Do Password Managers Improve Password Hygiene?

Do Password Managers Improve Password Hygiene?

David W. Ng
University of California, Berkeley
davidng@berkeley.edu

Jacky Ho
University of California, Berkeley
jacky.ho@berkeley.edu

Christian Hercules
University of California, Berkeley
cxhercules@berkeley.edu

Cristian Bravo-Lillo
Fintual
cristian@fintual.com

Stuart Schechter
Harvard University & DiceKeys
stuart@post.harvard.edu

Abstract

Password managers purport to increase users' security by improving password hygiene: generating unique random passwords when users create new accounts, replacing users' weak and reused passwords, and determining which sites are safe to send each password to. We conducted a study of password manager users to measure their password hygiene. While structured as a survey, we asked participants to upload anonymized screenshots with four hygiene statistics calculated by their password managers: the number of passwords their password manager classified as (1) reused, (2) weak, and (3) compromised, as well as (4) the total number of passwords stored.

Regardless of password manager, most participants had weak or reused passwords that they confessed they "should replace." Nearly a third (30%) had passwords that their password managers knew to be compromised and that the participants confessed they should replace. When creating new accounts, more than a third of participants using third-party password managers (29/81, 36%) and the majority of those using Chrome's password manager (48/61, 79%) preferred to "create a password myself" rather than "allow my password manager to create a random password for me."

We also asked how participants had generated the all-important "master" password used to protect the passwords stored by their password manager. A quarter (19/81) of those using third-party password managers confessed to re-using an existing password.

1 Introduction

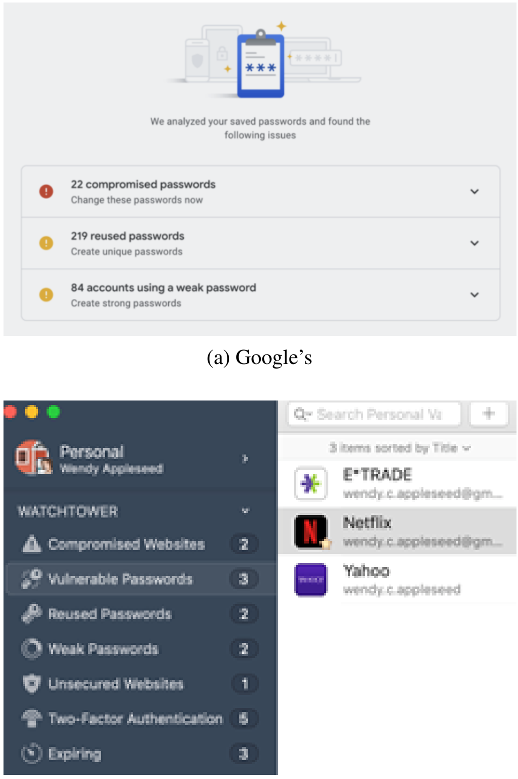
Password managers promise to *"solve poor password habits"* and *"protect you from breaches and other threats"* [1] leading to *"a safer life online"* [6]. And they can... if users replace their weak and re-used passwords with unique passwords randomly generated by their password managers. Those who adopt password managers, but who continue to rely on the weak and re-used passwords that they have memorized, may be no safer. We conducted the first large-scale investigation, across a wide swath of products, to investigate how frequently those who adopt password managers continue to suffer from poor password hygiene: relying on weak, re-used, and compromised passwords for accounts they confessed should have better passwords.

Such research is desperately needed because the developers of password managers do not measure whether their products are used as prescribed. Instrumenting products to report telemetry with hygiene statistics could be seen as antithetical to developers' generous privacy promises: that their *"first priority is safeguarding your data"* [10] and that *"your data is yours, and we don't want to know anything about it"* [1]. While developers privacy-preserving telemetry that users could opt into, they likely have no incentive to do so. The narrative that password managers improve security is being delivered by fawning articles in the popular press (many with affiliate links) as well as advocacy from security practitioners, academics, and influencers. They might reasonably expect that measured hygiene would fail to meet the high expectations of that narrative and could end up subverting it.

As independent researchers, we are at a disadvantage when investigating password manager use. To collect telemetry, we would need to get participants' consent to instrument the software they rely on to protect their most valuable secrets (as was attempted by Lyastani*et al.* [11]) which could bias participation to exclude those rationally unwilling to install software with access to monitor their passwords.

While we could ask users of password managers to report hygiene behaviors and statistics to us, self-reported password behaviors may diverge from true behaviors, as Wash *et al.* observed when asking participants if they re-used passwords or included special characters in passwords [16]. Collecting and entering password hygiene statistics requires participants' time, effort, and trust, and some participants may not report

Version 2022-12-24



We analyzed your saved passwords and found the following issues

22 compromised passwords
Change these passwords now

219 reused passwords
Create unique passwords

84 accounts using a weak password
Create strong passwords

(a) Google's

Personal
Wendy Appleseed

WATCHTOWER

Compromised Websites	2
Vulnerable Passwords	3
Reused Passwords	2
Weak Passwords	2
Unsecured Websites	1
Two-Factor Authentication	5
Expiring	3

Search Personal Va

3 items sorted by Title

E*TRADE
wendy.c.appleseed@gm...

Netflix
wendy.c.appleseed@gm...

Yahoo
wendy.c.appleseed

(b) 1Password's

Figure 1: Password manager reports with hygiene statistics.

them honestly or accurately.

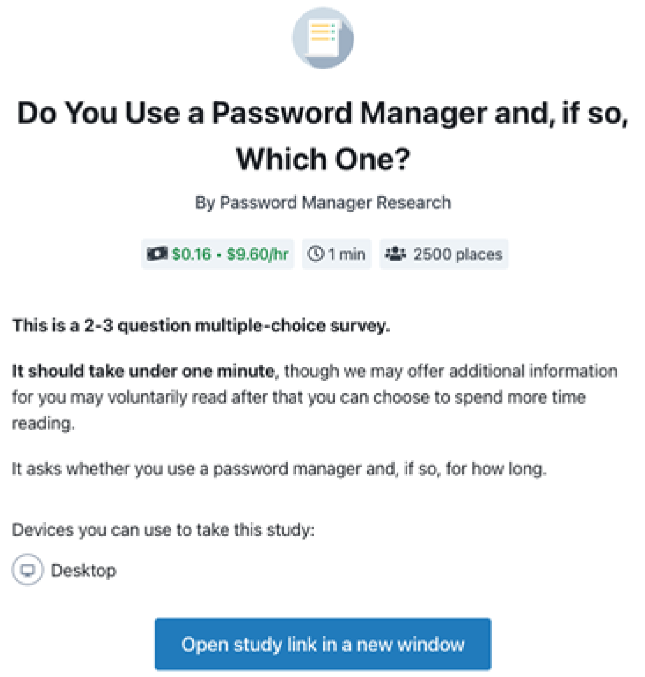
To reduce the risk of relying on participants as secondary sources of the hygiene statistics reported by their password managers, we asked participants to upload their primary sources: screenshots of these statistics displayed by their password manager, with sensitive information elided if necessary.

2 Experimental Design

We sought four password-hygiene statistics: the number of passwords their password manager classified as (1) reused, (2) weak, and (3) compromised, as well as (4) the total number of passwords stored. Most password managers display these statistics in security reports, often called "dashboards," as illustrated in Figure 1. Collecting anonymized screenshots allowed us to overcome many (but perhaps not all) instances in which participants might not actually be using a password manager, might not actually collect requested statistics, or might incorrectly report statistics into the survey form.

2.1 Screening

We identified users of password managers by running a short screening survey on the Prolific [14] participant recruitment platform. We titled our survey task "Do You Use a Password Manager and, if so, Which One?"



Do You Use a Password Manager and, if so, Which One?

By Password Manager Research

$0.16 · $9.60/hr | 1 min | 2500 places

This is a 2-3 question multiple-choice survey.

It should take under one minute, though we may offer additional information for you may voluntarily read after that you can choose to spend more time reading.

It asks whether you use a password manager and, if so, for how long.

Devices you can use to take this study:

Desktop

Figure 2: The advertisement for our study posted to Prolific.

To minimize the cost of screening, we designed the screening survey to take less than one minute and validated the time requirements using pilot studies. We promised that the task would be a 2–3 question multiple-choice survey, as illustrated in Figure 2. We paid $0.16 (see Figure 2), or $9.60/hour if participants used the full minute, per Prolific guidelines for "good" compensation. We did not identify the survey as a screener, as we did not want participants to try to identify which answers might lead to additional opportunities.

We used an option provided by prolific to display the advertisement only to participants using their browser's desktop mode, as participants using a mobile device would be less likely to have access to their password managers' desktop/web interface. Password managers' mobile apps for iOS and Android use an operating system mechanism to prevent screenshots, so participants without access to a desktop computer would be unable to capture screenshots.

We started our screening survey with two sentences to ensure that users of browser-based password managers knew they were, in fact, using a password manager.

> *A password manager is a program that saves your passwords and enters them for you. If you allow your web browser to save your passwords, you are using your browser's password manager.*

We then asked the participant which, if any password manager they used.

For this and all other survey questions, we share the exact wording of both the question and answer choices in a dedicated table, identified by the question number, in Section 3. This question about which password manager the screening participants used we label Question 1.

We screened out participants who were not using a password manager, asking them why they weren't using one (Question 2), before thanking them for completing the (screening) survey.

We also screened out those who were using a password manager that did not offer the hygiene statistics we needed—most notably Bitwarden's free edition, as their hygiene reports are a *premium* feature. And we screened out browser-based password managers other than Google's Chrome. Early pilots showed the great majority of respondents using password managers used the manager built into Chrome. To ensure a sufficient sample of users of third-party password managers, we down-sampled participants who used Chrome's password manager by inviting them to the full study with probability 1/15, based on the proportions we observed in those pilots. Since we had only asked one question of the Chrome participants we filtered out, and they had committed to answering 2–3, we took the opportunity to ask a follow-up question, about whether they let their password manager create passwords for them or created passwords themselves, which would be Question 16a for those completing the full study.

To those who reported creating passwords themselves, we asked why (free response) for the third and final question.

2.2 Consent

The one minute of participants' time that our $0.16 gratuity paid for was insufficient to convey everything they would need to know to consent to the full study. Paying a higher gratuity to all screening participants, including the majority we would be screening out before the consent, would consume budget that would be better spent on participants invited to the full study. So, we divided the consent into two stages.

For those screening participants we invited to the full study, we asked "Do you want to earn a USD $0.25 bonus spending one more minute learning about a USD $5.00 follow-up study?" Those who chose no were done, and to those who chose yes we explained what we would require in the full study.

To participate in the full USD $5.00 follow-up study:

- *You must take this survey on a computer (not a mobile phone or tablet) on which you have your password manager installed.*
- *You must be able to access your password manager through either a desktop or web interface, as this is necessary to capture a screenshot (iOS and Android apps prevent these screenshots).*
- *You must be willing to upload a screenshot of statistics generated by your password manager from your usage data. We will not ask you to upload any passwords or provide any information that would allow us to identify you. (The statistics we are looking for are three numbers, typically from 0 to 1000).*
- *The purpose of this study is [SIC[1]] determine whether people who [SIC[2]] password managers are benefiting from all their security features.*

Our $5 gratuity would pay for over 30 minutes at $9.60/hour, the rate Prolific's payment meter labels as "good," or 24 minutes at $12.50/hour, the top of Prolific's payment meter at which it labels pay as "great." During our initial pilot studies, participants spent a mean of 17 minutes and a median of 13 minutes and 43 seconds to complete, inclusive of the time for screening and consent for which participants received separate compensation. We had intended to present participants with the expected time to complete during the invitation and consent, but we either neglected to add this information or accidentally removed it during revisions of our survey.

We paid all those who declined the full study the $0.25 they were due, without noting whether they actually spent time learning about the follow-up study.

For those who were using a password manager and joined the full study, we would later ask *"How long have you been [SIC[3]] that password manager?"* (Question 24) so that we could filter out those who had not had at least two months to replace their weak, re-used, and compromised passwords.

2.3 Reporting of password statistics

When participants agreed to participate in the full study, our first request was for the screenshot(s) containing their password hygiene statistics. We provided product-specific instructions for finding these statistics. We explained how to take a screenshot, redact any private information, and upload it.

We then asked participants to self-report those same statistics, which we would then check against the screenshots.

2.4 Self-reported current hygiene

Having weak or reused passwords does not necessarily imply bad password hygiene. Rather, users might sensibly choose not to replace passwords that protect accounts that they have no reason to protect. To determine if participants had reported having weak or reused passwords for accounts they needed to protect, we asked if they felt that none of these passwords needed to be replaced, or if they felt they *should* replace "some" or "all" of them. (Question 8)

To those who responded that none of their weak or re-used passwords needed to be replaced (option 1 of Question 8), or

[1] we mistakenly elided the word "to"

[2] we mistakenly elided the word "use"

[3] We mistakenly elided the word "using"

that only some did (option 2), we asked why these passwords needn't be replaced (Question 9). To those who felt some (option 2) or all (option 3) of these passwords should be replaced, we asked why they had yet to replace them (Question 10).

We then asked those with compromised passwords whether any or all should be replaced (Question 11) with the same two follow ups (Questions 12 and 13).

Some participants might not have replaced weak or compromised passwords because they had not known they needed to, whereas others might have ignored prompts to do so. We thus asked whether participants had been encouraged by their password manager to replace their weak and re-used passwords (Question 14) as well as their compromised passwords (Question 15).

2.5 Effort expected to replace passwords

To understand the perceived effort of replacing passwords, we asked participants how much time they believed they would need to replace all their weak, re-used, and compromised passwords. We first asked for replacing all passwords, including any they did not feel they should change (the number they reported), and then excluding those they did not believe they should change—though we had not asked them to count the number they should (or should not) change.

We started by asking participants if their password manager offered "a feature to change passwords for common websites with a single click" (Question 20).

If the participant's password manager didn't have that feature, or they were unaware of it, we asked if they would use it if available (Question 21). If they did know about the feature, we asked if they had used it (Question 22) and if they expected to use it in the future (Question 23).

2.6 Master passwords and recovery

A user's master password must be simultaneously secure and memorable: its compromise can expose all of the passwords saved by the password manager, but if a user cannot remember it they may lose all these passwords. (For users of Chrome's password manager, their Google Account password is effectively their master password.) We asked participants how they created this all-important password (Question 3) what steps they had taken to ensure they didn't lose or forget it (Question 4), and what steps they had taken to ensure they could recover if they did lose or forget it (Question 5 for participants using Chrome's password manager and Question 6 for those using third-party password managers). For those who reported that they kept a copy of their master (Google Account) password on a device, we asked if they encrypted it (Question 7).

2.7 Demographics

We concluded the survey by asking the participant's age (in years) and gender identity (with options for "non-binary/third gender", "prefer to self-describe" with text entry, and "prefer not to say"). While we were not studying age or gender, this question proved useful as, following data collection, there were reports that Prolific's participant pool had been briefly skewed strongly young and female by a viral social media post [4].

Our last question was to report any problems with the survey which, despite its presence in pilot studies, did not result in the reporting of numerous language issues that sadly survived myriad proofreading by multiple native speakers of English. (See more in Section 4.1.)

2.8 Ethics and Safety

We designed our study to avoid collecting personally identifiable information or any other data that could put participants at risk. We instructed participants to remove potentially sensitive data from their screenshots and we verified during pilot studies that these instructions were effective. We also used pilot studies to ensure that we accurately represented the time required for the initial screening survey, consent step, and complete survey and paid fair wages. When evidence strongly indicated that participants had submitted screenshots that weren't their own or otherwise misled us, we followed up with them to give them a chance to explain what we had found before making the decision to withhold payment. When facing reasonable doubt, we paid participants rather than risk being in the wrong. We applied for and received approval from UC Berkeley's Office for Protection of Human Subjects (OPHS Protocol ID: 2020-11-13788). We observed no incidents of concern during the study.

3 Results

Prior to conducting our study we conducted a series of pilots. Our largest pilot, in March 2021, reached a cohort of 2491 screening participants, yielding 100 full participants. We previewed preliminary results from this pilot at the 2021 RSA Conference [12]. As we iterated on our methodology, we tested the changes using smaller cohorts to test further changes in August and early September.

We conducted our the full study from September 17th, 2021, to October 2nd, 2021. We used two roughly equally sized cohorts, monitoring response rates of the first cohort to ensure we stayed within budgetary constraints as we opened our study to the second cohort.

3.1 Initial screening

Of 5081 survey responses, we removed 9 because they had non-unique participant IDs and, in each instance, the participant made more progress in a different (typically subsequent) session. Each of the remaining 5072 sessions represents a unique Prolific participant identifier, which we treat as a unique participant. (We cannot guarantee that no human subject participated more than once using multiple Prolific accounts with different participant identifiers, though the screenshot requirement of our study would make it harder to do so than for most other studies.) Of those participants who completed the study, the mean time required was 17 minutes and 58 seconds and the median was 15 minutes and 14 seconds.

In Question 1 we detail the 5072 responses to our first screening question: which password manager participants used.

Roughly a quarter of screening participants, (1320, or 26%) responded *I'm not using a password manager*. When asked why (Question 2), a majority (58%) reported concerns with protecting the passwords in their password manager from others and nearly as many that their password manager might be compromised (46%), and over a quarter (28%) were distrusted developers of password managers. All security concerns were more commonly reported than a failure to appreciate the benefits of the product (25%) or ignorance of the products (10%).

The most common password manager in use was the one built into Chrome, which 1839 (36%) reported using. Second was Apple's Keychain, which 658 (13%) reported using (which we did not study as it did not report the required password hygiene statistics). Another 652 screening participants (13%) used one of the third-party password managers that reported the required hygiene statistics and qualified them for an invitation to the full study. Of those, 294 reported using 1Password, though its popularity might have been due to it being the first option listed (more on this later). Among those not included in the count of 652 participants using password managers that reported statistics were the 99 screening participants (2%) using Bitwarden's free edition—only the premium edition reported the hygiene statistics and so the free edition cannot help users identify and replace their weak and re-used passwords.

The random number generator we used to screen out 14/15 Chrome participants (inviting any given participant with probability $0.0\overline{6}$) granted invitations to 120 of the 1839 eligible Chrome participants from the screening survey (6.53%).

For those who answered that they weren't using a password manager, we had asked why

3.2 Two-stage consent

We detail participants' progress through the two-stage consent process, and their eventual responses, in Table 1.

None of the 120 Chrome invitees terminated the survey without answering the consent, whereas 126 of the 652 third-party invitees (19%) did. We had not hypothesized such a difference or planned to test for it, but since it represents a reason to question the validity of our data, we note that the probability of this happening by chance is < 0.00001 (Fisher's Exact Test for matrix 0, 120, 126, 652).

The most common reason given for declining the study was discomfort with uploading the screenshot, with roughly 15% of both Chrome and third-party invitees opting out for this reason. This is higher than we would have hoped, and we would like to know if participants were worried they might be associated with their bad hygiene (perhaps a legitimate risk) or if we had somehow failed to convince them that the screenshots were otherwise harmless. We hope researchers building on our methodology in the future can reduce the opt-out rate as it could bias the sample in favor of less risk averse participants.

At the completion of the two-stage consent, 71 of the 120 Chrome invitees (59%) consented and 190 of the 652 third-party invitees (29%) consented. Of the 30 percentage point difference, 19 were due to the third-party participants who did not answer the consent.

One explanation is that this reflects a real difference between users of Chrome's password manager and users of third-party password managers, perhaps because the latter are more privacy conscious. However, for those who answered the two-stage consent, nearly the same percentage (15% expressed privacy concerns.

Another explanation is that the difference might be caused by a bot or other collective of disingenuous Prolific users who were attempting to game the survey to be compensated for the screening portion and the initial ($0.25) bonus. We had not randomized the long list of options as they were hard for participants to navigate even when alphabetically ordered, and so we wondered if some disingenuous participants just chose the first answer (1Password). Few appear to have done so. Of the 294 participants who chose 1Password for the first question, 79 (27%) consented vs. 111 of 358 third-party invitees who had not chosen 1Password (31%).

Yet another explanation would be disingenuous respondents picking answers at random. This seems more plausible, as if there were disingenuous respondents picking answers to the first question at random, the Chrome invitees would be mostly unaffected. Since users of Chrome's password manager were so numerous, and we only invited one out of every 15 respondents who chose Chrome, we would automatically remove 14 of every 15 disingenuous Chrome participants.

3.3 Filtering those who completed the study

We filtered out 5 Chrome and 35 third-party participants because the screenshots they uploaded did not appear to be genuine. (Of these disingenuous participants, 23 of the 35 third-party participants had responded to the initial screen

Password Manager	*Screening Participants*		*Validated Participants*	
1Password	294	(6%)	14	(10%)
Bitwarden (Premium Edition)	13	(<1%)	4	(3%)
Bitwarden (Free Edition)	99	(2%)	*NA*	
Dashlane	42	(1%)	11	(8%)
KeePassXC	45	(1%)	11	(8%)
Keeper Password Manager	26	(1%)	0	(0%)
LastPass	136	(3%)	37	(26%)
Norton Password Manager	53	(1%)	2	(1%)
Password Boss	20	(<1%)	0	(0%)
RoboForm	10	(<1%)	2	(1%)
StickyPassword	9	(<1%)	0	(0%)
Zoho Vault	4	(<1%)	0	(0%)
The Password Manager Built into Google's Chrome Browser (Google Password Manager)	1839	(36%)	61	(43%)
The Password Manager Built into Apple's Safari Browser, MacOS, and iOS (Apple's Keychain)	658	(13%)	*NA*	
The Password Manager Built into Microsoft's Edge Browser	120	(2%)	*NA*	
The Password Manager Built into Mozilla's Firefox Browser	198	(2%)	*NA*	
Another Browser's Built-In Password Manager (please type the name below)	96	(2%)	*NA*	
Other Password Manager (please type the name below)	90	(2%)	*NA*	
I'm not using a password manager	1320	(26%)	*NA*	
Total	**5072**	(100%)	142	(100%)

Question 1: *"Which password manager are you using for your personal accounts? (if you use more than one, please report the one that manages the most accounts.)"* Answers tallied for the set of screening participants and for those participants who made it through the full study, had validated screenshots, and at least five passwords stored.

question with "1Password", the topmost response.)

We removed 5 Chrome participants and other 72 third-party participants because they had saved fewer than five passwords in their password managers and so did not appear to be using password managers for a significant portion of their passwords.

Two third-party participants did not complete the survey (all Chrome participants did), leaving us with 61 Chrome participants and 81 third-party participants.

3.4 Demographics

Since we did not pay participants in our screening-survey for the time and effort to answer demographic questions, we asked age and gender questions to those who were invited to the full study and chose to complete it. The responses of those who participated in the full study skewed young and male.

Specifically, 62 of 81 (77%) third-party participants and 33 of 61 (54%) Chrome participants identified as male. Only 18 of 81 third-party participants (22%) and 26 of 61 Chrome participants (43%) identified as female, with 2 Chrome participants (3%) and 1 (1%) third-party participant identifying as non-binary/third. No participant self-described or opted not to answer. The responses to our gender question do not indicate that our study was impacted by social media skewing Prolific's participant pool to be younger and more female, as occurred during other researchers' 2021 studies [4].

Being unaffected by that particular event does not mean that Prolific's participant pool is any more representative of the general population than it would be otherwise. Indeed, the 25th, 50th (median), and 75th percentile ages were 21, 24, and 32, far younger than the general population (which, in turn, may be different from the demographics of those who use password managers).

3.5 Master passwords and recovery

Like our demographic questions, we asked about master passwords and recovery toward the end of the survey. We examine these questions first because participants who were uncertain their master passwords were secure, or concerned they could lose access to their stored passwords, might be less willing to fully depend on their password manager to remember all their passwords and reluctant to adopt good hygiene.

When we asked third-party participants how they created their master password, the majority (50/81, 62%) answered "I created a password using only my mind." A quarter (19/81) re-used a password, only 5 of 81 (6%) answered "I used a random password suggested by my password manager," and another 5 (6%) answered using some other random process.

The passwords in Google's Chrome password manager are protected by users' Google Accounts, which are protected by passwords (thus somewhat equivalent to a master password)

[4] we mistakenly elided the words "you lose".

I am concerned that other people may access the computer, tablet, phone, or other device on which my passwords are saved.	764	(58%)
I worry that if malicious software compromises my devices my passwords may be compromised.	601	(46%)
I don't trust the companies that make the browser I use, or third-party password managers, with my passwords.	363	(28%)
I don't see any benefit in having my passwords saved and automatically entered.	328	(25%)
I didn't know I could save my passwords in my browser or with a password manager.	133	(10%)
Other (Please explain)	61	(5%)
Total participants responding	1315	(100%)

Question 2: *"Why are you not saving your passwords in your browser or other password manager? (Check all that apply)"*

	Chrome		**Third-party**	
I re-used a password I had already memorized.	30	(49%)	19	(24%)
I used a random password suggested by my password manager.	0	(0%)	5	(6%)
I created a password using only my mind.	28	(46%)	50	(62%)
I created a password using a physical randomness, software, or other non-mental process. (please explain)	0	(0%)	5	(6%)
Other (please explain)	3	(5%)	2	(2%)
Total	61	(100%)	81	(100%)

Question 3: *"How did you create and memorize the master password for your password manager?"*

I printed an "emergency kit" generated by my password manager.	1	(2%)	12	(15%)
I wrote down the master password onto paper or another physical medium.	27	(44%)	28	(35%)
I emailed my master password to myself.	7	(12%)	5	(6%)
I printed a copy of my master password.	2	(3%)	2	(3%)
I stored my master password in a file on my phone, computer, or other device.	17	(28%)	18	(22%)
I stored my master password in another way. (please explain)	6	(10%)	14	(17%)
I used a feature in my password manager to empower one or more trusted friends or family to help me recover my data.	2	(3%)	2	(3%)
I took one or more actions not listed above. (please explain)	7	(12%)	15	(19%)
Total participants responding	**61**	(100%)	**81**	(100%)

Question 4: *"Which of the following actions have you taken to ensure you don't forget your master password or lose access to your passwords? (Check all that apply)"*

I have printed all my passwords.	11	(19%)
I configured my password manager to give trusted contacts permission to recover my password data.	14	(24%)
I have given a copy of all the information I would need to recover Google Account to people I trust.	2	(3%)
I have a backup account / device authorized to access my Google Account and obtain my password data without the need to login.	34	(59%)
I have taken one or more other/different approaches. (Please describe in detail.)	7	(12%)
Total participants responding	**58**	(100%)

Question 5: *"What steps have you taken to ensure you can recover your passwords should you lose access to your google account. (Check all that apply)."* This question was optional.

I have printed all my passwords.	4	(5%)
I have printed an "emergency kit" provided by my password manager for use in recovering my password data.	15	(19%)
I configured my password manager to give trusted contacts permission to recover my password data.	8	(10%)
I have given a copy of all the information I would need to recover my password data with people I trust.	3	(4%)
I have installed my password manager on a backup device that can access my password data without the need for the master password.	17	(22%)
I have taken one or more other/different approaches. (Please describe in detail.)	25	(32%)
Total participants responding	**77**	(100%)

Question 6: *"What steps have you taken to ensure you can recover your passwords should* [SIC[4]] *your master password and/or the device(s) on which your password manager is installed. (Check all that apply)"* This question was optional.

	Chrome		Third-party	
Declined prior to learning about full study				
Terminated survey without answering consent 1	0	(0%)	126	(19%)
Declined to learn about full study	2	(2%)	26	(4%)
Terminated survey after consenting to learn but without answering consent to participate	1	(<1%)	9	(1%)
Declined to consent to full study				
I decline to participate and decline to provide a reason.	6	(5%)	49	(8%)
I am qualified and would like to participate, but I am not at a desktop computer right now. Please contact me later.	8	(7%)	34	(5%)
I am qualified and would like to participate, but I don't have time right now. Please contact me later.	3	(3%)	29	(4%)
I do not want to participate because the study doesn't pay as much as I'd like. (Enter the price you would participate for.)	0	(0%)	9	(1%)
I do not qualify because I do not have a desktop computer on which to perform the study tasks.	7	(6%)	26	(4%)
I do not qualify because I cannot use my password manager's desktop or web interface.	4	(3%)	39	(6%)
I am not comfortable uploading the screenshot of aggregate statistics or answering questions about my password manager.	17	(14%)	101	(15%)
I am unable to participate for other reasons. (Please tell us why.)	1	(<1%)	9	(1%)
I do not want to participate for other reasons. (Please tell us why.)	0	(0%)	5	(<1%)
Consented to the full study				
Yes, I am qualified to participate in the full $5.00 study and want to start immediately.	71	(59%)	190	(29%)
Total	120	(100%)	652	(100%)

Table 1: The two-step consent to (1) learn about the full study and (2) to participate in it.

	Chrome		Third-party	
Yes	4	(24%)	3	(17%)
No	12	(71%)	14	(78%)
I stored it in some files that were encrypted and some that were not.	1	(6%)	1	(6%)
Total	**17**	(100%)	**18**	(100%)

Question 7: *"You reported that you stored your password manager's master password in a file on your phone, computer, or other device. Did you encrypt the file you stored it in?"*

and potentially other authentication factors. The registration process for a Google Account does not offer a built-in random-password generator, so it's not surprising that no Chrome participant reported using a random password suggested by their password manager as a master (Google Account) password. Half of Chrome participants (30/61, 49%) re-used a password they had already memorized as their Google Account password. Almost everyone else (28/61, 46%) reported that they used a mentally-generated password.

The most commonly reported precaution taken to prevent losing access to master passwords (Question 4) was to write master passwords down.

The most common "backup" when a master password was lost (Questions 5 & 6 for Chrome & third-party managers respectively) was to have the password manager installed on multiple devices, with 34 of 58 (59%) of Chrome participants and 17 of 77 third-party participants (22%) reporting this. Only 15 of 77 of third-party participants (19%) printed out an emergency kit, even though most third-party password managers recommend doing so.

3.6 Hygiene for existing passwords

We graph the password statistics we collected in Figure 3: with Chrome participants in Figure 3a and third-party participants in Figure 3b. The green lines are the cumulative distribution functions (CDFs) of participants' total stored passwords, and the other lines are the CDFs of the number of weak, re-used, and compromised passwords. If participants had perfect hygiene, they would use their password manager for all their passwords (pushing the green line upward) and have no weak, re-used, or compromised passwords (maintaining the other lines flat at the 0 point on the x axis).

All but two participants (both third-party) had at least one re-used password (the yellow lines in Figure 3). All Chrome participants but one (60/61) had at least one password deemed weak by Google's password report, and nearly half of third-party participants had a password that their manager deemed weak (the red lines in Figure 3).

The mere presence of weak and reused passwords does not prove that participants hygiene put them at risk. Some participants may have removed all the weak and reused passwords for accounts they would not want to be compromised, and left the passwords for valueless accounts untouched. When we asked if they had weak or re-used passwords they should replace (Question 8), 52 of 60 (87%) Chrome participants with weak or re-used passwords confessed that they did[5] as

[5] This question was intended to be mandatory; however, one Chrome participant was able to leave this question unanswered.

	Chrome		Third-party	
Not asked/answered: no weak/re-used passwords reported	1	(2%)	2	(2%)
I do not need to replace any of the passwords reported as weak or reused.	8	(13%)	10	(12%)
I should replace some of the passwords reported as weak or reused.	42	(69%)	48	(59%)
I should replace all of the passwords reported as weak or reused.	10	(16%)	21	(26%)
Total participants	**61**	(100%)	**81**	(100%)

Question 8: *"Which statement best categorizes how you feel about replacing the passwords reported as weak or reused?"*

	Chrome		Third-party	
Some weak or reused passwords protect accounts that aren't worth protecting.	26	(40%)	43	(53%)
I prefer to have passwords I can remember without my password manager.	34	(56%)	23	(28%)
Other (please explain)	2	(3%)	6	(7%)

Question 9: *"Why do you feel it's okay to have some weak or reused passwords? (Check all that apply)"* Percentages are of total participants from the final row of Question 8.

	Chrome		Third-party	
I was not previously aware of these passwords being weak or reused.	27	(44%)	23	(28%)
I do not know how to replace these passwords.	0	(0%)	1	(1%)
I am worried something could go wrong when I replace these passwords.	9	(15%)	6	(7%)
The amount of work required to replace these passwords is overwhelming.	21	(34%)	30	(37%)
I have not gotten around to replacing these reused passwords.	16	(26%)	33	(41%)
Other (please explain)	6	(10%)	21	(26%)

Question 10: *"Why do you still have weak or reused passwords that you feel you should replace? (Check all that apply)"* Percentages are of total participants from the final row of Question 8, so 28% of participants using third-party password managers had weak or reused passwords that they reported not being previously aware of and that they should replace (top right).

did 69 of 79 (87%) of third-party participants with weak or re-used passwords.

For those participants who had some passwords they did not feel a need to replace, we asked why they didn't need replacing (Question 9). Factoring in free responses, most participants had at least some deleted or otherwise valueless accounts.

Still, roughly two-thirds of Chrome participants (34/52, 65%) and one third of third-party participants (23/69, 33%) reported wanting to keep passwords they "could remember." The allure of having passwords that can be entered from memory appears to be a significant impediment to improving password hygiene, as we will see again in Section 3.7.

For participants who had passwords they knew they should replace but hadn't yet, we asked why they still had them (Question 10). Again, they could choose more than one answer. The responses were fairly evenly divided between participants being unaware of the problem, being overwhelmed by the task of fixing it, or just not having gotten around to it yet.

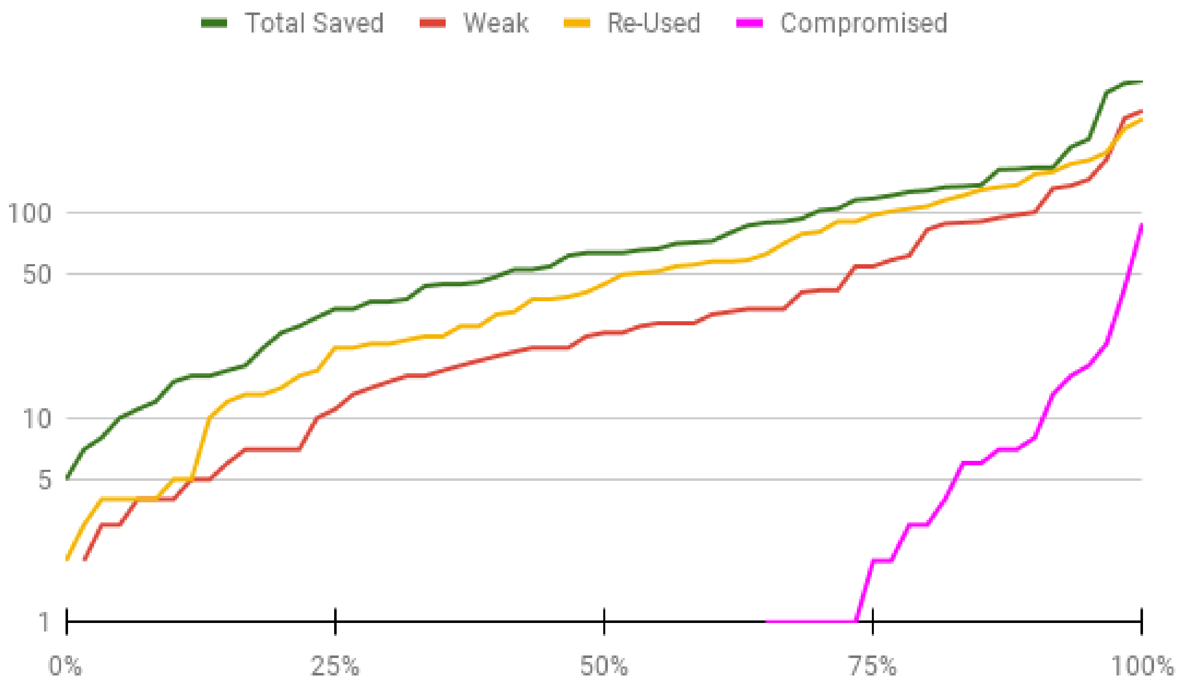
As shown by the purple lines in Figure 3, more than a quarter of both chrome and third-party participants' screenshots reported passwords that their password managers knew to be compromised! Among Chrome participants, 11 reported that *all* of these passwords should be replaced and another 7 reported that *some* should, for a total of 18 participants (30% of all 61 Chrome participants). Among third-party participants, 19 reported that *all* should be replaced and another 5 that *some* should be replaced, for a total of 24 (30% of all 81 third-party participants).

To those participants who confessed that they should replace some or all of their compromised passwords, we again asked why they had not replaced them (Question 13). Since they might have more than one reason, we allowed them to choose more than one answer. 11 out of 18 Chrome participants with compromised passwords (61%) and 16 out of 24 third-party participants with compromised passwords (67%) checked "I was not previously aware of these passwords being compromised" among their (potentially multiple) explanations.

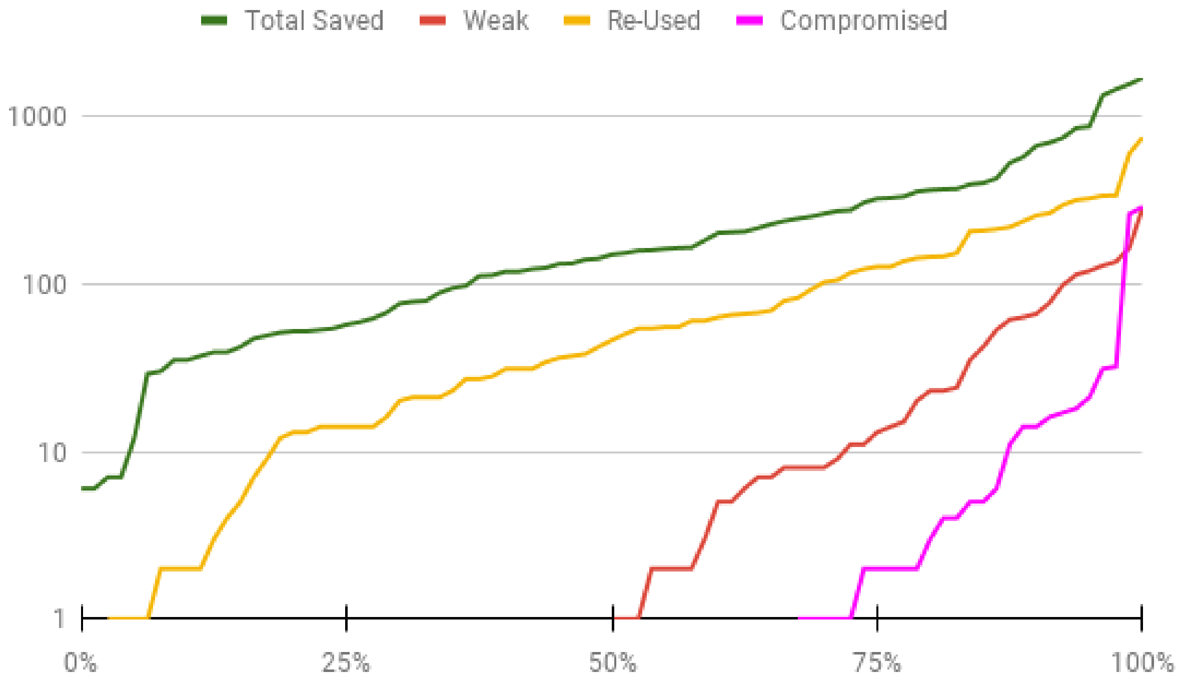
We asked the few participants who asserted that their compromised passwords need not be replaced, why not (Question 12). Factoring in free responses, most were protecting valueless accounts.

3.7 Hygiene for new passwords

When asked how they create passwords for new accounts (Question 16a), only 13 of 61 of Chrome participants (21%) responded that they would allow Chrome to create a random password for them. Most third-party participants reported that they allow their password manager to create random passwords for them: 50 of 81 (62%). Still, 29 of 81 (36%) of those who had made the effort to obtain and use a third-party password manager were creating passwords themselves. So,



(a) Chrome's password manager



(b) Third-party password managers

Figure 3: Cumulative distribution functions (CDFs) of participants' total number of saved passwords and those saved passwords that their password manager classified as weak, re-used, and compromised.At a given point, the x axis contains a fraction of participants whose count of [total/weak/re-used/compromised] passwords did not exceed the value in the y axis. For example, where the green line intersects the 100 mark on the Y axis, the X axis represents the cumulative percentage of participants with who had 100 total saved passwords or fewer.

	Chrome		Third-party	
Not asked/answered: no compromised passwords reported	37	(61%)	53	(65%)
I do not need to replace any of the passwords reported as compromised.	6	(10%)	4	(5%)
I should replace some of the passwords reported as compromised.	7	(11%)	5	(6%)
I should replace all of the passwords reported as compromised.	11	(18%)	19	(23%)
Total participants	**61**	(100%)	**81**	(100%)

Question 11: *"Which statement best categorizes how you feel about replacing the passwords reported as compromised?"*

	Chrome		Third-party	
Some compromised passwords protect accounts that aren't worth protecting.	6	(10%)	4	(5%)
I prefer to have passwords I can remember without my password manager.	7	(11%)	2	(2%)
Other (please explain[6])	2	(3%)	3	(4%)

Question 12: *"Why do you feel it's okay to have some compromised passwords? (Check all that apply)"* Percentages are of total participants from the final row of Question 11.

	Chrome		Third-party	
I was not previously aware of these passwords being compromised.	11	(18%)	16	(20%)
I do not know how to replace these passwords.	1	(2%)	1	(1%)
I am worried something could go wrong when I replace these passwords.	1	(2%)	2	(2%)
The amount of work required to replace these passwords is overwhelming.	4	(7%)	3	(4%)
I have not gotten around to replacing these reused passwords.	1	(2%)	6	(7%)
Other (please explain)	2	(3%)	1	(1%)

Question 13: *"Why do you still have compromised passwords that you feel you should replace? (Check all that apply)"* Percentages are of total participants from the final row of Question 11, so 20% of participants using third-party password managers had compromised passwords that they reported not being previously aware of and that they should replace (top right).

	Chrome		Third-party	
Never	25	(41%)	20	(25%)
When I first set up the password manager	6	(10%)	10	(12%)
More than a year ago	5	(8%)	9	(11%)
Within the past week	3	(5%)	12	(15%)
More than a month ago	16	(26%)	20	(25%)
More than a week ago	6	(10%)	10	(12%)
Total	**61**	(100%)	**81**	(100%)

Question 14: *"Prior to your participation in this study, when was the last time you can recall your password manager encouraging you to replace one or more of your weak or re-used passwords?"*

	Chrome		Third-party	
Never	5	(8%)	10	(12%)
When I first set up the password manager	2	(3%)	3	(4%)
More than a year ago	0	(0%)	7	(9%)
Within the past week	4	(7%)	0	(0%)
More than a month ago	10	(16%)	6	(7%)
More than a week ago	3	(5%)	2	(2%)
Total	**24**	(100%)	**28**	(100%)

Question 15: *"Prior to your participation in this study, when was the last time you can recall your password manager encouraging you to replace one or more of your compromised passwords?"*

even for new accounts, it is not safe to assume that those using password managers are practicing good hygiene.

We asked the same question of participants who used password managers but were screened out (Question 16b). The data are less trustworthy since disingenuous participants could not be filtered out.

More usefully, when we asked those participants "Why are you more likely to create a password for yourself than let your password manager create one for you?", two thirds (1585 of 2467, or 64%) wrote a free-form response that included the word "remember" suggesting that participants wanted passwords they could remember in the absence of their password manager. Others used words like "memory," and so again the allure of having passwords one can enter from memory appears to be an important barrier to better hygiene.

3.8 Feature awareness

We asked our participants if they had previously known that their password managers tracked weak and re-used passwords (Question 17). Roughly half of Chrome participants and roughly a third of third-party participants did not.

We also asked those participants who knew about the reports that warned them of weak and re-used passwords how

	Chrome	Third-party
Create a password myself and let my password manager save it	48 (79%)	29 (36%)
Allow my password manager to create a random password for me	13 (21%)	50 (62%)
Other (Please Explain)	0 (0%)	2 (2%)
Total	**61** (100%)	**81** (100%)

(a) answered by those who completed the study

	Chrome	Third-party
Create a password myself and let my password manager save it	2400 (86%)	33 (33%)
Allow my password manager to create a random password for me	346 (12%)	66 (67%)
Other (Please explain)	43 (2%)	0 (0%)
Total	**2789** (100%)	**99** (100%)

(b) answered by screening survey participants who reported using *any* browser-based or third-party password manager but who we screened out of the full study

Question 16: *"When you are creating an account on a website or changing your password, are you more likely to?"*

	Chrome	Third-party
Yes	28 (46%)	56 (69%)
No	33 (54%)	25 (31%)
Total	**61** (100%)	**81** (100%)

Question 17: *"Before this survey, did you know that your password manager tracks weak and re-used passwords?"*

	Chrome	Third-party
Very Frequently	3 (11%)	1 (2%)
Frequently	6 (21%)	12 (21%)
Rarely	15 (54%)	24 (43%)
Very Rarely	4 (14%)	15 (27%)
Never	0 (0%)	4 (7%)
Total	**28** (100%)	**56** (100%)

Question 18: *"How often do you encounter the screens that track your weak and re-used passwords?"*

	Chrome	Third-party
Very Frequently	2 (7%)	0 (0%)
Frequently	4 (14%)	6 (10%)
Rarely	13 (46%)	27 (48%)
Very Rarely	8 (29%)	16 (29%)
Never	1 (4%)	7 (13%)
Total	**28** (100%)	**56** (100%)

Question 19: *"How often do you take the time replace weak, re-used, or compromised passwords?"*

	Chrome	Third-party
Yes	33 (54%)	38 (47%)
No	28 (46%)	43 (53%)
Total	**61** (100%)	**81** (100%)

Question 20: *"Does your password manager offer a feature to change passwords for common websites with a single click?"*

frequently they had encountered them (Question 18) and how frequently they took the time to use these features to improve their passwords (Question 19). The majority reported that they rarely, very rarely, or never encountered the reports: 19 of 28 (68%) of Chrome participants and 43 of 56 (77%) of third-party participants. Only a select few frequently take time to improve their existing passwords.

Perhaps few participants replace their passwords because of the perceived effort. Some password managers, including Chrome and LastPass, can change passwords for popular websites with one click. We asked participants if their password manager offered such a feature (Question 20) and 46% of Chrome participants didn't know it did.

We asked participants who had not known about the reports if they planned to use them in the future to improve their passwords (Question 23). Many participants reported aspiring to do so, though the reliability of such self-reported data is suspect given that the purpose of the study was known, and participants could surely infer the most desirable answer. (We had asked because we had more participants unexpectedly answered "probably not" or "definitely not" we would have learned something unexpected and would want to investigate why.)

We asked participants who reported being aware of the password-replacement feature whether they used it (Question 22) and we asked those who did not know it existed if they would use such a feature (Question 21). Of those who didn't have the feature or know it existed, most said they would use it, but that is to be expected given the topic of our survey and the implicit desired response (there are demand effects).

[7] We repeated mistakenly the word "to" twice.

[8] We mistakenly elided the word "using".

	Chrome		Third-party	
Yes	18	(64%)	28	(65%)
No (Please explain)	2	(7%)	6	(14%)
I don’t know	8	(29%)	9	(21%)
Total	**28**	(100%)	**43**	(100%)

Question 21: *“If your password manager did offer a feature to change passwords for common websites with a single click, would you use it?”*

	Chrome		Third-party	
Yes	8	(24%)	19	(50%)
No (Please explain)	16	(48%)	12	(32%)
I don’t know	9	(27%)	7	(18%)
Total	**33**	(100%)	**38**	(100%)

Question 22: *“Have you used your password manager’s feature to to* [SIC[7]] *change passwords for one or more common websites with a single click?”*

3.9 Perceived time to replace passwords

We asked participants how much time they expected would be required to replace all their weak, re-used, and compromised passwords. In Figure 4, we plot their response to that question against the sum of the total number of weak, re-used, and compromised they reported having. Answers clustered around 60 minutes (one hour). Dividing the median time (50 minutes) by the median number of passwords (62) yields an estimate of about a minute per password.

4 Limitations

Our study had numerous flaws and limitations of note and readers should be take note of them to apply skepticism of our results.

4.1 Flaws in experimental design

Given that our survey was proofread by myriad native speakers of English and presented to thousands of pilot participants who were encouraged to share errors with us, we were surprised and disappointed by the number of language errors that survived and were only discovered after we conducted the final study. While participants seemed to have understood our questions despite these issues, it is impossible to prove that that was not the case.

In the screening task, we failed to inform participants how long the full study would take. Again, we saw no specific evidence that this may have biased our results towards the security-savvy or otherwise, but we cannot prove that it did not.

	Chrome		Third-party	
Definitely	18	(55%)	11	(44%)
Probably	6	(18%)	10	(42%)
Maybe	5	(15%)	3	(12%)
Probably not	4	(12%)	1	(4%)
Definitely not	0	(0%)	0	(0%)
Total	**33**	(100%)	**26**	(100%)

Question 23: *“Do you expect to use your password manager’s features to replace weak, re-used, or compromised passwords in the future?”*

	Chrome		Third-party	
Less than 2 months	0	(0%)	0	(0%)
Between 2 months to 1 year	5	(8%)	16	(20%)
Between 1 to 2 years	13	(21%)	25	(31%)
Between 2 to 3 years	6	(10%)	18	(22%)
Between 3 to 4 years	9	(15%)	2	(2%)
More than 4 years	28	(46%)	20	(25%)
Total	**61**	(100%)	**81**	(100%)

Question 24: *“How long have you been [SIC[8]] that password manager?”*

4.2 Selection (sampling) biases

As with any study, selection biases could have been present.

Some participants (15% who opted out at during the second step of the consent process) were unwilling to upload password-manager statistics even knowing these data did not contain any personally-identifiable information or pose a significant risk, or may not have understood that the data posed little risk. Reducing this opt-out rate should be a priority for future research.

Some Chrome users might not have understood that they were using a password manager. Some might have been more or less reluctant to participate than third-party participants. The third-party groups may have been more vulnerable to disingenuous participants since the down-sampling reduced the impact of those participants on Chrome. Thus, despite our best efforts to collect comparable data for Chrome and third-party participants, differences between these participant groups might not represent true differences in the products themselves or in those who choose them.

Our methodology required us to exclude participants using password managers exclusively via their mobile applications, as mobile operating systems block screenshots if applications forbid them, which most password managers do.

Our sampling might have excluded those who were not immediately available to take the survey on a desktop device. We sent 53 surveys individually to respondents who asked to be contacted later. Only 3 of the 53 responded and finished.

By excluding participants whose password managers don’t report hygiene statistics, including Bitwarden’s free edition,

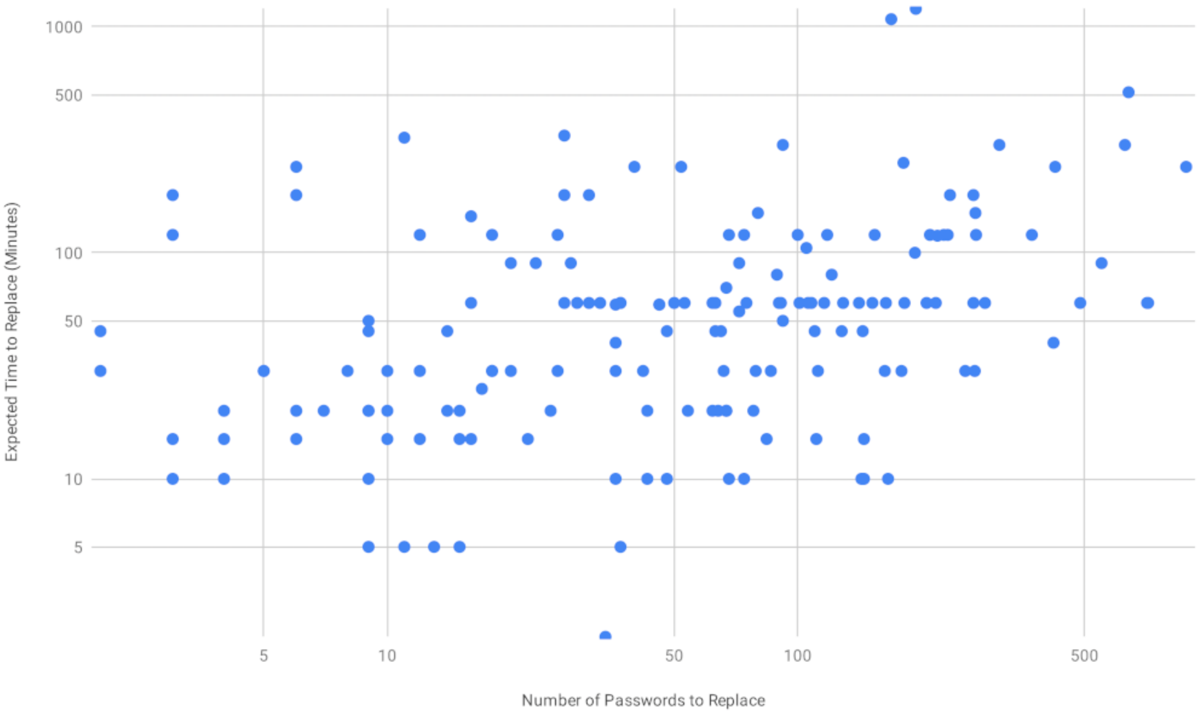


Figure 4: A scatter plot of participants' estimates of the time to replace all their weak, re-used, and compromised passwords vs. the number of those passwords they reported having.

we may have also biased responses to questions that did not require those hygiene statistics.

4.3 Data consistency

Different password managers may use different algorithms to classify passwords as weak, and even re-use may be subjective as different password managers may make different choices about whether a password associated with somewhat-related two domain names is being used for two accounts or one account.

4.4 Data collection and integrity

There were 5 participants (3 Bitwarden premium, 2 1Password) whose screenshots contained their weak, re-used and compromised passwords, but not the total number of passwords stored. If they correctly reported the numbers we could verify correctly we assumed that they reported the total number of passwords correctly even though we couldn't verify it.

We designed our methodology on the assumption that the most likely reason participants would fake a screenshot would be to receive a gratuity despite not actually using a password manager. Participants could have constructed forgeries by turning on their browser's debugger to modify HTML or by modifying images, however they would have little incentive to. The statistics themselves were not personal, and we provided the same gratuity regardless of whether participants were storing two passwords or 200. If they were not actually using a password manager, the work to identify the correct report and modify it was likely harder than installing the password manager. Still, it would be impossible for us to prove that we failed to detect some well-constructed forgeries of screenshots from disingenuous participants. Rather, we expect that if many participants were motivated to create forgeries, these forgeries would have a range of quality and we would have seen more poorly constructed forgeries.

Aside from the data collected from screenshots, many of our questions relied on participants to honestly and accurately report their behaviors, beliefs, and activities.

4.5 Improvements for future studies

Some peer reviewers who reviewed earlier drafts of this paper suggested we re-run our study. Our study budget has been exhausted and the masters students who conducted the study have since graduated. However, since millions use password managers, and those making often imply they improve hygiene without publishing statistics on real-world use, we believe it is important to share the results we have with the public.

To those considering reproducing our study, we would suggest improvements beyond correcting the unfortunate number of wording errors in our survey. Specifically, we would encourage researchers to ask participants to gauge how much of their decision to adopt a password manager was motivated by convenience and how much was security. If we had asked this question, we might be able to determine how many of participants with bad password hygiene had adopted password managers with the intent of improving their security. As mentioned earlier, we also would like to see researchers experiment to find ways to reduce participants' concerns with uploading screenshots to reduce the opt-out rate and resulting bias.

5 Related work

There is a dearth of research on password managers and, rather than pad our citation count with references to more general works on password, authentication, or usable security, we believe our best served by a more detailed exploration of the work that is available.

Studies of password managers date back as early as 2006 with Chiasson *et al.'s* examination of PwdHash and Password Multiplier [5] and Gaw and Felten's study of users' password-management habits [8]. By 2010, studies such as Karole *et al.* [9] examined password managers that look more like the ones we use today, including Apple's KeyChain and LastPass.

More recent work examines adoption of password managers, such as that of Alkaldi, Renaud, and Mackenzie, who focused on "autonomy and relatedness" [2].

To study password managers' impact on password hygiene, Lyastani *et al.* [11] recruited research participants via Mechanical Turk to install a browser plugin which recorded password-entry events. They were able to observe 128 password-entry events from password-manager users (all using LastPass). In those 128 events they observed only 100 unique passwords, indicating re-use was present. Limitations of their study include that they observed fewer than 3 passwords-entry events per participant during the study period and that the requirement that participants installing a browser extension may have biased the study in favor of less risk-averse participants with worse password hygiene.

We observed in Section 3.5 that few of our study participants used a random master password—the most secure option. Prior work shows that this is not because they are incapable of doing so. Bonneau and Schechter demonstrated that a surprisingly large fraction of the population can learn secrets with 56 bits of entropy given tens of short training sessions [3]. Doolani *et al.* have since shown that training may even be possible with one long session [7]. However, at the time of our study, no password manager provided any proven mechanism to help users learn a random master password.

Pearman *et al.* [13] previously compared those using browser-based and third-party password managers. They observed "higher levels of password reuse among users of [browser-based] password managers," consistent with our observations in Section 3.6. They also found that users of third-party password managers were more likely to use random passwords for new accounts, consistent with our observations in Section 3.7 (note the limitations of our results disclosed in Section 4.2).

Ray *et al.* [15] replicated the methodology of Pearman *et al.* for older adults, noting barriers such as the perception that they are "unlikely to create more passwords at their age." Because of the demographic limitations of our Prolific-recruited participant pool, we are unable to validate age-related differences.

6 Conclusion

We applaud password managers for introducing the market to automatically-generated passwords and workflows that help users to identify and replace weak, re-used, and compromised passwords.

Yet, there is a long way to go to improve password hygiene even among those using password managers. Almost all the users of password managers we surveyed kept re-used passwords they knew they should change. Roughly half of the users of third-party password managers still used passwords their password manager considered weak. Nearly 30% still had compromised passwords they knew they should replace. Many participants reported that they continue to prefer choosing passwords for new accounts that they can remember over strong random passwords generated by their password manager. Lastly, a quarter of those using third-party password managers protect their passwords with a "master" password that is itself re-used.

Interventions to improve hygiene are best designed with feedback from real-world use. If password managers are to deliver on their promise of improving password hygiene, they too should be measuring their efficacy.

Acknowledgements

This work was funded by the Center for Long Term Cybersecurity (CLTC) at UC Berkeley.

References

[1] 1password security. https://1password.com/security/. Accesed: 2022-02-04.

[2] Nora Alkaldi and Karen Renaud. Encouraging password manager adoption by meeting adopter self-determination needs. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[3] Joseph Bonneau and Stuart Schechter. Towards reliable storage of 56-bit secrets in human memory. In *23rd*

USENIX Security Symposium (USENIX Security 14), pages 607–623, 2014.

[4] Nick Charalambides. We recently went viral on tiktok - here's what we learned. https://bit.ly/3NnkAH4.

[5] Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. A usability study and critique of two password managers. In *USENIX Security Symposium*, volume 15, pages 1–16, 2006.

[6] Dashlane. https://www.dashlane.com/. Accesed: 2022-02-04.

[7] Jayesh Doolani, Matthew Wright, Rajesh Setty, and SM Taiabul Haque. Locimotion: Towards learning a strong authentication secret in a single session. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2021.

[8] Shirley Gaw and Edward W Felten. Password management strategies for online accounts. In *Proceedings of the second symposium on Usable privacy and security*, pages 44–55, 2006.

[9] Ambarish Karole, Nitesh Saxena, and Nicolas Christin. A comparative usability evaluation of traditional password managers. In *International Conference on Information Security and Cryptology*, pages 233–251. Springer, 2010.

[10] How lastpass works. https://www.lastpass.com/how-lastpass-works.

[11] Sanam Ghorbani Lyastani, Michael Schilling, Sascha Fahl, Michael Backes, and Sven Bugiel. Better managed than memorized? studying the impact of managers on password strength and reuse. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 203–220, 2018.

[12] David Ng and Stuart Schechter. Are password managers improving our password habits? In *RSA Conference*, 2021.

[13] Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Why people (don't) use password managers effectively. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pages 319–338, 2019.

[14] Prolific. https://www.prolific.co/.

[15] Hirak Ray, Flynn Wolf, Ravi Kuber, and Adam J Aviv. Why older adults (don't) use password managers. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 73–90, 2021.

[16] Rick Wash, Emilee Rader, and Chris Fennell. Can people self-report security accurately? agreement between self-report and behavioral measures. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 2228–2232, 2017.

A Free response answers to Question 9

Other responses to "Why do you feel it's okay to have some weak or reused passwords? (Check all that apply)"

Chrome participants:

- Some are passwords for website that do not contain any personal data nor directly an email address, so I don't really worry about waiting to change them - even though I know that is a security threat and I should replace them anyway asap
- When the website requires you to have a short amount of characters for a password.

Third-party participants:

- Some passwords reported as "reused" refer to the same Single-Sign Account that is accessed via multiple domains. My password manager is (and should be!) unable to detect this.
- Some are passwords that can't be changed.
- most of them is from local dev environment and no one except me has access to them
- For some websites that don't pertain to anything confidential or important, a weak, rememberable password suffices.
- I do not feel that it is ok, as it is a security risk, but I am too lazy to correct it
- I'm a developer. Most of those passwords are for sites that exist only on my machine

B Free response answers to Question 12

Other responses to "Why do you feel it's okay to have some compromised passwords? (Check all that apply)"

Chrome participants:

- Dead accounts, dead websites or accounts that aren't even mine
- These are all accounts that do not directly belong to me

Third-party participants:

- I checked and all of them are from deleted accounts. I should remove them from 1password completely
- Those accounts don't exist anymore
- When I followed the explanatory link I see one alert is a false positive. I won't be changing that.

C Peer reviews of prior drafts of this work

This paper will not appear at a peer-reviewed proceedings. Releasing our results in a timely manner is important because they may cause journalists and practitioners to question the assumptions they have made, in the absence of data, when when recommending password managers to consumers in hopes of improving password hygiene. We have waited to release the results for a better part of a year since conducting the study while awaiting feedback from both the Symposium on Usable Security and Privacy and USENIX Security. As a matter of full disclosure for those wondering what issues were disclosed by peer reviewers, you will find the totality of the feedback we received here.

In declining the work, reviewers concerns have focused primarily on style and subjective preferences for how we presented our work, or their subjective opinions of whether the research is important, and not scientific accuracy or integrity. Some reviewers have suggested that we re-run the study addressing as many of the limitations as we disclosed in Section 4. However, it is important to expeditiously release even imperfect results as they bring into questions assumptions that are being made in the absence of other data. Further, as we have no participant budget to re-run the study, and the students who ran the study have all graduated. The sooner we release our findings, the sooner others measuring the impact of password managers can learn from the limitations of our work as they design future studies.

C.1 Symposium on Usable Privacy and Security (SOUPS)

SOUPS 2022 Paper #20 Reviews and Comments
===

Paper #20 Do Password Managers Improve Password Hygiene?

Review #20A
===

* Updated: 13 May 2022 7:31:50am PDT

Overall merit

————-

4. Weak accept - While there may be some flaws, the paper has merit and we
should consider accepting it.

Reviewer expertise

——————

4. Expert: Historically an area of primary focus, or an area I have done
recent, significant work in.

Paper summary

This paper uses an online survey to investigate the quality of passwords users store in a password manager. This is done by investigating statistics reported by the password manager's credential audit tool. The survey also explores the reasons behind users' passwords' (in)security. This paper finds that in line with prior work by Lyastani et al., users manage a large number of insecure passwords. The paper concludes by stating that asking users to adopt a password manager is insufficient for them to gain the full security benefits available in a password manager.

Strengths

* The study was well described.
* Provides quantitative data showing that password managers' security functionality is underutilized.
* Fine-grained quantitative data about why users continue to manage weak passwords.

Weaknesses

* The paper uses the wrong style through, likely sidestepping the page limit.
* Sample size is rather small for a survey.
* Readability issues for the tables

Detailed comments for authors

Lyastani et al.'s work found that users continue to practice poor password hygiene even after using a password manager. This paper confirms these previous results. Moreover, it provides more details about the reasons users continue to practice poor hygiene. These two contributions merit acceptance in SOUPS. However, numerous issues with presentation clarity and a small number of methodological issues prevent me from rating this paper higher. Still, even with these flaws, I think this paper would be a good addition to the SOUPS program and could spur interesting discussion at the conference and future work.

### Presentation issues
These issues impeded the paper's readability, which is really unfortunate as I thought the actual results were really cool. Improving readability will help readers engaging with these great research results:

* The paper is not in the SOUPS style.
* Tables are not using the correct table environment, and are placed inline. This impacts the reading flow, as the text already describes high-level results, obviating the need to immediately look at results. I would much prefer them to be in floating environments that pulled the tables to the top for reference as needed, but not impede the flow of the text itself.
* The introduction is not self-contained. At the end of the introduction, I know what data the authors want to gather, but not how you did it, or a high-level understanding of your final results. As many readers will decide whether to read a paper based on a skim of the introduction, this stylistic choice is problematic. To encourage readers to delve into the very interesting results found in this paper, the introduction should be improved to address these shortcomings.
* The use of CDFs in Figures 2 and 3 was confusing. I think a non-cumulative graph would be preferable.
* Tables need to stay on a single page/column.
* Figure 4 is too small to read.

### Methodological issues
All human-factor studies have methodological issues. While I list several issues below, I do not think any of these rise to the level of requiring this paper to be rejected. I note them here for discussion by the PC and to allow the authors to thoughtfully address them in their limitations and future work discussion.

* The research method was intended to avoid only gathering responses from non-security-focused users. However, the results show that these results likely still have this bias. I'm willing to accept this bias, but I think the intro should be reframed to not make this bias sound like it would break the validity of the research, and hence the validity of this paper. **Not discussed in limitations.**
* The final sample size is rather small for a survey. However, based on my own personal experience gathering password manager users on crowdsourcing platforms I am not surprised by the difficulty of recruiting participants. As such, while a methodological limitation, I think it is likely to be faced by anyone doing this type of research. **Not discussed in limitations.**
* Plenty of minor textual mistakes in the study instrument. This is noted in the paper, but is still a little problematic. Discussed in limitations.
* It is not clear that a technical person couldn't doctor the screenshots. This would be as simple as dropping into the developer tools console. As such, I'm not sure the screenshot adds much value, but does allow for privacy loss (as was noted as having occurred several times in the paper). Not a huge issue. **Not discussed in limitations**
* The differing results between chrome and other third party users are interesting. This data would have been more interesting if further broken down by password manager. However, I doubt there were enough respondents to make this meaningful. **Could be included in Section 4.5.**

### Rebuttal
I thank the authors for clarifying the layout. Please note that formatting played no role in my rating or decision.

I appreciated the clarification regarding the limitations of Lyastani et al.'s work. I would recommend that those be made clear in a future revision of the paper (with care taken to not sound too harsh).

I was surprised that there was so much pushback on my recommendation for additional details on the limitations. After reviewing the rebuttal and the paper again, I stand by my statements that the issues I identified are not discussed in the limitations section. Related issues are. This aggressive pushback on my request for (minor) additions to the limitations sections made it hard for me to champion the paper, as it left me with little confidence that the requested changes would be made.

Overall, I think this paper is a "revise and resubmit". There is value in the results (agreed upon by all reviewers), but more polish is needed in the presentation. Unfortunately, at SOUPS, "revise and resubmit" is a "reject". I encourage the authors to update the paper and resubmit it elsewhere or at next year's SOUPS. With additional polish, I see no reason why it won't be published in the future.

Areas to address in response

——————————-

* How long is this paper when put into the proper format? I think it will likely be within reason to get it within the page limit, but I want to confirm.

Review #20B

===

Overall merit

————-

3. Weak reject - The paper has flaws, but I will not argue against it.

Reviewer expertise

——————

4. Expert: Historically an area of primary focus, or an area I have done
recent, significant work in.

Paper summary

————-

The paper presents the results of an online user study among password manager users. The goal of the study is to identify patterns in the password hygiene of users.

Strengths

———

* The used method including the multi-stage process is well thought out
* The methodology is mostly well laid out and easily understood
* Comprehensive descriptives on password manager usage and perceptions

Weaknesses

———-

* The closed-answer questions are derived from pilots that are not described in the paper
* The results read like a laundry-list of descriptive statistics
* There is hardly any discussion of and reflection on the results an their implications

Detailed comments for authors

——————————-

I am quite conflicted on this paper. The first part, including the method, is mostly well written and easily understood. The method seems sound and combines good approaches. Also, the descriptives presented in the results are quite comprehensive. Yet, I feel the paper was rushed to submission, leaving little time to reflect on the results and instead leading to cutting any discussion of the implications of the results (and also leading to formatting issues and using the wrong template).

## Results
My biggest concern with this paper is regarding the presentation of the results and their discussion. The results are mostly presented as a sorted laundry list of questions and answer responses with percentages. The table presentation form could have been changed into bar charts (in particular for the questions with yes/no and very frequently...never responses) to allow for an easier overview at a glance while also saving much space that could have been better used for a critical reflection of the results and a discussion of the larger implications of the results. More space for a proper discussion would have benefitted the paper greatly. Being only 1/4 page in length it reflects on three aspects: (1) few users have randomly generated master passwords where the paper cites research claiming that users of course should be able to remember such random strings (classical user blaming...); (2) support of higher reported reuse among users of browser-based password managers; (3) due to missing data they could not replicate findings concerning older adults. This is severely underwhelming. What can we learn from the results? How can we support users in improving their password practices? Many such questions arise for the reader but are left unaddressed in the paper.

## Method
The method refers to pilot tests for the selection of the closed question response options, but these pilot tests are

never described in any detail. For example, it is unclear how many of the options named in the pilots constitute the closed question response options. Unfortunately, the "other" responses from the main survey are also never elaborated upon. At the same time, the closed question response options seem quite specific and the question of overfitting the responses arises.

Considering the issues outlined above, I feel that the paper is not ready yet for publication. It shows much promise and an interesting research approach to a relevant topic and therefore holds the potential to bring valuable contributions if improved upon regarding the two aspects outlined before. But it just isn't there yet.

Minor

- On page 3, one of the questions has the annotation "boldface in original" and one of the words in that question is bold, which makes it unclear what was bold and what not in the survey.
- There is one line of text between the caption for Figure 4 and the Figure itself.
- This paper is using the wrong template.

Areas to address in response

----

* Why is there hardly any discussion of the results?
* How exactly were the closed question response options derived in pilots?

Review #20C

===

* Updated: 11 May 2022 9:09:21am PDT

Overall merit

----

3. Weak reject - The paper has flaws, but I will not argue against it.

Reviewer expertise

----

3. Knowledgeable: I know the area well (key related work is quite familiar
to me).

Paper summary

Paper explores whether password manager users utilise a number of features appropriately in the tool. Found that in fact many features are not used appropriately despite users relying on the password manager for everyday use.

Strengths

- Topical
- Potentially a very nice message that is actionable by the community and creators of password managers

Weaknesses

- Missing a literature review
- Some aspects of the study are not clearly motivated
- Key findings should be discussed in more detail, and put in the context of existing work.

Detailed comments for authors

This paper tackles an interesting and important area – password managers are promoted as important tools, but we do not know much about how they are actually used in practice by users. This study relies on the statistics produced by the password managers in order to determine whether their use actually results in better password hygiene. However, there are a number of issues with the presentation of the paper as well as some of the content.

The introduction for the paper does not follow a typical structure, and the motivation is more practical than academic. I would suggest that the authors use an academic research base for establishing the importance of their work, and it would be helpful to have clear contributions listed at the end of the section, in addition to an explicit research question(s) that will be answered in the paper.

The paper is also missing a detailed literature review. While some related literature is provided at the end of the paper, this is probably closer to a summary of findings situated in the literature (to be part of the Discussion) rather than an actual Background section – this should be between the Introduction and the Method, and cover the key areas of the field. While a number of key password manager papers are briefly touched upon, these can be expanded more and an evidence base for the need for password managers would also be beneficial (i.e., existing issues around password management, which then leads to password managers being a necessity).

The method is described in detail, although a clear overview of how the different steps of the study upfront (with a supporting diagram) would be helpful. The results are presented in quite a bit of detail. This is actually a weakness, in that all question results are presented as a table which takes up a lot of space that could be used to address some of the other issues (see Literature above and Findings below).

Findings – I feel that one of the key aspects of this work is obtaining the statistics from participants' password managers, yet this is not really discussed in much detail in the Findings. In fact, a graph is provided that is not very easy to understand, and very little text is dedicated to this part of the results despite possibly being the most novel aspect of the paper. At the very least the figure should have axes labels, and there should be a description of what the findings are in relation to the usage statistics and possibly some inferential statistics to better understand any potential discrepancies, in, e.g., reused password stored in the manager. A comparison with other literature discussing the password hygiene of users would also be beneficial here (or in the Discussion) to pinpoint whether users of password managers indeed have better hygiene (yet not perfect) or whether it is actually comparable.

The motivation to split the sample into third party password managers and built-in password managers (specifically Chrome) is not well motivated, although there is evidence to do so. The paper is discussed very briefly in Section 5, but should be used up front to motivate this design choice.

The Perceived Time to Replace passwords section is very short and does not really add much to the findings in its current state – what are the implications and is there any existing work that can be used to complement this finding?

Overall, while I think this can be a nice paper, in its current state it probably requires a bit too much work for this SOUPS cycle.

Minor:
- Question 1 table should probably be sorted by % rather than alphabetic order of password manager (like other tables)?
- Figure 3 (puppy picture) is indeed adorable!

Post Rebuttal

I thank the authors for their rebuttal. While I appreciate the offer to add literature, the texts (or examples) are not included in the rebuttal, which makes it difficult to assess what the section might look like. As such this feels as a major revisions paper, which SOUPS does not accept. No other issues were addressed in the rebuttal. As such, my score remains unchanged and I lean towards rejecting this paper.

Review #20D
===
* Updated: 13 May 2022 6:07:10am PDT

Overall merit
—————-
3. Weak reject - The paper has flaws, but I will not argue against it.

Reviewer expertise
——————
4. Expert: Historically an area of primary focus, or an area I have done
recent, significant work in.

Paper summary
—————-
The paper presents the results of a prolific-based online survey which investigated the password hygienes of password manager users. The sample included browser-based (Chrome) and app-based (like Bitwarden) users. The researchers analysed submitted password check statistics and found that users often store compromised, reused, and weak passwords. Overall, users of third-party password managers showed more secure password managing behaviour.

Strengths
———
+ The paper addresses a very important research problem. Indeed, password managers can only help to increase online security if they are used in a secure way.
+ The paper provides interesting descriptive insights into current password management behaviour and indicates a systematic difference between browser-based and app-based password management.

Weaknesses
———-
- The paper lacks completeness since the presented data is neither discussed nor interpreted.
- The related work section is rather short and does not support the reader in identifying the paper's contribution.
- The contribution seems rather small.

Detailed comments for authors
——————————
I'd like to thank the authors for submitting their paper to SOUPS 2022. I enjoyed reading the paper which is overall well written and easy to read. As stated above, I argue that this work addresses a critical problem space. The research approach seems feasible and the study was well conducted. While there is a lot to like about the paper, I'd argue that it should not be accepted in its current form. I'm not convinced that the contribution hits the bar for SOUPS paper.

I recommend to address the following aspects in order to improve the overall contribution.

1. Tone down presuming claims.

The paper hypothesises that vendors of password managers are not measuring user behaviour since "developers might reasonably fear that
measurements could only act to subvert the narrative that password managers improve security". The paper doesn't provide any references for this claim and I feel that the claim is unnecessary overall. Similarly, I'd recommend to tone down claims where the paper states that "participants are lying" since this doesn't seem respectful.

2. Discussion and implications.

The paper indeed presents interesting insights into password management behaviour. I'd argue that the systematic differences which have been found between browser-based storage and third-party users are particularly relevant. Unfortunately, there is no discussion of those results. The paper concludes with a very general claim:

"To improve hygiene, users need to understand that procuring the product is only the first step, and those building and marketing these products need to do more to ease and encourage proper use, as well as to measure the efficacy of these interventions."

I recommend to point out how the insights gained from this paper can help to improve the products. I'd like to read the authors' thoughts about the findings and how they inform future research and product development. In particular, the paper needs to better position itself in the light of previous work. As stated by the authors, most findings have been reported in previous work (e.g., tendency to use self-selected passwords).

3. Figures.

Figure 2-4 are somewhat difficult to parse. I would recommend to explain how the figures should be read. In addition, the authors might consider using a simpler representation of their data.

Nits:
- page 10 has an empty bracket ()
- page 11 has some format issues since the main text is mixed up with the figure caption.

== Post-rebuttal ==

I'd like to thank the authors for providing a response. Unfortunately, the rebuttal did not address my concerns. After reading the rebuttal I was still not sure how the paper would be improved based on the reviews.

Re: "participants are lying" - a quick search reveals that the paper stated : "[..] allowed us to overcome many (but perhaps not all) instances in which participants might be lying about using a password manager, might be lying about having collected the requested statistics, or might be incorrectly reporting statistics due to honest errors" - I would still recommend toning down such claims for the next submission.

Areas to address in response

———————————-

To improve the contribution of the paper, the authors need to clearly point out

- which novel insights we gain from the presented data AND
- how those insights inform future research or development.

Response by Stuart Schechter <stuart.schechter@gmail.com> (480 words)

———————————————————————————

—

We appreciate our reviewers' comments and insights.

We used the Word template from the SOUPS CFP to be inclusive of a collaborator uncomfortable with latex. We agree the prescribed template is atrocious. It gave us no unfair advantage. Reformatting to latex shrinks tables by 10-50%, as one can see by observing the excessive padding Word added to table cells. We can use the space to address reviewer feedback.

Our contribution is quantifying the (large) fraction of password manager users who continue to use weak, re-used, and compromised passwords, across a wide swath of products. Our work alerts about password managers not being the panacea they purport to be. Reviewer A suggests that Lyastani et al. already "found that users continue to practice poor password hygiene," but those findings came from only 128 password-entry events with 100 unique passwords (Table 5) from 49 participants (Table 11), all using LastPass. That's under 3 passwords observed per participant. From our reading, it's possible all those re-uses came from two participants.

Reviewer A is concerned that our limitations section fails to disclose the potential for sampling bias (it's in 4.1p2), potential failure to detect forgeries (see 4.4p1), and sample size (already well documented in Results). We can explain that developer tools allow doctor reports for web interfaces, but the motivation to provide doctored reports appears to be obtaining compensation without generating a report (those pretending to use managers could not generate one). Removing PII remains easier than doctoring for participants who actually use password managers. With regard to investigating differences between password managers, our participant budget of $5,000 did not afford us enough responses to make reliable comparisons. We would gladly tabulate unaggregated data in the appendix.

Reviewers C and D requested more related work. We are happy to cite work they would want added. Following the "common pitfalls" guidance attached to previous-year's CFPs, we cited only work warranting an individual description of why it is relevant, avoiding padding with superfluous references. That "common pitfalls" guidance also explains why some academic disciplines choose not to put related work between the introduction and methodology, as Reviewer C expects. Rather, like Reviewer A, we anticipate readers of our introduction will immediately want to know more about the methodology, so we get straight to it.

Reviewer B requested more discussion of results. We erred to allow readers to draw their own inferences, but agree we should add more direction and cleaner transitions. They also inquired about the use of the pilots. We read through free-response options in pilots to identify multiple choice options we failed to anticipate or that were misunderstood. We did not use formal bucketing approaches.

Reviewer D asked us not to say "participants are lying". We never used that phrase. Similarly, the middle two words of "developers might reasonably fear" clearly indicate we are positing a possible explanation and not making a "presuming claim."

C.2 USENIX Security

USENIX Security '23 Summer Paper #11 Reviews and Comments
===

Paper #11 Do Password Managers Improve Password Hygiene?

Review #11A
===

Paper summary

———————-

This paper investigates the password habits of password manager users. Participants in the study shared screenshots of their password manager apps, showing statistics such as the number of reused, weak, and compromised passwords. The researchers found that these numbers were relatively high and that many participants self-reported coming up with their own passwords, rather than relying on auto-generated ones from their password manager.

Detailed comments for authors

———————————————

This paper's topic is interesting and important. Adopting password managers is standard security advice, but relatively little research has examined whether they are used correctly and effectively.

I found the paper's research goals and methods compelling. Focusing the research questions on the number of reused, weak, or compromised passwords is apt, because the use of password managers is supposed to mitigate this, and it provides a concrete metric. Having data on this is helpful for knowing if password managers meet their goals.

I also found the methods for collecting this data to be clever and effective. By having participants share screenshots of their password managers, the data is likely to be reasonably authentic without compromising participants' privacy.

I do have some concerns about the paper, and they fall in two categories: data and presentation.

The main data issue is the one acknowledged by the paper: the large discrepancy between the participation rates of Chrome and non-Chrome users discussed in 3.2. I appreciate that the authors were forthright about this and put forward several potential explanatory theories. I'm inclined to believe the first hypothesis, that this a real difference among users, but with a discrepancy this significant, I think it's important to be sure it's not a data collection issue. If at all possible, I would suggest that the researchers replicate at least the consent portions of the study to see if this behavior continues to hold.

Another data-related question is: what proportion of final study participants used different types of password managers other than Chrome? (Question 1 lists only the number at the screening stage.)

I also did not see a justification for the sample size used in the study and how it was determined.

Relatedly, I was confused about the exact details of the participant numbers. The paper states that 71 and 190 Chrome and third-party participants consented to participation, but the final number of participants are 61 and 81, respectively. What happened to the other participants?

I was also unclear about the timeline of participation, as implied by this sentence: "We sent 53 surveys individually to respondents who asked to be contacted later." Under what circumstances did this happen?

More details about the pilot and how it determined the final survey would have been useful.

Additionally, the paper has many presentation issues, especially when it comes to focus and organization.

The title (Do Password Managers Improve Password Hygiene?) suggests a comparative approach (improve relative to what?), which the paper doesn't follow. I would instead suggest something like "The Password Hygiene of Password Manager Users."

The introduction, in my opinion, would be better if focused on the specifics of the study, rather than selling its contributions ("this study is so needed") or justifying its approach ("we are at a disadvantage"). In particular, I would strongly suggest that the introduction include the specific research questions the study sought to address.

Section 2 belabors the study's methods. The narrative style and the verbatim inclusion of consent, explanations, and questions made the overall procedures difficult to follow. I suggest significantly shortening this section, leaving only what's needed to understand the study. A flowchart or diagram could perhaps be useful for understanding the flow. Any other details (consent, explanations, questions) would be better placed in an appendix.

The results section includes all questions, answer choices, and response counts. This is great for reproducibility, but in my opinion makes the paper difficult to follow: it's very difficult to pay attention to all the tables and extract meaning from each one. My suggestion would be to pick out a few key results and focus on those. Additional questions (while still a valuable contribution) would again likely be better off in an appendix.

The related work could be improved by discussing how the study and its findings differ from prior research. Additionally, in light of this study's findings about poor password hygiene, the related work may benefit from covering the various literature on the password practices of people who don't use password managers and the implications of that. Also, there are some recent papers on password managers that the authors may want to include in the related work [1,2]. Finally, I would suggest putting the Related Work section second in the paper, to provide more background and context earlier.

The paper lacks a discussion, which could help synthesize the study's takeaways and situate them in the context of prior work.

There are many typos and misspellings, and a broken reference on page 2.

[1] https:/doi.org/10.1145/3491102.3517534
[2] https://www.usenix.org/conference/usenixsecurity22/presentation/mayer https://collinsmunyendo.github.io/papers/2022_USENIX_Password_Managers.pdf

Reasons to accept the paper

- Interesting topic
- Effective methods
- Novel results
- Thorough survey

Reasons not to accept the paper

- Potential data reliability issues
- Missing study details
- Numerous writing and presentation issues

Recommended decision
----

4. Reject

Writing quality
----

4. Needs improvement

Reviewer confidence
----

3. Highly confident (would try to convince others)

Review #11B
===

Paper summary

This paper presents the result of a survey of users who use password managers and what their password hygiene practices are: whether they have weak or reused passwords, whether they change passwords that have been compromised, etc. Survey responses are broken down among users of Chrome's password manager and third-party password managers (such as 1Password, LastPass, etc.), and are corroborated by screenshots of password manager UI reporting the number of weak, compromised, or reused passwords in use. The study finds that many password manager users have poor password hygiene and don't take full advantage of their password manager's features, such as generating strong random passwords for new accounts.

Detailed comments for authors

This paper has a strong premise: while password managers are often cited as an important end-user security measure, the fact that a user uses a password manager doesn't mean that they actually have an improved security posture. Their security only improves if they used the password manager's features as intended.

While I think the topic area and premise are promising, the contribution of the paper in its current form isn't enough for Usenix Security. As written, it could make a good workshop paper, or it could be expanded and built upon to make a solid Usenix paper. Here is some more detail about what could be improved:

* The survey could be more scientifically rigorous. Currently there are no clearly defined research questions, no hypotheses, and no control groups (e.g. users who don't use password managers at all). For example, while it is discouraging that half of third-party password manager users use weak passwords, if this statement is true of nearly all users who don't use password managers at all, the use of a password manager is still a big security improvement, even if it leaves many users with weak security still. Typically a survey like this with no hypothesis could make good exploratory research but doesn't stand on its own as a paper – for example, from your survey you could derive the hypothesis that password managers don't effectively steer users away from weak passwords relative to users who don't use password managers, and then conduct a more rigorous study to try to prove this hypothesis.
* There are no statistical tests or justification for the survey sample size. (This is related to the lack of hypothesis noted above.) As noted in Section 4, the demographics of the sample are not representative of the general population or even of password manager users, so it's not clear what conclusions (if any) we can draw from the work.
* The free-form answers should be coded rather than analyzed informally (see e.g. https://gradcoach.com/qualitative-data-coding-101/).
* The survey had many limitations, from spelling/grammatical errors to omitted questions. Section 4 does a good job describing and exploring these limitations, but many of them would be easily surmountable by re-running the survey.
* The related work section doesn't make it clear what the novel contributions of this survey are compared to prior work. Is the goal of this study to confirm prior findings (if so that should be noted explicitly), or does it bring something new?

Here are some smaller comments as well:
* The introduction struck me as unusual in its tone and format. Usually an introduction clearly defines the problem and why it's difficult, the novel ideas or data that the authors present in the paper, and the main contributions and findings. I'd encourage the authors to read some other related papers and model the introduction off them – it makes it easier for the reader if they follow a predictable pattern.
* Missing reference ("Section ??") in Section 2.1
* It's unclear why the survey results are stratified into Chrome users and third-party password managers. What about other browsers? And why lump all the third-party password managers together?
* Could the answers to all the survey questions be in randomized order? That would reduce biases.
* The authors go into perhaps more detail than necessary about their consent process and why they filtered out certain users (e.g. Table 1).
* Footnote 4 is a bit mysterious – why was a user able to leave a mandatory question unanswered?
* It would be interesting to also evaluate how well password managers protect users from phishing by refusing to fill credentials on phishing sites.

Required changes

---

* Re-run survey as a more rigorous scientific study (per suggestions above)
* Consider if there are ways to expand the contribution further, e.g. designing a better password manager UI based on your findings, or more deeply understanding why password managers don't improve users' security as much as they could
* Make the novelty of the contributions more clear/explicit

Reasons to accept the paper

+ The paper studies an important problem as password theft and phishing are ubiquitous attacks that affect lots of people, and very much unsolved.
+ The paper's premise – that the use of a password manager is not necessarily enough to make a person secure – is compelling.

Reasons not to accept the paper

- Contribution isn't large enough – no clear conclusions due to lack of representative sample, and unclear what the novelty is over existing work
- Lacks scientific rigor (no clearly stated research questions,

no hypotheses, no statistical tests or qualitative coding procedures)

Recommended decision

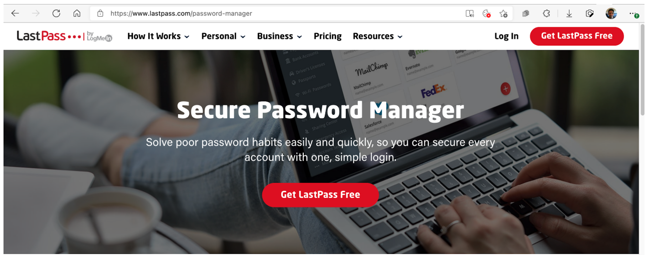
4. Reject

Writing quality

2. Well-written

Reviewer confidence

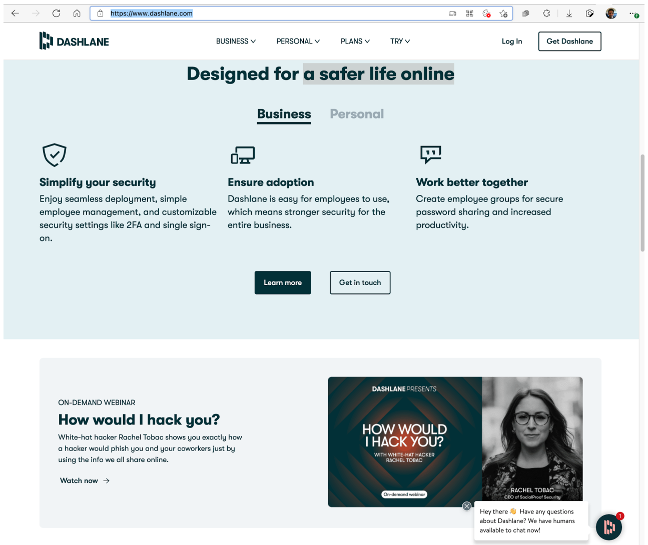
3. Highly confident (would try to convince others)



Secure Password Manager

Solve poor password habits easily and quickly, so you can secure every account with one, simple login.

(a) LastPass



Designed for a safer life online

Simplify your security

Enjoy seamless deployment, simple employee management, and customizable security settings like 2FA and single sign-on.

Ensure adoption

Dashlane is easy for employees to use, which means stronger security for the entire business.

Work better together

Create employee groups for secure password sharing and increased productivity.

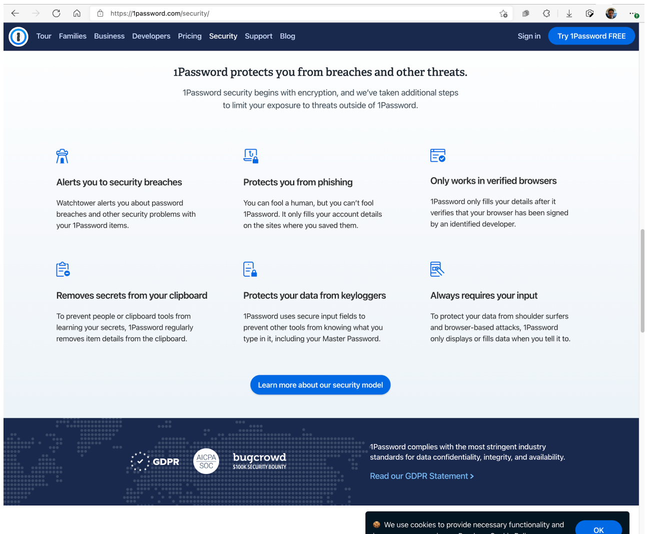
ON-DEMAND WEBINAR

How would I hack you?

White-hat hacker Rachel Tobac shows you exactly how a hacker would phish you and your coworkers just by using the info we all share online.

Hey there 👋 Have any questions about Dashlane? We have humans available to chat now!

(b) DashLane



1Password protects you from breaches and other threats.

1Password security begins with encryption, and we've taken additional steps to limit your exposure to threats outside of 1Password.

Alerts you to security breaches

Watchtower alerts you about password breaches and other security problems with your 1Password items.

Protects you from phishing

You can fool a human, but you can't fool 1Password. It only fills your account details on the sites where you saved them.

Only works in verified browsers

1Password only fills your details after it verifies that your browser has been signed by an identified developer.

Removes secrets from your clipboard

To prevent people or clipboard tools from learning your secrets, 1Password regularly removes item details from the clipboard.

Protects your data from keyloggers

1Password uses secure input fields to prevent other tools from knowing what you type in it, including your Master Password.

Always requires your input

To protect your data from shoulder surfers and browser-based attacks, 1Password only displays or fills data when you tell it to.

1Password complies with the most stringent industry standards for data confidentiality, integrity, and availability.

We use cookies to provide necessary functionality and

Figure 6: An image uploaded in response to our request for a screenshot of password hygiene statistics. We thus deemed the participant disingenuous (if admittedly adorable).