Unmasking the Security and Usability of Password Masking

Yuqi Hu
Georgia Institute of Technology
Atlanta, Georgia, USA
yhu476@gatech.edu

Suood Alroomi
Georgia Institute of Technology
Atlanta, Georgia, USA
Kuwait University
Kuwait City, Kuwait
roomi@gatech.edu

Sena Sahin
Georgia Institute of Technology
Atlanta, Georgia, USA
ssahin8@gatech.edu

Frank Li
Georgia Institute of Technology
Atlanta, Georgia, USA
frankli@gatech.edu

Abstract

Password masking, a practice where passwords are obscured during entry, is widely adopted for online authentication. However, its merits have been debated for over a decade, with questions about its security benefits and concerns about its usability impact. Yet to date, masking has received limited prior exploration.

In this work, we empirically investigate the security and usability impact of password masking. We first assess the masking practices of popular browsers and websites, demonstrating masking's ubiquity as well as its design diversity. Guided by our real-world observations, we then conduct a mixed-method evaluation of masking for both mobile and PC devices, combining a survey of over 200 participants on their experiences with and perspectives on masking along with user experiments of 600 participants performing password logins under varying masking conditions. Through our study, we uncover misconceptions about masking, masking's usability and security impact, and user preferences on masking's use and its design. Ultimately, our study establishes empirical grounding on how this popular technique manifests in practice, providing recommendations for its use moving forward.

CCS Concepts

• **Security and privacy → Usability in security and privacy**; **Authentication**.

Keywords

Password masking, Online authentication, Shoulder surfing, User studies

ACM Reference Format:
Yuqi Hu, Suood Alroomi, Sena Sahin, and Frank Li. 2024. Unmasking the Security and Usability of Password Masking. In *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS '24), October 14–18, 2024, Salt Lake City, UT, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3658644.3690333

CCS '24, October 14–18, 2024, Salt Lake City, UT, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0636-3/24/10
https://doi.org/10.1145/3658644.3690333

1 Introduction

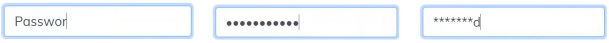
Password masking is a practice widely adopted for online authentication, where user passwords are hidden as they are typed (with different masking variations illustrated in Figure 1). While conceptually simple, the merits of password masking have been debated for well over a decade.

Proponents advocate for masking's necessity in preventing the unintended exposure of passwords during entry (i.e., protection against shoulder surfing). For example, masking is recommended by US NIST's latest Digital Identity Guidelines [14]. Similarly, unmasked passwords are considered a security weakness in the Common Weakness Enumeration (CWE) system (with CWE-549 defined as "Missing Password Field Masking").

However, detractors [6, 11, 20, 27] have questioned whether masking's security benefits are worth the usability impact, as users may be unable to observe and correct mistakes during password entry. For example, security specialist Bruce Schneier [1] argued that "Shoulder surfing isn't very common, and cleartext passwords greatly reduce errors. It has long annoyed me when I can't see what I type" [27]. Meanwhile, web usability expert Jakob Nielsen [2] maintains that "Usability suffers when users type in passwords and the only feedback they get is a row of bullets. Typically, masking passwords doesn't even increase security, but it does cost you business due to login failures" [20].

Despite masking's popularity and controversy, there has been limited prior investigation into the security and usability of password masking, as well as user perspectives on its use. To our knowledge, the most relevant prior work was conducted by Pidel and Neuhaus in 2019, where they monitored password entry with and without masking, but only for 10 study subjects using one masking design on mobile devices [21].

In this paper, we seek to close this research gap and establish a broader empirical grounding on the security and usability impacts of password masking. Our study entails three distinct methods of assessing password masking.


Passwor

•••••••••••

*******d

(a) No Masking **(b) Static Masking** **(c) Dynamic Masking**

Figure 1: Illustrations of different masking variations.

We begin by evaluating how masking is implemented by real-world websites, investigating sites within the Google CrUX Top 1K [13], as well as by popular mobile and PC browsers. Through our real-world assessment, we confirm masking's ubiquitous support by sites and browsers. We also observe varying masking designs across sites, as well as between PC and mobile browsers, demonstrating a lack of standardization on masking's use. These real-world observations motivate and guide our subsequent study phases.

We next conduct a user survey of over 200 participants, across both mobile and PC environments. Through this survey, we examine users' understanding of, experiences with, and perspectives on password masking, considering both usability and security aspects. We uncover that while masking's core purpose is understood by the majority of users, a sizable minority exhibit misunderstandings about masking's properties. We further characterize the usability pain points introduced by masking, as well as the situations where masking provides security benefits. Crucially, we also determine user preferences for masking's use and its design.

Finally, we experiment with 600 participants performing password logins under varying masking conditions for both mobile and PC. We simulate the account login workflow for an online payment service, and monitor login usability metrics such as the number of login attempts needed before success, time required for password entry, use of masking toggling, and errors in submitted passwords. We compare these metrics between the different masking conditions, observing that while masking appears to have some impact on login usability, the impact is overall limited.

Altogether, our study provides the first broad, systematic investigation of the security and usability impact of this commonplace security measure, informing the ongoing debate about its use. Furthermore, despite the simplicity of masking, we identify unexpected consequences of its use, including misconceptions about its functionality, both positive and negative interactions with password managers (with lessons for password manager designers), and influences on user password selection. We also enumerate the common settings where password shoulder surfing would occur in practice with the absence of password masking. By combining the results across our multiple methods, we provide empirically grounded recommendations about masking's use and its design. Our findings also illuminate how masking can serve as a unique case study for future research, particularly toward exploring user security education and the adoption of security mechanisms.

2 Related Work

Here, we discuss prior work related to password masking.

2.1 Password Masking Evaluation

Despite password masking's prominence, there has been very limited prior evaluation of password masking. The most relevant study was conducted by Pidel and Neuhaus[21], who performed a small-scale study of users entering passwords with and without masking, specifically on mobile devices. Participants entered their passwords three times across two sessions, one with masking and one without masking. This study was performed with 10 participants and found no significant differences in typing speed, error rates, or the need for corrections between masked and unmasked password entries.

Our study expands significantly upon this earlier investigation, by performing a significantly larger-scale experiment (with 600 participants) that explores multiple masking designs for both PC and mobile, and combining the insights derived with those from a large-scale user survey of password masking experiences and perspectives. Thus, our work seeks to provide a more comprehensive evaluation of password masking's security and usability impacts.

2.2 Shoulder Surfing

Password masking is designed to defend against the threat of password shoulder surfing, where an attacker learns of a victim's password through observing their password entry. Prior work has studied general shoulder surfing behavior, more broadly than for password entry. For example, Eiband et al. [10] surveyed 176 individuals about real-world shoulder surfing scenarios on mobile devices. Across diverse participant demographics, the study found that shoulder surfing primarily occurs opportunistically and non-maliciously, resulting in minimal tangible harm but a significant negative emotional impact on the individuals affected. Similarly, Saad et al. [26] simulated shoulder surfing scenarios using virtual reality, and through monitoring participants, observed likewise that shoulder surfing was occasional, opportunistic, and participants only partially recalled what they observed.

Such studies are informative about the threat model for password masking, but have focused on shoulder surfing generically. Unlike our study, these prior works do not shed light on user experiences and preferences with masking, nor its usability impact.

2.3 Improved Password Masking Designs

Several studies have considered alternative ways of obscuring password entry. Khamis et al. [15] compared several different visual filters with the traditional asterisks used in masking. They evaluated the visual filters' effectiveness during password entry, editing, and protection against shoulder surfing across different password types. The studies concluded that some of the visual filters improve editing speed, accuracy, and resistance to observation. Similarly, Alt et al. [3] explored visual masking for graphical passwords. The study monitored 26 participants using visually-masked graphical passwords for a month, observing a significant learning curve but enhanced long-term memorization. Their study highlights both the efficacy of graphical passwords, and the potential for masking with graphical passwords.

While the prior works provide promising directions for improving password masking in the future and shed some light on current issues with masking, they do not systematically investigate password masking as used in practice today.

3 Real-World Measurement

To provide grounding for our subsequent user studies, we first evaluated how password masking is implemented in practice, both by popular browsers and websites. Here, we sought to understand the popularity of password masking, as well as the common forms of masking deployed.

3.1 Password Masking by Browsers

A common way to implement password entry in an HTML form is through using a form `input` element with `type="password"`. Such password input elements can implement password masking by default, with variation across browsers and OSes [9]. We first investigate how various browsing environments implement masking for form password inputs.

Method. We evaluated the four most popular browsers[1] for PC [28] (Chrome, Safari, Edge, Firefox) and mobile platforms [29] (Chrome, Safari, Samsung, Opera). For PC, we investigated the browsers on Windows and Mac OS, while for mobile, we tested the browsers on Android and iOS (to the extent that a browser is supported by an OS; for example, Safari does not officially support Windows). For each browser, we opened a simple test HTML page with an HTML form containing a password input field and manually observed the masking implemented as we entered characters.

Results. We observed the following browser masking behaviors:

- All browsers implemented masking without a toggling option for password input elements.
- All PC browsers, on both Windows and Mac OS, applied *static masking*, where each character is masked as typed (see Figure 1b).
- All mobile browsers, on both Android and iOS, implemented *dynamic masking*, where a recently typed character is unmasked, but previously-typed characters are masked (see Figure 1c).

From these observations, we see that browsers broadly support masking by default for form password inputs[2], and different platforms (mobile versus PC) implement masking differently, with consistency across browsers on the same platform. Notably, mobile browsers offer partial visibility into the typed passwords, presumably reflecting that password entry can be more difficult within a mobile environment. Thus, in our subsequent user studies, we consider mobile and PC environments separately.

3.2 Password Masking by Websites

We next characterized masking as deployed by popular websites.

Method. We manually evaluated 100 domains randomly selected from the Google CrUX Top 1K domains list [13], snapshotted on September 30, 2023. We chose the CrUX top list as our focus is investigating masking as deployed by websites, and CrUX is based on browser web traffic telemetry. (Note that CrUX only provides domain rankings in ranges/buckets, rather than individual domain ranks; thus we lack more specific ranks for the sampled domains.)

For each domain, we manually visited the landing page and searched for an account login form or a link to an account login page. We further searched for the login URL on Google Search. If we found the login form, we entered arbitrary characters into the password input field to observe if and how masking was enabled. We followed this workflow on both a PC and mobile browser (specifically, Google Chrome on Mac OS and Android).

We note that an automated approach could afford a larger-scale measurement. However, recent automated methods [25] only capture a fraction of website logins, and without *a priori* knowledge of the different masking design variations, we still require manual inspection to categorize the masking deployments. Future work can investigate real-world website masking at broader scales.

Masking	Toggling	Toggle Design	No. Sites
Yes	No	-	41 (62%)
Yes	Yes	Opened Eye	15 (23%)
Yes	Yes	Closed Eye	6 (9%)
Yes	Yes	Show Password Button	4 (6%)
No	-	-	0

Table 1: Password masking by 100 websites randomly sampled from the Google CrUX Top 1K domains. We could manually evaluate the login password entry for 66 sites. We depict the number and percentage of these sites that deploy masking by default and a method to toggle the masking off. We also show the distribution of toggle button designs.

Results. Out of the 100 randomly sampled Top 1K sites, one was inaccessible. We also could not find an account login on 28 sites (mainly those with news/weather or free adult content). For 5 of the remaining sites, we identified their login page but could not proceed to password entry without valid account information. We did not attempt to create test accounts on these sites, as some did not offer public account registrations (e.g., financial sites).

Thus, we evaluated password masking for 66 top sites. Table 1 summarizes our findings. Our key observations are:

- All 66 sites employed password masking by default on their login forms (either through using a password form input element or implementing their own custom masking method). Besides consistently observing static masking on PC and dynamic masking on mobile, we did not observe masking variations across the two platforms for the same site.
- Out of 66 sites, 41 sites did not support a method to toggle off or disable password masking. Thus, users could only enter the password masked.
- Of the 25 sites allowing masking to be disabled, we observed various toggle designs. While masking was enabled, we found 15 sites with a toggle button depicting an open eye while 6 other sites used a closed eye image. We found 4 sites providing a "show password" checkbox for revealing the password.

Thus, our findings (even at a limited scale) demonstrate password masking's ubiquity and its diverse implementations (with and without toggling, and different toggling icon designs). In our subsequent user studies, we investigate how users experience and engage with masking, as well as preferences for masking forms.

> Password masking is widely supported by default, both by popular browsers and websites. Masking forms differ in design though, particularly in terms of static versus dynamic masking (between PC and mobile environments, respectively), as well as whether masking toggling is supported.

4 User Survey

We next surveyed 202 end users to investigate their understanding and experiences with password masking, as well as their preferences with masking methods. Our goals with this survey were to

[1] We consider the most popular browsers as of September 2023.

[2] Websites can still control password masking by using a form text input (`type="text"`) for password entry, and implementing password masking via JavaScript.

identify misunderstandings or unintended consequences of masking, preferences on masking's deployment, and perspectives on masking designs.

4.1 Method

We first discuss our method for conducting our user survey, including how we designed and deployed the survey, and analyzed the resulting data. This user survey was reviewed and approved by our organizations' Institutional Review Boards (IRBs).

4.1.1 Survey Design. We created an online survey deployed on the Qualtrics survey platform [24] to allow for scaling to a larger number of participants. The design of our survey instrument was informed by the existing debate about masking's security versus usability tradeoff [20, 21, 27]. However, we aimed to construct general questions about usability and security issues regarding masking, to avoid priming participants about specific masking concerns and afford broader participant responses. Our questions were further informed by our real-world measurements (see Section 3), such as incorporating the real-world masking designs observed. Also, as we observed masking differences between PC and mobile environments, we chose to develop two parallel versions of the survey, one specifically for each environment, asking participants both during the survey introduction and throughout the survey questions to respond only based on one particular environment.

Our final survey contained 22 questions in total (shown in Appendix A), including an attention question to help filter out survey responses from disengaged participants. Excluding the attention question, our survey consisted of five distinct sections:

(1) *Demographic Information:* The first section of the survey focused on gathering demographic data from participants, specifically age, gender, and education level. Additionally, participants were asked whether they use password auto-fill (e.g., a feature of many password managers) when logging into accounts, as auto-filling changes the user's password entry experience.
(2) *Past Experiences with Masking:* Next, we delved into participants' past encounters with password masking, prompting them to reflect on why websites use masking and their frequency of encountering masking during recent account logins.
(3) *Masking Usability Considerations:* We then explored the impact of password masking on authentication usability. Participants were asked to describe if and how masking introduced challenges during password entry.
(4) *Masking Security and Privacy Considerations:* The next section dived into participants' perceptions of the security and privacy implications associated with password masking. Participants were asked to explain situations where someone could observe their password entry, and how often those situations arose. They were further asked about their perspectives on the security and privacy benefits of masking, including across different categories of sites.
(5) *Preferred Masking Format:* Finally, participants were prompted to indicate their preferences regarding different masking settings, including default masking options and the use of toggle options to reveal masked passwords.

We relied upon both closed-ended and open-ended questions, typically using closed-ended questions for participants to express preferences or frequencies[3], and open-ended questions for participants to expand further on their perspectives.

Pilot Testing. During the survey design process, we conducted two rounds of pilot testing in February 2024, to assess and refine our survey instrument. In each round, we recruited 10 participants in total via Prolific (using the same recruiting method as our full-scale survey, as will be discussed in Section 4.1.2), with 5 participants each for the mobile and PC versions of our survey.

In our first pilot round, we observed some participants exhibiting confusion about password masking and its different forms. To address this, we simplified our survey language and incorporated visual aids (images and GIFs) to concretely illustrate password masking forms. Additionally, we refined the attention question to better identify disengaged participants, as some participants failed our initial attention question but provided meaningful and detailed responses otherwise. Finally, we observed some participant responses discussing the wrong environment (e.g., a participant in the PC version of the survey discussing experiences in a mobile setting). To more reliably elicit responses specifically for each environment, we introduced additional visual cues related to the survey's environment and reinforced the environment-specific context within the survey introduction and throughout the questions. Given the issues identified with this first pilot round, we did not include these pilot participants in our subsequent survey results.

With our updated survey instrument, we did not observe the same issues during our second pilot round (nor did we observe similar issues in the full-scale survey). We found that participants gave meaningful, detailed, and reliable responses. Thus, we chose to include the survey responses from the second pilot round in our subsequent analysis and concluded our pilot testing.

4.1.2 Recruitment and Survey Deployment. Our survey took approximately 7–8 minutes to complete. We recruited US-based adults (18+ years old) via the Prolific platform [23], offering $2 USD compensation upon completion (a $15/hour compensation rate). For our full-scale survey deployment, we recruited 200 participants in February and March 2024, 100 participants each for the mobile and PC versions of our survey. As a result, our total survey data consisted of 210 participants when including the survey responses from our second pilot round. However, we observed that 7 participants failed the attention check (5 on PC, 2 on mobile) and 1 participant on PC discussed the mobile environment, so we filtered those responses out. Thus, our final dataset consisted of 99 PC responses and 103 mobile responses, for a total of 202 participants.

4.1.3 Data Analysis. Our survey responses contain answers to both closed-ended and open-ended questions. We analyzed closed-ended responses using frequency analysis to identify patterns. We also calculate descriptive statistics on quantitative data to summarize central tendencies and variations in participants' responses.

To analyze the open-ended responses, we applied open coding to develop categories directly from data related to experiences and perceptions of password masking [8]. For each open-ended question, two researchers independently developed their own codebook

[3] We gauged frequencies primarily using 5-point Likert scales, so that participants need not recall exact numbers and instead report frequencies at an interpretable granularity.

		Mobile	PC
Age	18-24	12%	13%
	25–34	42%	35%
	35-44	19%	24%
	45-54	17%	9%
	55-64	4%	12%
	65+	6%	6%
Gender	Male	55%	54%
	Female	40%	43%
	Others	5%	3%
Education	High School or Less	15%	15%
	Associate / Some College	24%	21%
	Bachelor's	49%	47%
	Master's / Doctorate	13%	16%

Table 2: Demographics of our survey participants

to characterize participant responses, before meeting to discuss and converge on a final codebook. Subsequently, the two researchers independently coded the participant responses using the final codebook. Finally, the researchers met to converge on the final codes. Throughout both the codebook development and subsequent coding, any disagreements unresolved between the two researchers were further discussed and decided by two additional researchers.

In total, our analysis produced 77 codes (across 6 open-ended questions), with 818 coded segments for mobile and 759 for PC. To assess whether coding was consistent between the independent researchers, we computed the Kupper-Hafner inter-rater reliability scores [16], yielding an average agreement of 0.84. This score indicates largely consistent coding across the survey responses.

4.1.4 *Participant Demographics.*

Here we describe the demographics of our participants for both the mobile and PC versions of our survey, as detailed in Table 2.

Age. Our participants skewed towards younger individuals: 42% for mobile and 35% for PC were in the 25-34 age range. Additionally, 19% for mobile and 24% for PC were in the 35-44 age range.

Gender. Our participants skewed slightly towards males. For the two device environments, 54–55% identified as male.

Educational Background. Most of our participants had at least some college-level education. For mobile, 62% had a Bachelor's degree, while 47% had likewise for PC.

4.1.5 *Limitations.*

Our survey entails several key limitations, many inherent to survey-style investigations.

- We surveyed only US adults, so our results may not generalize globally. Future work can explore broader populations.
- Survey responses rely on participant recall, and may not accurately reflect real-world user behavior. To combat this, we requested participants reflect on logins over recent weeks.
- Participants may respond in a fashion they believe will be best perceived, rather than reflecting their true views (social desirability bias). We address this bias through common techniques such as survey anonymity, flexibility in responding to questions, and framing questions as neutral and general (rather than questions implying masking's costs or benefits).
- In our analysis, we provide the number of study subjects who gave a particular response to our questions. However, we caution that such counts are not necessarily reliable indicators of real-world prevalence. Future larger-scale studies of masking can replicate or expand upon our work.

4.2 Results

Here, we analyzed our survey responses, divided between the mobile (labeled as **M**) and PC (labeled as **P**) environments, seeking to understand user experiences, perceptions, and preferences with password masking. In most cases, we observed similar responses for mobile and PC; we particularly highlight cases where mobile and PC responses notably differ.

4.2.1 *Past Experiences.*

We first investigate our participants' comprehension of and experience with password masking.

Comprehension. We asked our participants an open-ended question about why websites use password masking, to assess whether they understood its core purpose. The majority of our survey participants demonstrated comprehension of masking's contributions. In total, 144 participants (71%), 75 for mobile and 69 for PC, explicitly discussed masking as providing some form of shoulder surfing protection (although not necessarily using such terminology). For example, PC participant P88 said that password masking is used "To prevent people from looking over your shoulder and seeing your password," while mobile participant M28 described masking's use as "To hide from screen peepers." Similarly, 131 participants (65%), 78 for mobile and 53 for PC, described masking as making users feel more secure. For example, M69 stated that "Website uses password masking to give a sense of security. Others can't see what you typed in for your password."

We did observe 44 participants (22%) who expressed some misconceptions about the protections offered by masking, with three primary misunderstandings.

1. *Website Hacking Protection:* 26 participants (13%) believed that masking helped protect against website hackers. P36 said that websites use masking "because of hackers and security breaches."
2. *Spyware Protection:* 14 participants (7%) believed masking protects password entry against spyware. For example, P97 believes that "If there is a keystroke logging spyware on your machine this [masking] might eliminate that use."
3. *Secure Password Transmission and Storage:* 3 participants indicated that masking protects the password during transmission or storage, such as P84 saying that masking "keeps the data they collect secure while transmitting."

While password masking visually hides the password during entry, the cleartext password is ultimately still accessible to a browser (or malware, such as a keylogger or malicious browser extension). Password transmission and storage also do not involve masking, and instead rely upon other mechanisms (e.g., TLS, password salting, and hashing).

Beyond misconceptions, 14 participants (7%) expressed uncertainty about masking's purpose. M87 said that they were "Unsure and I've always wanted to know why." Meanwhile, M70 said "I don't really understand because I'm not tech savvy." Thus, despite being widely used, masking is not fully understood by all users, indicating a potential lack of awareness.

Frequency of Masking Encounters. We asked our participants to recall how frequently they encountered password masking during logins in the prior few weeks, using a 5-point Likert scale (1=Never, 5=Always). In total, 99 participants (49%), 45 for mobile, and 44 for PC, indicated that they always encountered masking when logging in, while an additional 80 participants (40%) encountered masking the majority of logins. Thus broadly, participants frequently experienced masking, although a small minority (approximately 10%) reported logging in without masking for less than half of logins, an observation that runs counter to the popularity of masking on top websites (from Section 3).

Users frequently encounter masking and broadly understand its purpose. However, some users do not fully understand masking's uses or exhibit misconceptions about the limits of masking's security protections.

4.2.2 Usability Considerations. Next, we explore the usability considerations with masking.

Usability Impact. We asked our participants to assess how often they had issues or challenges when entering passwords with masking (over the prior few weeks), using a 5-point Likert scale (1=Never, 5 = Always). Overall, we observed that most participants reported occasional issues, although mobile participants reported a higher rate of challenges than PC participants. For PC, 28 participants said that they never had issues, while 58 indicated they sometimes had issues but less than half of the time. Only 10 participants indicated having issues half the time or more. Meanwhile, for mobile, 17 participants reported never having issues, 57 recalled occasionally having problems, and 23 indicated issues happened in roughly half of their password entries. Thus, for mobile participants, fewer recalled avoiding masking issues altogether and the remaining participants reported more issues.

We also asked participants to compare the difficulty of password entry with and without masking, again using a 5-point Likert scale (1 = Much Easier, 5 = Much Harder). We found that participants either found the two conditions equivalent or masking slightly harder. In total, 86 participants (43%) indicated that masking did not notably impact password entry difficulty. Meanwhile, 72 participants (36%) indicated masking made password entry somewhat more difficult (4 out of 5 on the Likert scale). Mobile and PC responses were similar although mobile participants again leaned towards more difficulties, with 12 mobile participants saying password entry was much harder with masking than without, compared to 5 for PC respondents. Surprisingly, 27 participants (13%) responded that masking made password entry easier (14 for PC, 13 for mobile). We lack further insight into this situation, as we did not ask an open-ended question for participants to further explain their rating (in the interest of limiting our survey length, and as we expected many responses to mirror our existing question about experienced issues). We hypothesize though that either some participants responded with the comparison direction reversed (comparing no masking with masking), or that they largely felt neutral about masking's usability impact but positive about its other benefits, thus rating masking as overall making password entry easier or better.

Usability Issues. To delve into usability issues introduced by masking, we asked our participants an open-ended question to describe issues or challenges they had encountered with masking.

Matching with existing expectations that masking obscures typos [20, 21, 27], we did observe that the primary issues discussed related to password entry typos, mentioned by 184 of our respondents. Furthermore, 34 of our participants discussed how such typos forced login retries. For example, P1 said "Sometimes I'll type too fast and miss a key. Because it [the password] is hidden, I don't catch it and will have to retype my entire login info all over." Typos also included cases where caps lock is enabled, but users cannot see the resulting erroneous password, as mentioned by 3 respondents. P60 stated "I cannot tell if I have capslock on, or what the issue is when I cannot successfully log in." Several users (10) indicated that such typos were frustrating. P34 exemplified this feeling, saying "I get annoyed when I type it in wrong. Most of the time I have to click the little eye icon anyway to make sure it's correct."

As with our usability impact analysis above, we observed 13 participants discussing how masking impacted mobile more than PC. For example, M38 explained "typing on a touchscreen keyboard isn't as accurate as typing on a physical keyboard for me."

Interestingly, 13 participants also commented on how masking interacted with password managers. In one direction, users indicated that by using a password manager to auto-fill passwords, they avoided usability issues from masking. For example, M25 stated "I don't really have any problems, it auto-fills passwords and they are masked." In the other direction, participants described masking as negatively interacting with password managers. M19 described how masking has prevented them from determining that "Just an old password being stored [by a password manager]". Meanwhile, M38 said that masking "caused issues when I'm saving a password to my password manager; I'll go back and see that the password was saved as •••••• instead of the actual characters." We dig deeper into password managers and masking next.

Masking and Password Managers. Given the potential interactions between password managers and masking, we further investigate whether password managers reduce masking issues.

To start, we asked our participants a 5-point Likert scale question about how often they used password auto-fill features, such as those provided by many password managers, when logging in (1 = Never, 5 = Always). For both PC (42 participants, 42%) and mobile (42 participants, 41%), the most common response was most of the time (4 out of 5). The remaining participants were roughly evenly distributed across the different frequencies. Thus, password auto-filling was common amongst our participants, although there was also wide diversity in its use.

We then computed the Pearson correlation coefficients between the frequency of password auto-fill with the frequency of password masking issues, as well as with the ease of login with masking compared to no masking (all 5-point Likert scale questions). We did not observe notable correlations in any cases. The Pearson coefficient with the largest absolute value was $r = -0.16$, between auto-fill frequency and frequency of password masking issues for mobile, which indicates a low degree of correlation. All other correlation coefficients were even closer to 0. While this outcome may seem surprising, we note that only a small number of participants mentioned how masking interacted with a password manager. Furthermore, our results suggest that masking issues were occasional and often in situations without auto-fill. Thus, we hypothesize that while auto-filling may help users avoid typos when entering their passwords, it can introduce different issues (as discussed earlier) and users still often need to enter passwords without auto-fill and deal with masking issues.

Masking and Password Creation. Prior expectations of masking were that it made password entry more difficult [20, 21, 27], which our survey results confirm (although the extent of usability issues may be less than some would have believed). Given this expectation, we also hypothesized that some users may change the passwords that they use to reduce masking-induced issues. Thus, we asked participants if masking influenced their password choice. In total, 164 participants (81%), 83 for PC and 81 for mobile, indicated that masking *did not* influence password choice.

We asked participants to explain their answers. The majority of respondents (97) who said that masking did not influence their password choice explained that they adhered to the same password selection process regardless of masking. They discussed prioritizing secure passwords, ease of memorization, adherence to a consistent pattern, or in some cases simply reusing the same password. As exemplified by P16, "Because password strength and security is far more important than the ease of entering said password with password masking." Similarly, P74 noted, "I use the same password schema regardless. It's so I can remember, and that doesn't change if it's masked or not." Another interesting explanation, discussed by 18 participants, is that the user purely relies on password managers for creating passwords, and thus the password chosen is not affected by masking's presence. For example, P70 said "All of my passwords are auto-generated by my password manager."

Among those who responded that masking does influence their password choice, we saw 24 participants who described choosing easier-to-type passwords due to masking. For example, M81 explained "If I can't see what I'm typing, I make the password simpler." Similarly, P18 said "I would choose passwords that I can easily remember over auto-generated passwords with a lot of random letters and symbols that are hard to remember and easy to mess up." In another direction, some respondents indicated that masking gave them more freedom in password selection. P13 said "I can make up a strange password and not be judged by it by anyone around me." Meanwhile, P15 said "if others saw it all the time I would make it [the password] something cooler and less embarrassing." We further saw some respondents indicate that masking made them pick more complex or longer passwords. M64 said that masking "increases my awareness about my password length." Meanwhile, M48 explained "Password masking makes me more confident in the security of a site, but also influences me to make a more complex password. I guess the stars make me a little nervous."

> Most users encounter usability issues with password masking only occasionally, finding password entry with masking similar or only slightly harder than without masking. Mobile users do exhibit more frequent issues than PC users though.
>
> The central masking usability concerns surround typos during password entry, although we observed some initial evidence that password managers affect masking experiences both negatively and positively. We observed also that for a minority of users, masking influences password selection.

4.2.3 Security and Privacy Considerations. We further explored user experiences and perspectives on masking's security and privacy impact.

Security-Sensitive Situations. We first asked our participants a 5-point Likert scale question about how often during logins (in the prior few weeks) they found themselves in a setting where someone else might observe their password entry (1 = Never, 5 = Always). Most participants (123, 61%), for both PC (61) and mobile (62), indicated that such situations arose only occasionally (2 out of 5). Meanwhile, 48 participants (24%), 28 for PC and 20 for mobile, never recalled being in such settings. The remaining 31 participants indicated that they were often in a potential shoulder-surfing situation, for at least half of their logins. Thus, we see that shoulder surfing situations do arise in practice, but often infrequently (we note mobile respondents indicate slightly higher prevalence).

We asked an open-ended question for participants to discuss example situations. Three common situations were:

(1) **Public Setting.** The most popular situation was in a public setting, discussed by 49 participants for mobile and 35 for PC. P1 described working in public, saying "When I am doing my school work at a cafe or library, there are people walking by all the time and could peep at my screen and see my password, especially if it isn't masked." Similarly, M12 described accessing accounts on public transportation, stating "Perhaps I'm in a bus or a train and I need to quickly access my bank account." Three mobile participants also explicitly mentioned public cameras, such as M46 explaining "I also know that on rare occasions, my screen can be seen and read by nearby security cameras."

(2) **Logins near Family or Friends.** The next most common setting, described by 25 participants for mobile and 31 for PC, was in an intimate setting where a user authenticates in the presence of a family member or friend. For example, P76 said "Sometimes my wife might be behind me at home." Meanwhile, M69 stated "When I'm sending money to a friend for lunch in front of them. I have to log into my bank so they can see me type in my password."

(3) **Work Environment.** The third popular environment was during work, mentioned by 21 participants for both mobile and PC. M36 gave an example, "When I'm at work, it's a fairly big office and sometimes there are people who are right behind me that I haven't yet heard. I am often surprised by someone standing right behind me." P25 illustrated similarly, "I work in a field that requires me to use my password on a work computer and there is always somebody around. I enter my password an estimated 40 times a day."

Two other interesting situations discussed include educational settings (12 participants in total) and during screen sharing, such as during online video calls or when presenting one's screen (13 participants). P77 gave an example of the education setting, saying "In class when I am sitting right next to someone." Meanwhile, P102 described the screen sharing situation, saying "If I am logging into something while showing a coworker something on my screen." Thus, there are a variety of real-world situations where shoulder surfing is a realistic risk, although participants ultimately reported such situations being occasional.

Perceived Security Benefits. We also asked our participants if they felt more secure/private with masking compared to without. Our participants overall indicated that they felt neutral (26 for PC, 20 for mobile) or more secure/private with masking (67 each for both PC and mobile). Thus, only a small minority of participants (10% in total for both environments) indicated that they did not feel more secure with masking.

We also asked our participants to choose categories of sites that they felt masking would be most important on, including an open-ended "Others" option (although we observed very limited use of this option). Participants could choose to select multiple categories or no categories. These site categories included social media, financial platforms, email services, shopping sites, healthcare platforms, work accounts, entertainment services, and educational platforms. Here, we sought to understand the type of sites that users prioritized masking on. We found that financial platforms were the most popular choice, with 97 participants for mobile and 94 for PC. Following closely behind were email services (76 participants for mobile, 74 for PC), and work accounts (72 participants for mobile, 77 for PC). Meanwhile, approximately 60% of participants for both PC and mobile selected social media, health, and shopping sites. On the other extreme, only about a quarter of participants chose educational and entertainment sites.

Users typically find themselves at risk of shoulder surfing only occasionally but described a variety of such settings, including in public, at work, and with family/friends.

Users overall feel more secure and private with masking, and particularly valued masking on financial websites (as well as email and work accounts, to a lesser extent).

4.2.4 Masking Method Preferences. In our final set of survey questions, we asked our participants about their masking design preferences. We first asked about whether participants preferred masking on or off by default, and if they wanted a masking toggle option. Specifically, we asked participants their preferences between no masking, masking on by default with or without a toggle option, and masking off by default with an option to toggle on. The first three options we offered reflected design options seen in practice during our real-world measurements (from Section 3), but we offered the last choice as it reflected a final option in the design space when considering masking and toggling (note that no masking is equivalent to masking off by default without a toggle option).

A strong majority of 167 (82%) participants, for both PC (82) and mobile (85), preferred masking on by default with an option to toggle off. Meanwhile, only 16 participants in total selected masking on by default without a toggle option, 13 wanted masking off by default with an option to toggle on, and only 5 chose no masking.

We also asked our participants that if masking was enabled, whether they preferred every character consistently being masked when typed (static masking) versus the last character typed being unmasked temporarily before becoming masked (dynamic masking). Here, we observed that dynamic masking was more commonly preferred, but more so for mobile than for PC. For PC, 48 participants wanted static masking while 51 preferred dynamic; for mobile, 33 selected static compared to 69 for dynamic.

We asked an open-ended question to our participants to elaborate on their design preferences. We observed widespread discussion of perceived safety, by 131 participants (64%), as a major justification for having masking on by default. For example, P11 said "I feel more comfortable with them masked in case someone is watching me type these passwords in." Another interesting and related reason was to avoid socially awkward situations. P68 exemplified this by saying "It [masking] does remove the awkwardness of them feeling like they have to turn their head or if you do not trust the person, asking them to do so." Respondents also wanted a toggling option as visual aid during password entry (132 participants, 65%) and user control (73 participants, 36%). P12 commented on how toggling helped with masking, "That way I can quickly check to make sure the password is correct." Several participants explicitly mentioned a fear of getting locked out due to mistyping the password multiple times. M38 opined that "Having no ability to view what you're typing whatsoever is incredibly inconvenient and had resulted in me being locked out of my account because I didn't know caps lock was on or something similar."

For participants opting for other design options, we observed various explanations. Some indicated they log in when isolated so masking (at least by default) was not necessary. For example, P73 preferred masking off by default with an option to toggle on (an option we provided that we did not observe used in practice), saying "I live alone, I work from home, I don't need masking most of the time. I like having the option to turn it on if needed (if I'm sharing a screen for example)." Others felt that masking does not provide any real security benefits, including M3: "It's more of a security theater than anything else. If somebody's paying close attention they can still see your password." Finally, a few participants commented about how masking is inconvenient. M84 said "Masking's inconvenient and not helpful for me at all so I would prefer not to have it, but I'm never given that option."

Users broadly prefer masking enabled by default, with an option to toggle it off. Between static versus dynamic masking, PC users are roughly split while over two-thirds of mobile users prefer dynamic masking.

5 User Experiment

Complementing our survey study, we also conducted a user experiment to empirically evaluate users authenticating with and without password masking. Our survey provides user perspectives and self-reported experiences regarding masking while our user experiment can provide quantitative data on real-world behavior. Our user experiment focused on how masking impacts authentication usability metrics (we discuss directions for evaluating real-world security in Section 6).

5.1 Method

We first describe our user experiment method, which simulates the login workflow on a website. Our experiment study was approved by our organizations' IRBs.

5.1.1 Experiment Design. Our experiment entailed users entering passwords under different masking conditions, while we monitored for password entry usability issues (e.g., time to complete, errors). We evaluated the same five masking variations as investigated in our survey (from Section 4.2.4) and as seen in practice (from Section 3), for both mobile and PC environments. Specifically, we evaluated no masking, static masking with and without a toggle option, and dynamic masking with and without a toggle option. As a result, our experiment involved 10 workflows that were identical except in the masking condition tested. Each participant was randomly assigned to one workflow.

Experiment Steps. We designed our experiment for online participants, running a publicly accessible website under our research team's management. Thus, participants would arrive at our website and proceed through the following steps.

(1) *Explanation and Consent.* The participant first encountered an experiment explanation page that gave a brief summary of our experiment steps, followed by a consent page. We explained that our study was investigating online authentication behavior and that participants would create an account on a toy financial website (Cosmos Transfer, modeled after online payment services) and later log in. We requested that the participant treat this account as if it were a real financial one, and we also explicitly explained that we would not collect or save their account credentials, to encourage them to select realistic passwords.
(2) *Account Creation.* After consenting, the participant was directed to an account creation page for Cosmos Transfer and prompted to create an account. We displayed our simple password policy: the password length had to be 8 characters or longer. This policy aligns with modern recommendations [14, 30], which excludes character composition requirements. The account creation form also presented a zxcvbn-base password strength meter [32] to nudge (but not enforce) users towards stronger passwords.
(3) *Distractor Task.* After creating an account, the participant was directed to a page indicating that Cosmos Transfer required verification that the user was human. The participant was prompted to complete a 9-piece jigsaw puzzle, which required putting together puzzle pieces to form a complete picture. This page served as a distractor task, to logically and temporally separate the account creation task from the subsequent login task (particularly so participants did not log in with a password immediately after selecting it). Prior work has effectively applied such distractor tasks during similar authentication studies [18].
(4) *Account Login.* After puzzle completion, the participant was directed to an account login page and prompted to authenticate. Here, the password entry field employed the specific masking condition being evaluated. The participant progressed either upon successful login or upon five failed login attempts.
(5) *Completion.* At the completion page, we thanked the participant and provided instructions for receiving the study compensation.

Experiment Telemetry. During the experiment, we collected telemetry during both the account creation and login pages. For each participant, we tracked a unique user ID (which does not contain personal information) to map the telemetry to the participant.

On the account creation page, once the participant selected a password, we temporarily stored the password purely client-side within the browser's session storage (which is automatically cleared when the page session ends, such as when the browser or tab is closed). The participant-chosen password was never transmitted. We recorded the password length and the password strength as measured by zxcvbn [32] (on a scale of 0 to 4). This data allows us to assess whether participants chose overly weak passwords, and whether the population of chosen passwords differed significantly across the workflows. We note that while individual users will vary in the length and strength of their selected passwords, by randomly assigning users to different masking conditions, the distribution of password characteristics should be commensurate across workflows/conditions.

On the account login page, we recorded the following password entry metrics for each login attempt:

- *Password entry duration:* measured from when the first password character was typed to when the last character was typed.
- *Number of characters entered and deleted:* measured as the total number of characters entered into the password field, as well as the total number of characters that were deleted (including if the participant deleted all characters entered). The difference between these two values is the final password length submitted.
- *Masking toggle count:* measured as the number of times the masking toggle was used (also revealing the final toggle state).
- *Login success or failure:* determined based on whether the submitted password matched the correct password (stored in the browser's session storage).

If a login attempt failed, we compared the submitted password with the correct one (stored in the browser's session storage), computing three string distance functions:

- Levenshtein, which computes distance when allowing for insertions, deletions, or substitutions.
- Damerau–Levenshtein, a variant of Levenshtein distance that further permits transpositions.
- Hamming, which computes the number of character positions that differ. Note that this distance is only applicable for two strings of identical length.

We also assessed whether the submitted password exhibited one of five common classes of password typos, as identified by prior work [7]. These typos included adding an extraneous character at the password start or end, flipping the case of the password's first letter, flipping the case of all password letters, or missing shift for symbol characters at the password end.

Implementation Details. We made several key implementation decisions to enhance the validity of the study.

- On the account creation and login pages, we disallowed copy-pasting into the password field, so that participants would directly enter the password.
- On the account creation and login pages, we implemented the password field as HTML input elements of *type=text*, rather than as *type=password*. The masking (both static and dynamic) was implemented instead in JavaScript. This decision provided two key benefits:

  1) We could control the specific masking type used, as different

browsers implement a default masking form for password input fields (as found in Section 3).
2) We could mitigate password manager interactions (such as saving and auto-filling in the password), as the password field was not readily identifiable.

- Using JavaScript, we fingerprinted the device and browser for two reasons:
1) We could confirm whether a mobile or PC participant was indeed on an appropriate device.
2) We identified that certain UI elements appeared differently on various devices, and thus could customize the UI appropriately.

5.1.2 Experiment Deployment. We recruited experiment participants via Prolific [23]. We had 10 Prolific tasks, one for each experiment workflow (which differed in the masking condition tested). Prolific participants were directed to our experiment website, and upon reaching the experiment completion page, were redirected back to Prolific to receive compensation. The experiment workflow took on average 3.5 minutes to complete, with the distractor puzzle requiring the majority of the time. We compensated participants $0.88, a rate of $15/hour (the same rate as our survey compensation).

In total, we recruited 50 participants per workflow, as well as 10 participants per workflow during pilot deployment, to test and debug our experiment website. As our pilot deployment did not reveal issues that impacted the validity of the collected data, we included the collected data, resulting in 60 participants per workflow, for a total of 600 experiment participants.

5.1.3 Limitations. Our experiment, while simulating a real-world login workflow, is ultimately an artificial setup. Participants may not have behaved as they would during password authentication on a real website, especially being aware that they are part of a study. We attempted to mitigate these issues by adopting a website design mimicking a real-world financial service, encouraging participants to treat the created account like a real financial account, ensuring study subjects that their account credentials would not be collected, and discussing that we were generally studying authentication without emphasizing masking. Furthermore, as we will discuss in Section 5.2.1, our analysis of the password characteristics from participants suggests that they picked meaningful passwords, providing support for the external validity of our results.

Our experiment is also at a moderate scale, but large enough to identify workflow differences with notable effect sizes. However, a larger-scale experiment may be able to identify further differences with smaller effect sizes, which can still have practical impact (e.g., for popular websites).

5.2 Results

Here, we analyzed the password entry usability metrics across masking conditions, for both PC and mobile environments.

5.2.1 Data Quality Analysis. We first assess the quality of our collected telemetry and identify data requiring filtering.

Password Characteristics. We analyzed the distribution of zxcvbn-based password strengths [32] for the participant-selected passwords. We observed that our participants chose reasonably strong passwords: 24% chose passwords of the highest strength level 4, 32% for level 3, 24% for level 2, and 18% for levels 1 or 0. For comparison, we also analyzed the strengths of passwords in the `000webhost` and `RockYou` password dumps [17]. For `000webhost`, 17% of passwords were level 4, 25% level 3, 32% level 2, and 27% levels 1 and 0; for `RockYou`, 7% were level 4, 21% level 3, 36% level 2, and 36% levels 1 and 0. Thus, the distribution of password strengths for our participant-chosen passwords skewed to stronger passwords than in leaks, suggesting realistically strong passwords.

Furthermore, we analyzed the length distribution of participant-chosen passwords (recall that our policy required a minimum password length of 8). We observed that participants were not primarily choosing as short of a password as possible, with only 21% of participants chose length 8 passwords. Meanwhile, we also observed only 3% of passwords exceeding length 20, indicating that it is unlikely participants used randomly generated passwords (e.g., as generated by password managers).

Overall, we did not observe password characteristics that would suggest unrealistic or weak password selection. When broken down by individual workflows, we also observed similar password strength and length distributions, indicating that the population of passwords chosen by participants in different workflows were characteristically similar.

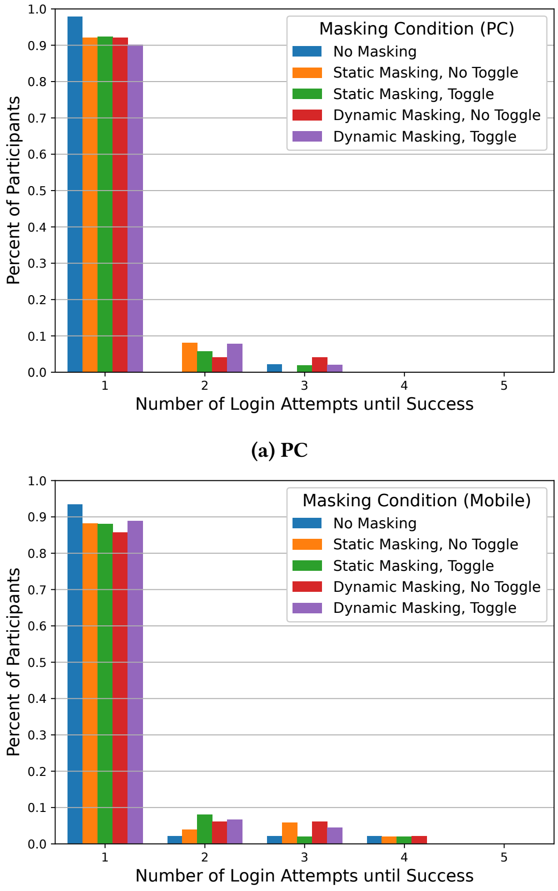
Failed Logins. Across the full experiment, we observed 15 participants (2.5%) fail to log in within 5 attempts (including 8 participants in workflows without masking). While we lacked visibility into the participants' real and submitted passwords, we observed that the edit distances between the real and submitted passwords for these participants' login attempts were high (greater than 2). We believe that these participants likely did not accurately remember their originally created passwords, or were not making realistic attempts to log in. This belief is reinforced by the scarcity of successful logins requiring more than three attempts. Across the entire experiment, we observed no participant successfully log in only on the fifth attempt, and 4 participants successfully log in on the fourth attempt. Thus, we filtered this small number of participants who failed to successfully log in from our subsequent analysis, as their behaviors likely were unrealistic or reflected failed password recall rather than masking influences.

5.2.2 Usability Impact of Masking. Here, we characterize the login performance of participants under varying masking conditions.

Login Attempts. We analyzed the number of login attempts that participants required before successfully logging in. Figure 2 depicts the distribution of login attempts across masking conditions, for both PC and mobile.

We saw that across all conditions and both device environments, the vast majority of participants successfully logged in on their first attempt. We also observed that password entry for mobile does appear more challenging than for PC, as more participants required multiple attempts to succeed across all masking conditions (with no PC participants requiring more than three attempts, while 4 mobile participants required four attempts).

Comparing the different masking conditions, we observed that for both PC and mobile, participants were slightly more successfully logging in without masking compared to any of the masking conditions. Meanwhile, participants experiencing different masking variants performed similarly. For PC, 97% of *No Masking* participants logged in on their first attempt, compared to 90–92% for



(a) PC

(b) Mobile

Figure 2: Number of login attempts until success, for different masking conditions on both PC and mobile environments.

participants experiencing some form of masking. (We note that prior work evaluating Amazon MTurk workers typing cleartext passwords observed a password typo rate of about 3% [7], commensurate with our findings.) Similarly, for mobile, 93% of *No Masking* participants succeeded immediately compared to 85–88% for workflows using masking. These results align with our survey results (from Section 4.2.2), which indicated that masking usability issues were occasional and that password entry with masking was similar or only slightly more difficult than without masking.

We evaluated whether the differences observed between the masking conditions were statistically significant, for each device environment. We used the Mann-Whitney U test to compare masking conditions pairwise, as this test is non-parametric and our sample distributions are not normally distributed. As we were making multiple statistical comparisons, we applied the Holm-Bonferroni method for multi-test correction. Across all pairwise comparisons for each device environment, we did not find any statistically significant differences. This result does not mean that there are no differences between the masking conditions, but rather the effect size (if existing) is too small to detect with our sample size. Nonetheless, this statistical result reinforced our conclusion that the impact of masking on login attempts is limited.

Password Entry Timing. Next, we investigated whether masking impacts the time it takes to enter passwords. We note that password timing is inherently related to password characteristics such as length. However, as discussed in Section 5.2.1, the passwords chosen by participants in different experiment workflows shared similar length and strength distributions. Thus, we could compare aggregate statistics between workflow participants.

We evaluated the password entry timing for the first login attempt for all participants. We excluded subsequent login attempts as users may adjust their typing behavior due to the prior failed logins, rather than reflecting the influence of masking. We computed the timing medians for each condition's participants, instead of averages, as the timing distributions are not normally distributed, and the average is more heavily influenced by outliers. Furthermore, we observed that the timings varied widely across participants.

For PC, the median password entry time ranged from 3.54 to 4.86 seconds (with *No Masking* participants exhibiting a median of 4.04 seconds). For mobile, the median password entry time ranged from 4.53 to 5.6 seconds (the median for *No Masking* participants was 4.64 seconds).

We again saw that password entry is typically slower on mobile than on PC. However, we did not observe a consistent pattern when comparing no masking with masking, nor toggling versus without (even when considering the 25th and 75th timing percentiles, in addition to the median). Computing the Mann-Whitney U test to pairwise compare masking conditions for each device environment, and applying Holm-Bonferroni multi-test corrections, we again did not observe statistically significant differences between the timing distributions. Thus, we conclude that password masking does not significantly influence password entry timing, and the timing variations are dominated by the diversity in typing behaviors.

Toggle Use. We analyzed how often participants used the masking toggle when available.

We first consider the use of the toggle during the initial login attempt. We observed that for PC, 15% of *Static Masking* and 14% of *Dynamic Masking* participants used the toggle button on their first login. For mobile, 15% and 18% of *Static Masking* and *Dynamic Masking* participants toggled masking, respectively.

We also considered the participants who required multiple login attempts but had not used the toggle option during their first login (but could have). We find that 20% of such participants used the toggle option in their subsequent login attempts.

Thus, toggling is indeed used by a non-trivial minority of participants, although it is not used in the majority of login attempts.

Character Deletion. We analyzed how often participants delete characters during password entry. On the one hand, one might expect more deletions without masking, as a user can observe and correct typos. On the other hand, users with masking may delete part of an entered password (possibly restarting) due to uncertainty about the correctness of the typed password.

We evaluated the proportion of participants who deleted characters on their first login attempt. Across the masking conditions for PC, 12–25% of the participants deleted characters (19% for *No*

Masking users). For masking conditions on mobile, 20–27% of participants deleted characters (22% for *No Masking*). Thus, we observed a slightly higher prevalence of deletions for mobile than for PC. However, we did not find a consistent pattern across masking and toggling conditions that suggested they significantly impact character deletions (nor do we find significant differences using the same statistical evaluations as done earlier). We hypothesize that both effects discussed above may be in play to similar degrees, where no masking permits some careful users to correct typos, while masking may trigger some users to retype parts of their password.

Typos. Finally, we provided a brief analysis of the typos made by our participants during failed login attempts. Overall, as most participant successfully logged in on their initial attempt, we have limited data on failed attempts. We could not automatically classify the typo category for 47% of failed login attempts. The most prominent detected typo (35% of failed login attempts) was the mis-capitalization of the first password character, followed by an extra character at the password end (8%). We did observe at least one participant making each of the other detected typo categories though (Section 5.1). Given the low number of participants per workflow that required multiple login attempts, we lacked sufficient scale to meaningfully compare across masking conditions. However, we note that the typos were distributed broadly across workflows, including one *No Masking* participant mis-capitalizing the first character and another adding an extra character at the password end (highlighting that even without masking, users must pay sufficient attention to the entered characters to detect errors). Overall, we lack evidence that masking impacts the type of typos made during password entry.

> Masking has some, but ultimately limited, impact on login success. It does not seem to notably impact password entry timing, character deletions, or the type of typos made.
>
> A substantial minority of users do utilize toggling.

6 Concluding Discussion

In this paper, we conducted the first systematic investigation of password masking, analyzing its practical impact through multiple methods. We combined insights from a measurement of password masking by real-world browsers and sites, a user survey of over 200 users, and an experiment with 600 participants. Our study revealed user understanding of masking, its usability impacts (both as reported by users and through experimental data), security considerations, and user preferences for masking designs. Here we synthesize the lessons learned and directions for future work.

Future Investigation into Password Shoulder Surfing. We first recognize one key limitation of our work, which points to a direction for future research. While we assess masking's usability impact both through user self-reporting and observing real-world user behavior, we only evaluate masking's security considerations from user self-reporting. We note though that our user experiment results align closely with our survey findings, providing some confidence that the self-reported experiences are reliable.

Ultimately, monitoring real-world user behavior for password shoulder surfing presents a considerable challenge. Behavior can change under observation, and the opportunistic nature of shoulder surfing makes it difficult to control in a real-world setting. Moreover, it may happen in restricted or private environments, such as workplaces, schools, or homes. While some prior work on general shoulder surfing has similarly relied on user reporting [10], some other works have tried simulating potential shoulder surfing scenarios in a controlled environment (e.g., through using virtual reality [26]). Future work on shoulder surfing of authentication (or other security-relevant) information could adopt such approaches.

We emphasize, however, that to fully understand the security benefits of password masking, we should not only evaluate instances when shoulder surfers observe a user's password but also assess the security or privacy consequences of such information leakage. In a worst-case scenario, every instance of password leakage could be considered a security breach. However, future work should explore the consequences of password observation for a more realistic analysis.

To Mask or Not Mask? Next, we tackle the most central question about masking. From our empirical evaluation of browsers and websites (Section 3), the answer for the vast majority of browsers and sites is already yes. Yet to our knowledge, there exists no public data or documentation justifying these decisions on masking. Thus, it is not certain what drives this decision. The goal of our study was to provide empirical grounding on user experiences and behaviors with masking.

Based on our study's findings, *we recommend that browsers and websites support password masking.* The usability costs seem largely balanced against the security benefits. Both our survey and user experiment indicate that usability issues are only occasional (Sections 4.2.2 and 5.2). Meanwhile, our survey found that masking's security benefits are also only occasional (Section 4.2.3). We note that prior work on general shoulder surfing [10, 26] has also found shoulder surfing occurs infrequently and typically opportunistically, aligning with our results. However, we found that 97% of our participants preferred some form of masking over no masking (Section 4.2.4). Furthermore, we observed a number of real-world scenarios where users do require masking (Section 4.2.3), including in intimate partner situations. Finally, we note that password autofill can help avoid masking's usability concerns, and was widely used by users (Section 4.2.2). While the usability impact is still non-trivial, we ultimately argue that, as many browsers and sites already provide, some form of masking should be available to users.

Recommended Masking Format. In Section 4.2.4, we found that the vast majority (over 80%) of participants preferred masking on by default with an option to toggle off. We also observed a slight preference for dynamic masking over static masking, particularly by mobile users. From our experiments in Section 5.2, we found minimal performance differences between static and dynamic masking, as well as with or without toggling. However, we do observe that about 15% of users do toggle masking when available, and an even higher rate (20%) do so after failing an initial login attempt. Thus, given user preferences and behaviors, *we recommend dynamic masking on by default, with a toggle option available.*

Despite our recommendation, we do believe another design is interesting and warrants exploration. Conceptually, masking off by default with a toggle option could have reduced usability impact (as many users indicated that they often log in alone), while still supporting masking for security-sensitive situations. However, having the less secure state as the default might lead to user errors, such as forgetting to enable masking when needed. Furthermore, we did observe some users also commenting that explicitly enabling masking could be a negatively perceived action, such as signaling a lack of trust to the nearby person. Despite users not preferring this design (Section 4.2.4), evaluating how it may be used by users in practice may be illuminating.

User Education. Password masking is arguably one of the simplest security/privacy mechanisms in use, especially when compared with other security/privacy apparatuses visible to users, such as TLS (particularly through HTTPS), secure messaging (including email encryption), application permissions, and even passwords themselves (along with more recent authentication developments, such as passwordless authentication). Indeed, in Section 4.2.1, we did find that the majority of users understood masking's purpose. Yet over a quarter of our participants either misunderstood some of the masking's benefits or lacked a clear understanding.

Our finding indicates that many users could benefit from further education about password masking. Beyond this, masking is an intriguing case study on user security education. Given that such a relatively simple and widely adopted security technique is still misunderstood by a large portion of users, how can we educate users so that nearly all users understand the technique properly? If we cannot achieve that goal with masking, is there hope for educating users about even more complex security topics? There may be a unique opportunity for future work to perform a controlled study on user security education, focusing on masking specifically. Anecdotally, we found that online resources for masking are largely oriented towards web developers. Thus one effort that could be explored is providing resources geared towards end users.

Exploring the Adoption of Security Mechanisms. Our finding in Section 3 that password masking is ubiquitous across popular browsers and websites should be unsurprising to most web users. However, masking is arguably controversial and to our knowledge, not discussed by most authentication guidelines. While NIST's latest authentication guideline does briefly mention masking [14], its older versions did not [4, 5], nor do other common guidelines (e.g., OWASP [22], UK's NCSC [19]). This begs the question, how did this security mechanism become so commonplace (especially compared to the slower adoption of other security techniques, such as HTTPS [12] or Content Security Policy [31])? Future work could investigate why and how password masking was adopted, which may offer a ripe opportunity to study the factors driving the successful adoption of security techniques.

External Impacts of Masking. Our study found that in some cases, password masking influenced password manager use as well as user password selection (Section 4.2.2), demonstrating that even simple security mechanisms can impact external dependencies.

Our observations can guide password managers in better accommodating masking, given its prevalence. Password managers should ensure that passwords are saved correctly when users create accounts or change passwords. If users manually save passwords by copy-pasting from masked fields, there is a risk that only the masked symbols (e.g., dots) are copied instead of the actual password. To prevent this, password managers could either block copy-pasting or include a reminder to unmask the password before copying.

Future work can also dig deeper into password masking's influence on password creation. We found that a non-trivial minority (about 20%) of users discussed such an influence. Further work can be done to characterize the impact, which we observe going in both directions. In some cases, users discussed masking as leading to weaker passwords, while in other cases, masking afforded more personalized (and potentially stronger) passwords. If masking has a significant impact on the security of passwords themselves (such as has been identified for password resets [33]), that may change masking's cost-benefit analysis.

7 Acknowledgements

We thank the anonymous reviewers for their constructive feedback. The second author was supported by the Kuwait University Scholarship. This work was also supported in part by the National Science Foundation award CNS-2055549. The opinions expressed in this paper do not necessarily reflect those of the research sponsors.

References

[1] 2024. Bruce Schneier. https://en.wikipedia.org/wiki/Bruce_Schneier.

[2] 2024. Jakob Nielsen. https://en.wikipedia.org/wiki/Jakob_Nielsen_(usability_consultant).

[3] Florian Alt, Mateusz Mikusz, Stefan Schneegass, and Andreas Bulling. 2016. Memorability of cued-recall graphical passwords with saliency masks. In *International Conference on Mobile and Ubiquitous Multimedia*.

[4] William E. Burr, Donna F. Dodson, Elaine M. Newton, Ray A. Perlner, W. Timothy Polk, Sarbari Gupta, and Emad A. Nabbus. 2011. Electronic Authentication Guideline. *NIST Special Publication 800-63-1* (2011).

[5] William E. Burr, Donna F. Dodson, Elaine M. Newton, Ray A. Perlner, W. Timothy Polk, Sarbari Gupta, and Emad A. Nabbus. 2013. Electronic Authentication Guideline. *NIST Special Publication 800-63-2* (2013).

[6] Tony Caccavo. 2022. "Masked"' Passwords Don't Work the Way You Think. https://teampassword.com/blog/masked-passwords-dont-work.

[7] Rahul Chatterjee, Anish Athayle, Devdatta Akhawe, Ari Juels, and Thomas Ristenpart. 2016. pASSWORD tYPOS and How to Correct Them Securely. In *IEEE Symposium on Security and Privacy (SP)*.

[8] Juliet Corbin and Anselm Strauss. 2014. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* Sage Publications.

[9] MDN Web Docs. 2024. <input type="password">. https://developer.mozilla.org/en-US/docs/Web/HTML/Element/input/password.

[10] Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *CHI Conference on Human Factors in Computing Systems.*

[11] Jessica Enders. 2015. Masking Passwords: Help or Hindrance? https://www.sitepoint.com/masking-passwords-help-or-hindrance/.

[12] Adrienne Porter Felt, Richard Barnes, April King, Chris Palmer, Chris Bentzel, and Parisa Tabriz. 2017. Measuring HTTPS Adoption on the Web. In *USENIX Security Symposium.*

[13] Google. 2024. Overview of CrUX. https://developer.chrome.com/docs/crux/.

[14] P Grassi, Michael E Garcia, and James L Fenton. 2017. Digital Identity Guidelines. *NIST Special Publication 800-63C* (2017).

[15] Mohamed Khamis, Tobias Seitz, Leonhard Mertl, Alice Nguyen, Mario Schneller, and Zhe Li. 2019. Passquerade: Improving Error Correction of Text Passwords on Mobile Devices by using Graphic Filters for Password Masking. In *CHI Conference on Human Factors in Computing Systems.*

[16] Lawrence L Kupper and Kerry B Hafner. 1989. On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics* 45, 3 (1989), 957–967.

[17] Daniel Miessler. 2020. SecLists/Passwords/Common-Credentials. https://github.com/danielmiessler/SecLists/tree/master/Passwords/Common-Credentials

[18] Collins W. Munyendo, Philipp Markert, Alexandra Nisenoff, Miles Grant, Elena Korkes, Blase Ur, and Adam J. Aviv. 2022. "The Same PIN, Just Longer": On the (In)Security of Upgrading PINs from 4 to 6 Digits. In *USENIX Security Symposium.*

[19] National Cyber Security Centre (NCSC). 2018. Password administration for system owners. https://www.ncsc.gov.uk/collection/passwords.

[20] Jakob Nielsen. 2009. Stop Password Masking. https://www.nngroup.com/articles/stop-password-masking/.

[21] Catlin Pidel and Stephan Neuhaus. 2019. BREAKING: Password Entry Is Fine. In *HCI for Cybersecurity, Privacy and Trust.* 67–80.

[22] The Open Web Application Security Project. 2024. Authentication Cheat Sheet. https://cheatsheetseries.owasp.org/cheatsheets/Authentication_Cheat_Sheet.html.

[23] Prolific. 2024. Prolific | Quickly find research participants you can trust. https://www.prolific.com/.

[24] Qualtrics. 2024. Qualtrics XM: The Leading Experience Management Software. https://www.qualtrics.com/.

[25] Suood Al Roomi and Frank Li. 2023. A Large-Scale Measurement of Website Login Policies. In *USENIX Security Symposium.*

[26] Alia Saad, Jonathan Liebers, Uwe Gruenefeld, Florian Alt, and Stefan Schneegass. 2021. Understanding Bystanders' Tendency to Shoulder Surf Smartphones Using 360-degree Videos in Virtual Reality. In *International Conference on Mobile Human-Computer Interaction.*

[27] Bruce Schneier. 2009. The Problem with Password Masking. https://www.schneier.com/blog/archives/2009/06/the_problem_wit_2.html.

[28] Statcounter. 2024. Desktop Browser Market Share Worldwide. https://gs.statcounter.com/browser-market-share/desktop/worldwide.

[29] Statcounter. 2024. Mobile Browser Market Share Worldwide. https://gs.statcounter.com/browser-market-share/mobile/worldwide.

[30] Joshua Tan, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2020. Practical recommendations for stronger, more usable passwords combining minimum-strength, minimum-length, and blocklist requirements. In *ACM Conference on Computer and Communications Security (CCS).*

[31] Lukas Weichselbaum, Michele Spagnuolo, Sebastian Lekies, and Artur Janc. 2016. CSP Is Dead, Long Live CSP! On the Insecurity of Whitelists and the Future of Content Security Policy. In *ACM Conference on Computer and Communications Security (CCS).*

[32] Daniel Lowe Wheeler. 2016. zxcvbn: Low-Budget Password Strength Estimation. In *USENIX Security Symposium.*

[33] Yinqian Zhang, Fabian Monrose, and Michael K. Reiter. 2010. The security of modern password expiration: an algorithmic framework and empirical analysis. In *ACM Conference on Computer and Communications Security (CCS).*

A Survey Instrument

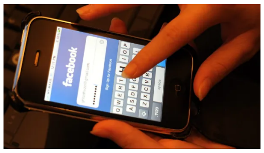
We deployed two survey versions, one for mobile and one for PC. The questions are identical, except for framing specifically for the relevant environment. Here, we present the mobile version.

Introduction

Welcome to our study on password masking!

As shown in the example below, password masking is a practice where the characters of a password are hidden as they are entered (e.g., so it looks like you're entering "***" even though you're typing your actual password).

This research aims to better understand your experiences and perceptions regarding different password masking techniques when using your phones or tablets. When answering the questions, **please only discuss your experiences or preferences when using phones and tablets specifically**, and not other devices. Your participation is invaluable in helping us improve the design of password masking in the future.



Thank you for participating in our study, and we appreciate your time and insights!

Demographics

Q1 How old are you?

- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- 55-64 years old
- 65+ years old

Q2 What's your gender?

- Male
- Female
- Non-binary / third gender
- Prefer not to say

Q3 What is the highest level of education you have completed?

- Some high school or less
- High school diploma or GED
- Some college or associate degree
- Bachelor's degree
- Graduate or professional degree
- Doctoral degree

Q4 Specifically during logins on your phones or tablets, over the past few weeks, how often do you use a password auto-fill feature (such as provided by a password manager) when you were logging in?

- Never
- Sometimes
- About half the time
- Most of the time
- Always

Past Experiences with Masking

Q5 Why do you think websites use password masking?

- Answer: ______

Q6 Specifically during logins on your phones or tablets, over the past few weeks, how often did you encounter masked input fields when you were logging in?

- Never
- Sometimes
- About half the time
- Most of the time
- Always

Impact on Usability

Q7 Specifically during logins on your phones or tablets, can you describe any issues or challenges you have encountered when entering passwords with masking?

- Answer: ______

Q8 Specifically during logins on your phones or tablets, over the past few weeks, how often did you encounter issues when entering passwords with masking?

- Never
- Sometimes
- About half the time
- Most of the time

- Always

Q9 Specifically during logins on your phones or tablets, how easy is it to enter passwords with masking compared to without masking?

- Much easier
- Somewhat easier
- Roughly the same
- Somewhat more difficult
- Much more difficult

Q10 Does password masking influence the passwords you select?

- Yes
- No

Q11 Could you explain your answer to the last question?

- Answer: ______

Perception of Security and Privacy

Q12 Specifically during logins on your phones or tablets, how often are you in a setting/situation where someone else might be able to see you type your password?

- Never
- Sometimes
- About half the time
- Most of the time
- Always

Q13 Specifically during logins on your phones or tablets, can you explain some of the situations/settings you encounter where someone else might be able to see you type your password?

- Answer: ______

Q14 Specifically during logins on your phones or tablets, you feel more secure or private when accessing websites that mask their passwords compared to those that don't.

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

Q15 Could you explain your answer to the last question?

- Answer: ______

Q16 (Select all that apply) What type of websites or online platforms do you find it most important to have password masking for?

- Social Media (e.g., Facebook, Twitter, Instagram)
- Banking or Financial Platforms
- Email Services (e.g., Gmail, Outlook)
- Shopping Websites (e.g., Amazon, eBay)
- Healthcare Platforms
- Work or Professional Accounts
- Entertainment Services (e.g., Netflix, Spotify)
- Educational Platforms
- Other (Please specify)

Q17 Specifically during logins on your phones or tablets, how important is it for you to have your passwords visually hidden, even if it involves some inconvenience?

- Not at all important
- Slightly important
- Moderately important
- Very important
- Extremely important

Q18. Attention Check Specifically during logins on your phones or tablets, what's the most secure way to get a charger? Please just select "Avoid purchasing it".

- Go to convenient store
- Online shopping
- Avoid purchasing it
- Borrow it from your friends

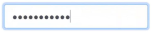
Preferred Masking Format

Q19 Specifically during logins on your phones or tablets, when toggle options are available to reveal your masked passwords, how often do you use them?

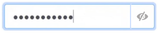
- Never
- Sometimes
- About half the time
- Most of the time
- Always

Q20 Specifically during logins on your phones or tablets, which form of masking would you prefer?

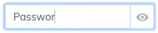
- Masking on by default, without an option to view the typed password


•••••••••••

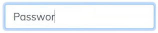
- Masking on by default, with an option to view the typed password


•••••••••••

- Masking off by default, with the option to mask the typed password


Passwor

- No masking

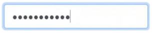

Passwor

Q21 Could you explain your answer to the last question?

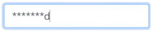
- Answer: ______

Q22 Specifically during logins on your phones or tablets, if masking is enabled, which form of masking do you prefer?

- Every character is always masked when typed


•••••••••••

- Last character typed is unmasked temporarily (e.g., for a second) before becoming masked


*******d