.18	0.26	-0.70	0.48
q14:pwm	-0.07	0.24	-0.30	0.76

em: Entry method; q9: Creation strategy; q14: Storage strategy; in-situ: Plugin questionnaire

Table 8: Logistic multi-level regression model predicting reuse. Estimates are in relation to manually entered passwords by a human and refer to the corresponding logit transformed odds ratios. Statistically significant predictors are shaded.

5 Discussion

5.1 Password Managers' Impact

In general, our participants showed very similar password strength and reuse characteristics as in prior studies [51, 66] and our analysis could also reaffirm prior results, such as rampant password reuse.

Our study adds novel insights to the existing literature by considering the exact password entry methods and by painting a more complete picture by considering the users' password creation strategies. We found that almost all participants entered passwords with more than one entry method. Further, we discovered that every entry method showed reused passwords, although the ratio of reused passwords differs significantly between the entry methods. More than 80% of Chrome auto-filled passwords were reused, while only 47% of the passwords entered with LastPass' plugins were reused in some way, and even only 22% of the copied/pasted passwords. Similarly, we noticed that low-strength passwords have been entered with all entry methods, where LastPass had on average the strongest passwords (mean zxcvbn score of 2.80). Interestingly, manually entered passwords and Chrome auto-filled passwords were on a par with the overall password strength but showed above average reuse rates.

For our participants, we discovered a dichotomous distribution of self-reported creation strategies. Participants indicated using a password generator right now or in the

recent past, or clearly described mental algorithms and similar methods for human-generated passwords. Taking a differentiated view based on the creation strategies, we find that users of a password generator are closer to a desirable situation with stronger, less reused passwords, although being far from ideal. Only a negligible fraction of participants mentioned analog tools or alternative strategies (like two-factor authentication). Two-factor authentication (2FA), in particular, might be a valuable feature for future, targeted investigations, but for our study, we excluded 2FA since most (major) websites still lack support for 2FA and even for services offering 2FA support the userbase has only little adapted to it [46].

Using regression modeling, we put our data together to a more complete view of password managers' influence. Our models suggest that the interaction between the creation strategy and the entry methods has a significant influence on the password strength. If the passwords are entered with technical support (auto-fill, password manager plugin, or copy&paste), this results in stronger passwords under the condition that technical means were already used when generating the passwords in the first place. Thus, password managers that provide users with password creation features indeed positively influence the overall password strength in the ecosystem. All the more, it is curious that Chrome, as the primary tool to access websites, has the password generation feature disabled by default [7]. Future work could investigate and compare Apple's walled-garden ecosystem, where the Safari browser has this feature enabled by default. Another, maybe surprising, result of our modeling is that the self-reported password strength was a significant predictor for the measured password strength, suggesting that our participants have a clear view on the strength of the entered password. This is in contradiction to prior results of lab studies, like [62], and we think it is worth investigating why users in the wild are so much better at judging their own password strength.

Our models further suggest, that the use of password generators and the website value also significantly reduced the chance of password reuse. More interestingly, however, is that the password storage strategies have different influence independently of an interaction with the creation strategy. Using a password manager plugin or copy&pasting passwords reduced password reuse, while Chrome's auto-fill aggravated reuse. In other words, we observed that users were able to *manually* create more unique passwords when managing their passwords digitally or with a manager, but not with Chrome auto-fill.

The benefit of password managers is also put into better perspective when considering particular strategies in our $\text{Group}_{\text{Human}}$. We noticed that users tend to have a "self-centered" view when it comes to password uniqueness (i.e., personal vs. global), but are unaware of the fact that an attacker would not be concerned with *personal* uniqueness of passwords. A large fraction of users reported to *"come up with [a password **they**] have never used before"* or to *"try to think of something that [**they**] have never used before."* Those results also align with prior studies [56, 52, 38]. While our participants were able to correctly judge the strength of their entered passwords, their creation strategies indicate an incomplete understanding of uniqueness. In the future, the influence of services like *Have I Been Pwned*[6], which are increasingly integrated into password creation forms and managers, onto the users' understanding of uniqueness and password reuse could be studied.

Another interesting question that comes from our study is why users of password managers ($\text{Group}_{\text{PWM}}$) still reuse passwords and employ weak passwords. There could be different reasons, on which we can only speculate at this point. For instance, users might employ a default password for low-value websites, however, we could not find any evidence in our data set for a correlation between website value and strength or reuse for $\text{Group}_{\text{PWM}}$. Another explanation could be that those passwords existed prior to starting using a password manager and were never replaced (e.g., LastPass introduced features[7] for automatically updating "legacy passwords" in 2014), or maybe those are passwords that are also required on devices not managed by the user (e.g., computer pool devices at the university). Thus, we think it would interesting to investigate this question more focused.

Further, in light of the high relevance of copy&paste for strong and unique passwords, our results can also underline the "Cobra effect" [35, 36, 47] of disabling paste functionality for password fields on websites to encourage the use of 2FA or password managers. Based on our data, we consider those users who mainly use copy&paste to enter their passwords to be a very interesting subgroup that would be worth further research (e.g., which storage strategies are exactly pursued or motivation to abstain from managers). Unfortunately, there were too few copy&paste users in our current dataset to make any further reliable statements about them separately.

In summary, password managers indeed provide benefits to the users' password strength and uniqueness. Although both benefits can be achieved separately, our data suggest that the integrated workflow of 3rd party password managers for generation and storage provides the highest benefits. More troublesome is that our results suggest that the most widely used manager, Chrome's auto-filling feature, has only a positive effect on password strength when used in conjunction with an additional generator and even shows an aggravating effect on password

[6] `https://haveibeenpwned.com`

[7] `https://blog.lastpass.com/2014/12/introducing-auto-password-changing-with.html/`

reuse. The conclusion we draw from this, is that research should investigate how such integrated workflows can be brought to more users, e.g., by better understanding and tackling the reasons why users abstain from using password managers in the first place.

5.2 Threats to validity

As with other human-subject and field studies, we cannot eliminate all threats to the validity of our study. We targeted Google Chrome users, which had in general [6] the highest market share, also among our survey participants. Further, we recruited only experienced US workers on Amazon MTurk, which might not be representative for any population or other cultures (external validity), however, our demographics and password statistics show alignment with prior studies. Furthermore, we collected our data *in the wild*, which yields a high ecological validity and avoids common problems of password lab studies [41], but on the downside does not give control over all variables (internal validity). We asked our participants to behave naturally and also tried to encourage this behavior through transparency, availability, and above average payment, however, like closest related work [66, 51] we cannot exclude that some participants behaved unusually.

6 Conclusion

Passwords are the de-facto authentication scheme on the internet. Since users are very often referred to password managers as a technical solution for creating guessing-resistant, unique passwords, it is important to understand the impact that those managers *actually* have on users' passwords. Studying this impact requires in the first place an approach that is able to detect potential effects of managers. This paper's first contribution is an addition to the existing methodology, which for the first time allowed measuring the influence of managers on password strength and reuse *in the wild*. By combining insights into users' password storage and creation strategies within situ collected password metrics, we create a more complete view of passwords. We applied this methodology in a study with 170 workers from Amazon MTurk and were able to show that password managers indeed influence password security. More importantly, we were further able to study factors that affect the password strength and reuse. We found that users that rely on technical support for password creation had both stronger and more unique passwords, even if entered through other channels than a manager. We also found that Chrome's auto-fill option aggravated the password reuse problem. For future work, we see different alleys. For instance, investigating how different, even novel forms of password generators can be integrated with users' strategies. Moreover, one could apply our approach to explore password managers' influence in other ecosystems, such as Apple's walled-garden ecosystem or mobile password managers.

Acknowledgements. We like to thank our anonymous reviewers for their valuable comments and feedback.

This work was partially supported by the German Federal Ministry of Education and Research (BMBF) through funding for the Center for IT-Security, Privacy, and Accountability (CISPA) (VFIT/FKZ: 16KIS0345).

References

[1] 1Password. `https://1password.com/`.

[2] Alexa Web Information Service: Developer Guide (API Version 2005-07-11). `https://docs.aws.amazon.com/AlexaWebInfoService/latest/`.

[3] Github: dropbox/zxcvbn. `https://github.com/dropbox/zxcvbn`.

[4] Kaspersky Password Manager. `https://www.kaspersky.com/password-manager`.

[5] LastPass. `https://www.lastpass.com`.

[6] W3Counter: Browser & Platform Market Share (November 2017). `https://www.w3counter.com/globalstats.php`.

[7] How To Enable and Use Password Generator in Google Chrome. `https://edgetalk.net/enable-use-password-generator-google-chrome/`, Sept. 2016.

[8] BAILEY, D. V., DÜRMUTH, M., AND PAAR, C. Statistics on password re-use and adaptive strength for financial accounts. In *Proc. 9th International Conference on Security and Cryptography for Networks (SCN'14)* (2014).

[9] BATES, D., MAECHLER, M., BOLKER, B., WALKER, S., CHRISTENSEN, R. H. B., SINGMANN, H., DAI, B., GROTHENDIECK, G., AND GREEN, P. lme4: Linear mixed-effects models using 'eigen' and s4. `https://cran.r-project.org/web/packages/lme4/index.html`.

[10] BONNEAU, J. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *Proc. 33rd IEEE Symposium on Security and Privacy (SP '12)* (2012), IEEE Computer Society.

[11] BONNEAU, J., HERLEY, C., OORSCHOT, P. C. V., AND STAJANO, F. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Proc. 33rd IEEE Symposium on Security and Privacy (SP '12)* (2012), IEEE Computer Society.

[12] BONNEAU, J., AND PREIBUSCH, S. The password thicket: technical and market failures in human authentication on the web. In *9th Workshop on the Economocs of Info Security (WEIS'10)* (2010).

[13] CASTELLUCCIA, C., DÜRMUTH, M., AND PERITO, D. Adaptive password-strength meters from markov models. In *Proc. 19th Annual Network and Distributed System Security Symposium (NDSS '12)* (2012), The Internet Society.

[14] CHIASSON, S., VAN OORSCHOT, P. C., AND BIDDLE, R. A usability study and critique of two password managers. In *Proc. 15th USENIX Security Symposium (SEC '06)* (2006), USENIX Association.

[15] CHRISTENSEN, R. H. B. ordinal: Regression models for ordinal data. `https://cran.r-project.org/web/packages/ordinal/index.html`.

[16] DAS, A., BONNEAU, J., CAESAR, M., BORISOV, N., AND WANG, X. The tangled web of password reuse. In *Proc. 21th Annual Network and Distributed System Security Symposium (NDSS '14)* (2014), The Internet Society.

[17] DE CARNÉ DE CARNAVALET, X., AND MANNAN, M. From very weak to very strong: Analyzing password-strength meters. In *Proc. 21th Annual Network and Distributed System Security Symposium (NDSS '14)* (2014), The Internet Society.

[18] DELL'AMICO, M., MICHIARDI, P., AND ROUDIER, Y. Password strength: An empirical analysis. In *Proc. 29th Conference on Information Communications (INFOCOM'10)* (2010), IEEE Press.

[19] DÜRMUTH, M., ANGELSTORF, F., CASTELLUCCIA, C., PERITO, D., AND CHAABANE, A. Omen: Faster password guessing using an ordered markov enumerator. In *Proc. 7th International Symposium on Engineering Secure Software and Systems (ESSoS 2015)* (2015), Springer.

[20] DYNAMO WIKI. Guidelines for academic requesters (version 2.0). `http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters`. Last visited: 11/10/17.

[21] E. SHANNON, C. Prediction and entropy of printed english.

[22] EGELMAN, S., AND PEER, E. Scaling the security wall: Developing a security behavior intentions scale (SeBIS). In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'15)* (2015), ACM.

[23] EGELMAN, S., SOTIRAKOPOULOS, A., MUSLUKHOV, I., BEZNOSOV, K., AND HERLEY, C. Does my password go up to eleven?: The impact of password meters on password selection. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'13)* (2013), ACM.

[24] FAHL, S., HARBACH, M., ACAR, Y., AND SMITH, M. On the ecological validity of a password study. In *Proc. 9th Symposium on Usable Privacy and Security (SOUPS'13)* (2013), ACM.

[25] FIELD, A., AND MILES, J. *Discovering Statistics Using R*. Sage Publications Ltd., 5 2012.

[26] FLORENCIO, D., AND HERLEY, C. A large-scale study of web password habits. In *Proc. 16th International Conference on World Wide Web (WWW'07)* (2007), ACM.

[27] GAW, S., AND FELTEN, E. W. Password management strategies for online accounts. In *Proc. 2nd Symposium on Usable Privacy and Security (SOUPS'06)* (2006), ACM.

[28] GRASSI, P. A., FENTON, J. L., NEWTON, E. M., PERLNER, R. A., REGENSCHEID, A. R., BURR, W. E., AND RICHER, J. P. NIST SP800–63B: Digital authentication guideline (Authentication and Lifecycle Management), June 2017. Last visited: 10/11/17.

[29] HAYASHI, E., AND HONG, J. A diary study of password usage in daily life. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'11)* (2011), ACM.

[30] HERLEY, C., AND VAN OORSCHOT, P. A research agenda acknowledging the persistence of passwords. *IEEE Security and Privacy 10*, 1 (Jan. 2012), 28–36.

[31] HITAJ, B., GASTI, P., ATENIESE, G., AND PÉREZ-CRUZ, F. Passgan: A deep learning approach for password guessing. *CoRR abs/1709.00440* (2017).

[32] HOX, J. J., MOERBEEK, M., AND VAN DE SCHOOT, R. *Multilevel Analysis: Techniques and Applications, Third Edition (Quantitative Methodology)*, 3 ed. Quantitative Methodology Series, 9 2017. An optional note.

[33] HUANG, J. L., BOWLING, N. A., LIU, M., AND LI, Y. Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology 30*, 2 (2015), 299–311.

[34] HUNT, T. The science of password selection. `https://www.troyhunt.com/science-of-password-selection/`, July 2011.

[35] HUNT, T. The "cobra effect" that is disabling paste on password fields. `https://www.troyhunt.com/the-cobra-effect-that-is-disabling/`, May 2014.

[36] HUNT, T. It's not about "supporting password managers", it's about not consciously breaking security. `https://www.troyhunt.com/its-not-about-supporting-password/`, July 2015.

[37] HUNT, T. Password reuse, credential stuffing and another billion records in have i been pwned. `https://www.troyhunt.com/password-reuse-credential-stuffing-and-another-1-billion-records-in-have-i-been-pwned/`, May 2017.

[38] INGLESANT, P. G., AND SASSE, M. A. The true cost of unusable password policies: Password use in the wild. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'10)* (2010), ACM.

[39] KAROLE, A., SAXENA, N., AND CHRISTIN, N. A comparative usability evaluation of traditional password managers. In *Proceedings of the 13th International Conference on Information Security and Cryptology* (2011), Springer-Verlag.

[40] KELLEY, P. G., KOMANDURI, S., MAZUREK, M. L., SHAY, R., VIDAS, T., BAUER, L., CHRISTIN, N., CRANOR, L. F., AND LOPEZ, J. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *Proc. 33rd IEEE Symposium on Security and Privacy (SP '12)* (2012), IEEE Computer Society.

[41] KOMANDURI, S., SHAY, R., KELLEY, P. G., MAZUREK, M. L., BAUER, L., CHRISTIN, N., CRANOR, L. F., AND EGELMAN, S. Of passwords and people: Measuring the effect of password-composition policies. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'11)* (2011), ACM.

[42] KUMARAGURU, P., AND CRANOR, L. F. Privacy Indexes: A Survey of Westin's Studies. Tech. rep., 2005.

[43] LI, Z., HE, W., AKHAWE, D., AND SONG, D. The emperor's new password manager: Security analysis of web-based password managers. In *Proc. 23rd USENIX Security Symposium (SEC' 14)* (2014), USENIX Association.

[44] MCMILLAN, R. The world's first computer password? it was useless too. `https://www.wired.com/2012/01/computer-password/`, 2012.

[45] MELICHER, W., UR, B., SEGRETI, S. M., KOMANDURI, S., BAUER, L., CHRISTIN, N., AND CRANOR, L. F. Fast, lean, and accurate: Modeling password guessability using neural networks. In *Proc. 24th USENIX Security Symposium (SEC' 16)* (2016), USENIX Association.

[46] MILKA, G. Anatomy of account takeover. In *Enigma 2018* (2018), USENIX Association.

[47] MOORE, P. Don't let them paste passwords... `https://paul.reviews/dont-let-them-paste-passwords/`, July 2015.

[48] MORRIS, R., AND THOMPSON, K. Password security: A case history. *Commun. ACM 22*, 11 (Nov. 1979), 594–597.

[49] NOTOATMODJO, G., AND THOMBORSON, C. Passwords and perceptions. In *Proc. 7th Australasian Conference on Information Security - Volume 98* (2009), Australian Computer Society, Inc.

[50] PARSONS, K., CALIC, D., PATTINSON, M., BUTAVICIUS, M., MCCORMAC, A., AND ZWAANS, T. The human aspects of information security questionnaire (HAIS-Q). *Comput. Secur. 66*, C (May 2017), 40–51.

[51] PEARMAN, S., THOMAS, J., NAEINI, P. E., HABIB, H., BAUER, L., CHRISTIN, N., CRANOR, L. F., EGELMAN, S., AND FORGET, A. Let's go in for a closer look: Observing passwords in their natural habitat. In *Proc. 24th ACM Conference on Computer and Communication Security (CCS '17)* (2017), ACM.

[52] RINN, C., SUMMERS, K., RHODES, E., VIROTHAISAKUN, J., AND CHISNELL, D. Password creation strategies across high- and low-literacy web users. In *Proc. 78th ASIS&T Annual Meeting (ASIST'15)* (2015), American Society for Information Science.

[53] RUBENKING, N. J. The best password managers of 2017. `http://uk.pcmag.com/password-managers-products/4296/guide/the-best-password-managers-of-2017`, Nov. 2017.

[54] SHAY, R., KOMANDURI, S., KELLEY, P. G., LEON, P. G., MAZUREK, M. L., BAUER, L., CHRISTIN, N., AND CRANOR, L. F. Encountering stronger password requirements: User attitudes and behaviors. In *Proc. 6th Symposium on Usable Privacy and Security (SOUPS'10)* (2010), ACM.

[55] SILVER, D., JANA, S., BONEH, D., CHEN, E., AND JACKSON, C. Password managers: Attacks and defenses. In *Proc. 23rd USENIX Security Symposium (SEC' 14)* (2014), USENIX Association.

[56] STOBERT, E., AND BIDDLE, R. The password life cycle: User behaviour in managing passwords. In *Proc. 10th Symposium on Usable Privacy and Security (SOUPS'14)* (2014), USENIX Association.

[57] STOCK, B., AND JOHNS, M. Protecting users against xss-based password manager abuse. In *Proc. 9th ACM Symposium on Information, Computer and Communication Security (ASIACCS '14)* (2014), ACM.

[58] TAM, L., GLASSMAN, M., AND VANDENWAUVER, M. The psychology of password management: a tradeoff between security and convenience. *Behaviour & Information Technology 29*, 3 (2010), 233–244.

[59] THE UNIVERSITY OF CHICAGO – IT SERVICES. Strengthen your passwords or passphrases and keep them secure. `https://uchicago.service-now.com/it?id=kb_article&kb=KB00015347`, Oct. 2017.

[60] UR, B., ALFIERI, F., AUNG, M., BAUER, L., CHRISTIN, N., COLNAGO, J., CRANOR, L. F., DIXON, H., NAEINI, P. E., HABIB, H., JOHNSON, N., AND MELICHER, W. Design and evaluation of a data-driven password meter. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'17)* (2017), ACM.

[61] UR, B., KELLEY, P. G., KOMANDURI, S., LEE, J., MAASS, M., MAZUREK, M. L., PASSARO, T., SHAY, R., VIDAS, T., BAUER, L., CHRISTIN, N., AND CRANOR, L. F. How does your password measure up? the effect of strength meters on password creation. In *Proc. 21st USENIX Security Symposium (SEC '12)* (2012), USENIX Association.

[62] UR, B., NOMA, F., BEES, J., SEGRETI, S. M., SHAY, R., BAUER, L., CHRISTIN, N., AND CRANOR, L. F. "i added '!' at the end to make it secure": Observing password creation in the lab. In *Proc. 11th Symposium on Usable Privacy and Security (SOUPS'15)* (2015), USENIX Association.

[63] UR, B., SEGRETI, S. M., BAUER, L., CHRISTIN, N., CRANOR, L. F., KOMANDURI, S., KURILOVA, D., MAZUREK, M. L., MELICHER, W., AND SHAY, R. Measuring real-world accuracies and biases in modeling password guessability. In *Proc. 24th USENIX Security Symposium (SEC' 15)* (2015), USENIX Association.

[64] U.S. CENSUS BUREAU. 2010 Census National Summary File of Redistricting Data. `https://www.census.gov/2010census/data/`, 2011.

[65] WANG, D., AND WANG, P. The emperor's new password creation policies: An evaluation of leading web services and the effect of role in resisting against online guessing. In *Proc. 20th European Symposium on Research in Computer Security (ESORICS'15)* (2015), Springer.

[66] WASH, R., RADER, E., BERMAN, R., AND WELLMER, Z. Understanding password choices: How frequently entered passwords are re-used across websites. In *Proc. 12th Symposium on Usable Privacy and Security (SOUPS'16)* (2016), USENIX Association.

[67] WEIR, M., AGGARWAL, S., COLLINS, M., AND STERN, H. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proc. 17th ACM Conference on Computer and Communication Security (CCS '10)* (2010), ACM.

[68] WHEELER, D. L. zxcvbn: Low-budget password strength estimation. In *Proc. 24th USENIX Security Symposium (SEC' 16)* (2016), USENIX Association.

[69] WOODRUFF, A., PIHUR, V., CONSOLVO, S., BRANDIMARTE, L., AND ACQUISTI, A. Would a privacy fundamentalist sell their DNA for $1000...if nothing bad happened as a result? the westin categories, behavioral intentions, and consequences. In *Proc. 10th Symposium on Usable Privacy and Security (SOUPS'14)* (2014), USENIX Association.

[70] YAN, J., BLACKWELL, A., ANDERSON, R., AND GRANT, A. Password memorability and security: Empirical results. *IEEE Security and Privacy 2*, 5 (Sept. 2004), 25–31.

[71] ZHAO, R., YUE, C., AND SUN, K. Vulnerability and risk analysis of two commercial browser and cloud based password managers. *ASE Science Journal 1*, 4 (2013), 1–15.

A Sampling Survey Questions

Q1: *For each of the following statements, how strongly do you agree or disagree?*
a1: Consumer have lost all control over how personal information is collected and used by companies.
a2: Most businesses handle the personal information they collect about consumers in a proper and confidential way.
a3: Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.
(i) Strongly disagree, (ii) Somewhat disagree, (iii) Somewhat agree, (iv) Strongly agree
Q2: *On how many different Internet sites do you have a user account that is secured with a password? (If you are not sure about the number please estimate the number)* (FreeText)
Q3: *Has ever one of your passwords been leaked or been stolen?*
(i) Yes, (ii) No, (iii) I am not aware of that, (iv) I do not care
Q4: *How strongly do you agree or disagree:?*
b1. Passwords are useless, because hackers can steal my data either way. (i) Strongly disagree, (ii) Somewhat disagree, (iii) Somewhat agree, (iv) Strongly agree
b2. I don't care about my passwords' strength, because I don't have anything to hide. (i) Strongly disagree, (ii) Somewhat disagree, (iii) Somewhat agree, (iv) Strongly agree
Q5: *What characterizes in your opinion a strong/secure password?* (FreeText)
Q6: *Please rate the strength of the following passwords?*

c1. thHisiSaSecUrePassWord
c2. Pa$sWordsk123
c3. AiWuutaiveep9j
c4. !@#$%&̂*()
c5. 12/07/2017
(i) Very weak, (ii) Weak, (iii) Moderate strength, (iv) Strong, (v) Very strong

Q7: *I have never used a computer?* (i) I have never, (ii) I do

Q8: *How would you rate your ability to create strong passwords?*
(i) 5 (high ability), (ii) 4, (iii) 3, (iv) 2, (v) 1 (low ability)

Q9: *How do you proceed if you have to create a new password? (What is your strategy?)* (FreeText)

Q10: *I try to create secure passwords.....*
(i) for all my accounts and websites, (ii) for my email accounts, (iii) for online shopping, (iv) for online booking/reservation, (v) for social networks, (vi) No answer, (vii) Other

Q11: *I make a point of changing my passwords on websites that are critical to my privacy every..... (choose the closest match)*
(i) Day, (ii) Week, (iii) Two weeks, (iv) Month, (v) 6 month, (vi) Year, (vii) Never, (viii) Other

Q12: *Do you use the same password for different email accounts, websites, or devices?* (i) Yes, (ii) No

Q13: *Do you use any of the following strategies for creating your password or part of your password, anywhere, at any time in the last year...* (i) I used the name of celebrities as a password or as a part of a password, (ii) I used the name of family members as a password or as a part of a password, (iii) I used literature (book, poetry, etc.) as a password or as a part of a password, (iv) I used familiar numbers (street address, employee number, etc) as a password or as a part of a password, (v) I used random characters as a password, (vi) I used a password manager to generate passwords, (vii) No answer, (viii) Other

Q14: *How do you remember all of your passwords?* (i) I write them down on paper (notebook, day planner, etc), (ii) I try to remember them (human memory), (iii) I use computer files (Word document, Excel sheet, text file, etc), (iv) I use encrypted computer files (e.g. CryptoPad), (v) I store my passwords on my mobile phone or PDA, (vi) I use 3rd party password manager (save in extra program, e.g. LastPass, keepass, 1Password, etc.), (vii) I use website cookies (Website checkbox: "Remember my password on this computer"), (viii) I use the same password for more than one purpose, (ix) I use browser built-in password manager (i.e saved in browser), (x) I use a variation of a past password (eg. password1 and then password2 and then password3, etc.), (xi) No answer, (xii) Other

Q15: *Have you ever used a computer program to generate your passwords?* (i) Yes, (ii) No

Q16: *When creating a new password, which do you regard as most important: choosing a password that is easy to remember for future use (ease of remembering) or the password's security?*
(i) Always ease of remembering, (ii) Mostly ease of remembering, (iii) Mostly security, (iv) Always security, (v) Other

Q17: *When you create a new password, which of the following factors do you consider? The password ....*
(i) does not contain dictionary words, (ii) is in a foreign (non-English) language, (iii) is not related to the site (i.e., the name of the site), (iv) includes numbers, (v) includes special characters (e.g. "&" or "!"), (vi) is at least eight (8) characters long, (vii) None of the above: I didn't think about it, (viii) No answer, (ix) Other

Q18: *My home planet is Earth?* (i) Yes, (ii) No

Q19: *Do you use the "save password" feature of your browser?* (i) Yes, (ii) No

Q20: *Do you use any kind of extra password manager program (for instance, LastPass, 1Password, Keepass, Dashlane, etc.)?* (i) Yes, (ii) No

Q21: *Which password manager(s) do you use? (You can write one name per line)* (FreeText)

Q22: *Please give us a short description of your impression of using your browser's password saving feature and/or of using extra password managers* (FreeText)

Q23: *How many passwords do you keep in your password manager(s) and browser's saved passwords? (if you don't know the exact number, please estimate the number)* (FreeText)

B Zxcvbn Score

To better understand zxcvbn's scoring, we used zxcvbn to score 200 million unique passwords collected from `hashes.org`, where we measured the zxcvbn score and the corresponding guesses in log10. The results in Table 9 show that each score has a corresponding cutoff for guesses, e.g., score 2 requires between 10^3–10^6 guesses.

Score	#Passwords	Mean	SD	Min	25%	50%	75%	Max
0	122,296	2.69	0.42	0.30	2.48	2.92	3.00	3.00
1	34,496,960	5.34	0.59	3.00	5.00	5.44	5.87	6.00
2	69,090,776	7.15	0.66	6.00	6.61	7.00	7.87	8.00
3	57,256,840	8.87	0.65	8.00	8.28	8.87	9.36	10.00
4	39,789,207	12.51	2.29	10.00	11.00	12.00	13.36	32.00

Table 9: Zxcvbn scores and estimated no. of guesses (in log10) for 200 million unique passwords from *hashes.org*.

C Website category vs. website value

Commonly the website category is used as a proxy for the website value. Since we collected both, we can provide insights into this general assumption. Figure 8 shows the self-reported value per domain. For instance, in >70% of logged passwords for a financial domain, the user reported a very high value for that domain. Similarly, in >60% of a logged passwords for news websites, the users (strongly) disagreed that this domain has a high value.

D Known Password Manager Plugins

Chrome plugins are identified through a 32 characters long UUID that can be retrieved from Google's Chrome Web Store. Table 10 lists the password manager plugins that our study plugin can detect based on their UUID. Plugins not in this list are reported as *"Unknown plugin."*

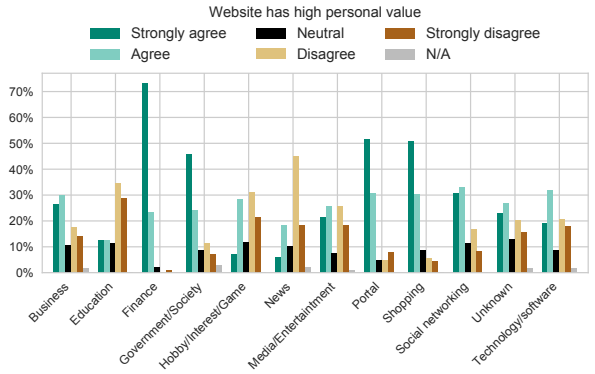

Figure 8: Self-reported website value per website category

Name	UUID
Dashlane	`fdjamakpfbbddfjaooikfcpapjohcfmg`
LastPass	`hdokiejnpimakedhajhdlcegeplioahd`
1Password	`aomjjhallfgjeglblehebfpbcfeobpgk`
Roboform	`pnlccmojcmeohlpggmfnbbiapkmbliob`
Enpass	`kmcfomidfpdkfieipokbalgegidffkal`
Zoho Vault	`igkpcodhieompeloncfnbekccinhapdb`
Norton Identity Safe	`iikflkcanblccfahdhdonehdalibjnif`
KeePass	`ompiailgknfdndiefoaoiligalphfdae`

Table 10: UUIDs of plugins detected by our study plugin

E Demographics of participant groups

Table 11 presents the demographics of our two participant groups according their password creation strategies.

F Model fit

All models in the building process were compared according to the corresponding akaike information criterion (AIC), which is an estimator of the relative quality of statistical models for a given set of data. Additionally, the models were statistically compared using likelihood-ratio tests, which were evaluated using a Chi-squared distribution. The final model is selected based on AIC as well as their ability to describe the empirical data better than the previous models. Tables 12 and 13 present the goodness of fit for the relevant steps in the model building process.

	Human	PWM
Number of participants		
	121	49
Gender		
Female	59 (48.76%)	14 (28.57%)
Male	62 (51.24%)	35 (71.43%)
Age group		
18–30	48 (39.67%)	16 (32.65%)
31–40	39 (32.23%)	24 (48.98%)
41–50	27 (22.31%)	5 (10.20%)
51–60	5 (4.13%)	3 (6.12%)
61–70	2 (1.65%)	0
≥71	0 0	1 (2.04%)
Computer science background		
Yes	10 (8.26%)	17 (34.69%)
No	111 (91.74%)	32 (65.13%)
Education level		
Less than high school	0	1 (2.04%)
High school graduate	22 (18.18%)	4 (8.16%)
Some college, no degree	28 (23.14%)	6 (12.24%)
Associate's degree	27 (22.31%)	7 (14.29%)
Bachelor degree	35 (28.93%)	27 (55.10%)
Ph.D	0	1 (2.04%)
Graduate/prof. degree	9 (7.44%)	3 (6.12%)
Ethnicity		
White/Caucasian	91 (75.21%)	32 (65.31%)
Black/African American	15 (12.40%)	10 (20.41%)
Asian	5 (4.13%)	4 (8.16%)
Hispanic/Latino	10 (8.26%)	2 (4.08%)
Multiracial	0	1 (2.04%)
Privacy concern (Westin index)		
Privacy fanatic	45 (37.19%)	21 (42.86%)
Privacy unconcerned	15 (12.40%)	16 (32.65%)
Privacy pragmatist	61 (50.41%)	12 (24.49%)
Attitude about passwords		
Pessimist	1 (0.83%)	2 (4.08%)
Optimist	88 (72.73%)	44 (89.80%)
Conflicted	32 (26.45%)	3 (6.12%)
Prior password leaked experienced		
No	53 (43.80%)	19 (38.78%)
Yes	44 (36.36%)	14 (28.57%)
Not aware of	24 (19.83%)	16 (32.65%)

Table 11: Demographics of our two participant categories.

	AIC	logLik	df	Pr(>Chisq)
simple regression	5080.6	-2536.3		
multi-level base	4536.7	-2263.4	1	<0.001
+ login level	4316.3	-2147.1	6	<0.001
+ user level	4320.4	-2143.2	6	0.2494034
+ interactions	4309.5	-2133.7	4	<0.001

Table 12: Goodness of fit for models predicting ZCVBN scores

	AIC	logLik	Df	Pr(>Chisq)
simple regression	1959.7	-978.84		
multi-level base	1794.6	-895.28	1	< 0.001
+ login level	1694.9	-839.46	6	< 0.001
+ user level	1684.7	-828.37	6	<0.01
+ interactions	1687.6	-825.80	4	0.27351

Table 13: Goodness of fit for models predicting password reuse