Multiple Password Interference in Text Passwords and Click-Based Graphical Passwords

Sonia Chiasson[1], Alain Forget[1], Elizabeth Stobert[2],
P.C. van Oorschot[1], Robert Biddle[1]
[1]School of Computer Science, [2]Department of Psychology
Carleton University, Ottawa, Canada
{chiasson, aforget, paulv}@scs.carleton.ca
estobert@connect.carleton.ca, robert_biddle@carleton.ca

ABSTRACT

The underlying issues relating to the usability and security of multiple passwords are largely unexplored. However, we know that people generally have difficulty remembering multiple passwords. This reduces security since users reuse the same password for different systems or reveal other passwords as they try to log in. We report on a laboratory study comparing recall of multiple text passwords with recall of multiple click-based graphical passwords. In a one-hour session (short-term), we found that participants in the graphical password condition coped significantly better than those in the text password condition. In particular, they made fewer errors when recalling their passwords, did not resort to creating passwords directly related to account names, and did not use similar passwords across multiple accounts. After two weeks, participants in the two conditions had recall success rates that were not statistically different from each other, but those with text passwords made more recall errors than participants with graphical passwords. In our study, click-based graphical passwords were significantly less susceptible to multiple password interference in the short-term, while having comparable usability to text passwords in most other respects.

Categories and Subject Descriptors

K.6.5 [**Management of computing and information systems**]: Security and protection: Authentication

General Terms

Security, Human Factors

Keywords

authentication, graphical passwords, multiple password interference, usable security

1. INTRODUCTION

Special consideration is required to design usable, understandable, and manageable security features. At first glance, it seems like applying standard usability and Human-Computer Interaction (HCI) principles should suffice, but security constraints make this problematic. Most importantly, some design features that might make a system more usable would also make it less secure. Addressing these security weaknesses can too easily render the software unusable again. Even worse, one might argue that an unusable security system is *inherently* insecure, since users will then misuse or bypass the security mechanisms. One must also consider how the design affects the observable behaviour of legitimate users, in case such behaviour could be exploited by attackers. The challenge is to design software that is both secure and usable [10].

In this paper, we address an important issue in user authentication software: the memorability of multiple passwords. Authentication software supports legitimate users in gaining access to systems or resources by verifying their credentials. We focus on passwords, the most common form of credentials. The problem with passwords is making them easy for legitimate users to remember, but difficult for attackers to guess. Alternatives to passwords include physical tokens or biometrics; these also have problems, such as cost, management, and privacy, which we will not address in the paper. As passwords are the most common method of authentication, the password problem is important, and is made worse by the increasing number of users and the number of different systems they access [15, 18]. In particular, users now need to remember not just one password, but many. This places a significant memory load on users, leading them to choose (and reuse) simple passwords that are easy for attackers to guess. Despite this reality, there has been little work investigating the issues relating to multiple passwords.

Our current work is motivated by recent proposals for alternative kinds of passwords, particularly click-based graphical passwords [3, 39]. In such systems, the user does not enter a text password using a keyboard, but instead clicks on particular points on an image. Such graphical passwords are intended to take advantage of the human ability to more easily recognize and recall images than textual information [29]. We wished to study whether this approach had advantages over text passwords when multiple distinct passwords were necessary. We were concerned about the potential for multiple password interference, where remembering a password for one system might affect the user's memory of a password for another system. As described in the cognitive psychology literature [2], *memory interference* is "the impaired ability to remember an item when it is similar to other items stored in memory".

Our study was conducted in a laboratory setting where 65 participants were assigned to use either textual or graphical passwords. They created six distinct passwords for several different "accounts", and later had to recall the passwords for each account in a different order than they were created. Twenty-six of those participants also returned after two weeks to test recall of these passwords. In the

case of graphical passwords, each account was associated with a different image, so participants had one image per password. Literature on memory research acknowledges that ceiling effects, such as high success rates that mask differences between conditions, are a problem which must be overcome with careful experimental design in order to get meaningful results [22]. While our study does not mirror real-life usage of passwords, the intent of the experiment was to highlight differences in performance.

We found that in the short-term, participants had more difficulty recalling multiple text passwords than multiple graphical passwords. However, after two weeks, both groups had significant difficulty remembering their passwords and there was no statistical difference in success rates. We further found that participants in the text condition could more easily recall their passwords when they used insecure password practices, such as choosing passwords that followed a common pattern or that were obviously associated with account names. For example, 40% of text passwords were obviously related to their associated account. These results constitute evidence for an important advantage inherent in click-based graphical passwords – built-in cueing that helps with memorability – while text password systems encourage users to adopt insecure coping strategies. In post-hoc analysis, we found that males were more successful than females with graphical passwords and that males were more likely to use account-related text passwords. These results align with psychology literature showing that males perform better at visual-spatial tasks and more weakly in linguistic tasks than females [5,23]. The results of this paper motivate further study in field settings and deeper examination of the underlying human factors issues involved in using these authentication mechanisms.

The remainder of the paper is divided as follows. Section 2 provides background on the type of graphical password system used, memory cueing, and multiple password interference. In Sections 3 and 4, we outline our hypotheses, describe the methodology of our study, and present the results. We discuss validation of our hypotheses in Section 5. Lastly, we offer some discussion and concluding remarks in Sections 6 and 7.

2. BACKGROUND AND RELATED WORK

Security is rarely a user's primary task [37], and typically involves an extra step in addition to the main task, such as having to log in to read one's email. Users need security features to be as non-disruptive as possible, but still need them to work properly to preserve integrity and privacy. A second unusual characteristic of security software is that it attracts illegitimate users of the system who are actively trying to gain unauthorized access. These attackers will take advantage of all information available. Usable security software must therefore offer assistance to legitimate users, without giving assistance to attackers. In particular, this changes the nature of feedback in interaction design, which must inform legitimate users while revealing no useful information to others.

With any authentication system where users are expected to recall information to log in, there is a risk of memory interference. Multiple password interference occurs when users must remember passwords for many systems and the memories of the different passwords interfere with each other. Studies have shown that users typically create easy-to-guess text passwords and reuse these passwords across several accounts [1, 15, 18]. When trying to log in, they will cycle through their passwords until they find one that works. Gaw and Felten [18] report that users in their lab study tried an average of 2.43 passwords before a correct login. This may be under-reporting the problem, however, because users in their study were only allowed 90 seconds per account. While this trial-and-error approach helps users deal with password systems and multiple password interference, revealing all of one's passwords at every login can amplify security risks, for example in the presence of key loggers or when passwords are sent to phishing sites.

One proposed solution to the password problem is to use a password manager. With a password manager, users typically have one master password and the password manager creates, stores, and enters passwords for individual accounts on behalf of the user. The individual passwords are typically much more random than what users would select on their own and are thus stronger against attack. However, implementations of some password managers have usability problems [9] that can leave users even more vulnerable than when they were managing passwords themselves. A second drawback is that a centralized scheme has a new single point of failure: if attackers gain access to the master password, they now have control over all of the user's accounts. While password managers may be appropriate in some circumstances, authentication schemes that are both secure and memorable are still needed.

We are interested in the graphical password approach. It has been suggested that graphical passwords may be less susceptible to multiple password interference since humans have better memory for recognizing and recalling images than text [24, 31]. Surveys of graphical passwords circa 2005 are available from Suo et al. [32] and from Monrose and Reiter [26]. Proposed schemes include click-based graphical passwords such as PassPoints [39]. Many of these have the added advantage of presenting a cue to the user to help trigger the appropriate memory. Cued-recall has been established as an easier memory task than uncued recall [29, 34]. With cued-recall, the system provides a cue to help prompt the user's memory of the password (or a portion thereof). This is a desirable usability feature that reduces the memory load on users. With click-based graphical passwords, a password consists of user-selected click-points on the images presented. Therefore, the images act as mnemonic cues to remember the corresponding click-points.

In PassPoints, users are presented with an image, and a password consists of 5 click-points on the image (see Figure 1). To log in, users must select the same 5 click-points in the same order. The system allows for a tolerance area around each click-point so that approximately correct login attempts are accepted. Several user studies and security analyses have been conducted on PassPoints [6, 13, 20, 33, 38–40]. While these have found PassPoints to be generally usable, security concerns have been raised because users tend to select predictable passwords which are exploitable in dictionary attacks [13, 30, 33]. Newer click-based graphical password schemes, such as Persuasive Cued Click-Points [7,8], address two important security concerns with respect to user selected passwords [11]: they offer a significant reduction in hotspots (i.e., areas of the image that have higher probability of being selected by users) and in the use of click-point patterns (such as selecting click-points that form a straight line across the image). These characteristics significantly reduce vulnerability to dictionary attacks. The present paper uses the better-known PassPoints scheme for these interference tests, in order to leverage a more closely-examined and understood password scheme and to build on existing results [6] on interference between two passwords only (see below).

A few studies have compared text passwords to graphical passwords, but in these cases, users only had one password to remember (either text or graphical). Wiedenbeck et al. [39] compared user performance of text passwords and PassPoints in a lab study. Their results were mixed, but slightly favoured text passwords. Komanduri and Hutchings's study [21] compared text passwords to their newly proposed picture-password scheme. They found better memorability for their picture-passwords although the results were not statistically significant due to a small user sample.

Figure 1: A PassPoints password consists of 5 ordered click-points (the numbered labels do not appear in practice).

To our knowledge, there is little published work examining the problem of multiple password interference, despite the growing number of passwords held by most users. Moncur and Lepatre [25] compared VIP, a recognition-based graphical password scheme in which users select their images from a set of decoys, to a graphical variation of 4-digit PINs (Personal Identification Numbers). While they showed that users were slightly more likely to recall 5 VIP passwords than PINs, it is unclear how this compares to text passwords. Passwords in their study were not associated with any "accounts", and the study did not take into consideration serial memory effects. Recently, Everitt et al. [14] investigated how interference and frequency of access affected memorability of 4 recognition-based PassFaces [11] passwords. They found that infrequently accessed passwords were more difficult to remember and that users performed better when they had a chance to practice each new password individually over several days rather than learning several at once. Both of these studies focus solely on usability and do not consider the security of the schemes themselves or whether changes in user behaviour when dealing with multiple passwords may affect the security of the system (although passwords were assigned in the PassFaces study to eliminate user choice).

Multiple password interference was also examined as part of a field study of PassPoints [6]. In this study, a subset of participants had two distinct passwords to remember (on two different images). These participants had lower login success rates than those with only one password. We are not aware of comparable studies for regular text passwords, so it is unknown how this performance decrease compares with text passwords. Vu et al. [36] conducted lab studies examining the effect of various text password restrictions on memorability when multiple passwords were used. Their results include that users with five passwords had more difficulty than those who had only three, that some users selected passwords with obvious connection to their accounts, and that password restrictions were not sufficient for encouraging secure text password selection.

3. STUDY DETAILS

We hypothesized that click-based graphical passwords would be easier for users to recall than text passwords when users had multiple passwords to remember. In other words, there would be less interference from multiple unique graphical passwords than multiple unique text passwords. Although many variants of graphical passwords and text passwords were available, we began our investigation with regular text passwords, where users were free to select any password, and PassPoints, the click-based graphical system that had been most closely evaluated to-date. Our experiment compared multiple password interference for these two conditions: the Text condition (MText) and the PassPoints condition (MPP).

Our specific hypotheses with respect to multiple password interference were:

1. Participants will have lower recall success rates with text passwords than with PassPoints passwords.
2. Participants in the Text condition are more likely than PassPoints participants to use patterns across their own passwords.
3. Participants will recall text passwords more slowly than PassPoints passwords.
4. Participants in the Text condition are more likely than PassPoints participants to create passwords that are directly related to their corresponding accounts.
5. Participants in the Text condition will make more recall errors than participants in the PassPoints condition.

We conducted a lab study with 65 participants (26 males and 39 females). Participants completed their sessions individually. This study used a between-subjects design and had two conditions; half of the participants were randomly assigned to the Text password condition and half to the PassPoints password condition. All participants were familiar with text passwords, but no participant in the PassPoints condition had any previous experience with graphical passwords. Participants were primarily university students from various degree programs. All were regular internet users, but none were experts in computer security.

Programs for the Text and PassPoints conditions were implemented as stand-alone Windows applications and displayed on a 17-inch screen. The PassPoints application used 451x331 pixel images, tolerance areas of 19x19 pixels, and 5 click-points. This configuration is consistent with previous studies [6, 39]. The images are identified as: Cars, Mural, Philadelphia, Pool, Statue, and Truck (Figures 2 to 7). These were images from a previous PassPoints lab study [6] and shown to have average to good usability and security. A PassPoints system using this configuration has a theoretical password space of 2^{44} possible passwords. The Text password system enforced an 8-character minimum, but no other restrictions were imposed; this gives a theoretical password space of 2^{52}. While 32 special characters are available on a standard keyboard, most users use a very small subset of these special characters. In fact, there is evidence that they do not even realize that these can be included in a password or know how to type them [16]. While these special characters are included in the theoretical password space, it is highly improbable that passwords contain any of them. The theoretical password space of PassPoints could be enlarged to match that of text passwords through different system configuration, but we chose to maintain compatibility with previous studies. Reducing the text password limit to less than 8 characters also seemed to be a poor alternative.

3.1 Methodology

Our study included two lab-based sessions. Session 1 took one hour and was completed by all 65 participants. For Session 2, participant returned to the lab and tried to recall their previously created passwords. The second session occurred after two weeks and was completed by 26 participants.[1]

[1]Session 2 was added to our methodology after we examined the initial results. 26 out of 28 participants recruited after this methodology change completed Session 2.

Figure 2: Cars image [4]

Figure 3: Mural image [40]

Figure 4: Philadelphia image [40]

Figure 5: Pool image [27]

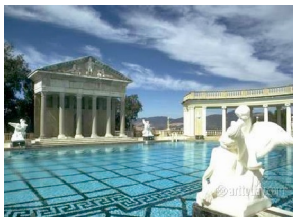

Figure 6: Statue image [17]

Figure 7: Truck image [17]

Table 1: Methodology

Phase		Number of trials	Steps
I.	Practice	2 trials. This data was not used in the analysis.	Create Confirm Answer Questions Distraction Login
II.	Password Generation	6 trials. Accounts were presented in the same order for all participants.	Create Confirm Answer Questions Distraction Login
III.	Retention	6 trials. Account order was shuffled according to the Latin square.	Recall-1
IV.	2-week Retention	6 trials. Account order was the same as for the Retention phase.	Recall-2

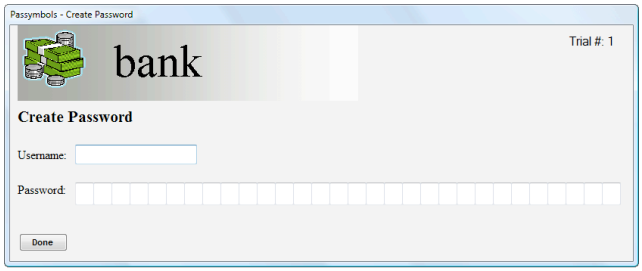

Figure 8: Password creation interface (Text condition).

3.1.1 Session 1

The initial one-hour session was divided into three phases: Practice, Password Generation, and Retention, as shown in Table 1. First, participants completed a Practice phase with two trials. For each trial, they created, confirmed, and logged in with one password. This phase was used to explain the process and familiarize participants with the user interface. During Practice phase, participants were told that they did not need to remember their practice passwords and would not be asked about them again.

In the Password Generation phase, participants completed 6 trials where they created distinct passwords, each associated with a different pre-defined "account": bank, email, instant messenger, library, online dating, and work. The accounts were identified by coloured banners at the top of the application window that included a unique icon and the account name (see Figures 8 and 9). In this phase, the accounts were presented to all participants in the same order. For PassPoints, each account was associated with a distinct image, so participants never had more than one password per image. Furthermore, the PassPoints accounts were consistently paired with the same images (although in a real implementation, a system would use different images for different participants and offer a new image if a user reset a password). This design decision allowed us to gather enough data on each image to assess whether hotspots and patterns occurred as a result of having multiple passwords, and to remove a potential confounding variable. Participants were asked to pick realistic passwords that they could remember but that would be difficult for others to guess. They were further told that they would need to remember these passwords later and reminded that each password was created for a specific account. In total, 395 account passwords were created: 204 in the Text condition and 191 in the PassPoints condition.

During the Practice and Password Generation phases, participants used the following 5-step process for each of their accounts.

1. Create. Participants created a password for the given account. Those in the Text condition had an 8-character minimum and had to enter their password twice (see Figure 8). The text passwords were visible during password creation. Those in the PassPoints condition were required to click on 5 different click-points on the provided image (see Figure 9). The selected click-points were never made visible to participants at any time during the study.

2. Confirm. Participants confirmed their password. For text passwords, participants entered their password, now echoed only as asterisks. PassPoints participants had to enter their password by clicking on the same 5 ordered click-points, within a 19x19 pixel tolerance area of the original click-points. If participants could not remember their password, they could return to the Create step.

3. Answer Questions. Participants responded to two 10-point Likert-scale questions about the perceived difficulty of creating and remembering the current password.

4. Perform Distraction Task. A 30-second distraction task was used to simulate a longer passage of time [19]. Participants in the Text condition counted backwards by 3s from a random 4-digit number, while participants in the PassPoints condition completed a Mental Rotations Test (MRT) puzzle [28]. Different tasks were used to clear textual working memory and visual working memory.

Figure 9: Password creation interface (PassPoints condition).

5. Login. Participants re-entered their password. They could re-try as many times as necessary to get it correct. If they forgot their password, they could return to the Create step.

After a short break where participants completed a demographics questionnaire, they moved on to the Retention phase of the study. The Retention phase tested whether participants could recall each of their 6 passwords. The application prompted participants to log in by displaying the account name and banner along with an entry field for their username and a password entry field or an image, depending on whether participants were in the Text or PassPoints condition. Participants could re-try as many times as they wished if they made mistakes, and an additional button was available: "I don't remember creating a password for this account". In fact, we never asked participants to enter passwords for which they had no account, but they may have forgotten that they created a password for a particular account.

In the Retention phase, the accounts were presented in shuffled order based on a Latin square, where each row represented a participant and each column represented an account. Pre-testing showed that one hour was sufficient for the generation and later recall of 6 passwords, therefore we used a 6×6 Latin Square for determining presentation order during the Retention phase. The Latin square ensures that the presentation order of the accounts is balanced across all participants, avoiding possible bias that might otherwise result from serial order effects.

To complete Session 1, participants answered a paper questionnaire aimed at gathering their perception of the password system and their general attitudes towards passwords.

3.1.2 Session 2

Twenty-six participants returned to the lab 12 to 15 days after their Session 1 to complete the second part of the study. This 2-week Retention phase followed the same procedure as the initial Retention phase and accounts were presented in the same order. Participants could re-try as many times as necessary to recall their password or decide that they had forgotten it. At the end of the session, participants completed a short paper questionnaire.

4. RESULTS

As is the standard approach in analyzing data from user studies, we used various statistical tests to assess whether differences in the data reflect actual differences between conditions or whether these may have occurred by chance. Three types of statistical tests for significance were used in the present analysis, each intended to determine whether the groups being analyzed were distinct from each other with respect to the factor being tested. We used t-tests to compare variance of the means between two groups, Mann-Whitney tests to compare ordered categorical data (e.g., Likert scale responses, where the choices are discrete and ordered, but it cannot be assumed that participants view all pairs of adjacent levels as equidistant), and Chi-square (χ^2) tests to compare non-ordered categorical or nominal data (e.g., comparing login success/fail ratios). In all cases, we regard a value of $p < .05$ as indicating that the groups being tested are different from each other with at least 95% probability, making the result statistically significant. In the tables, "not significant" indicates that the test revealed no statistically significant differences between the two conditions (i.e., $p > .05$).

4.1 Success Rates

We first examine success rates as a measure of participants' performance. The success rate is the number of successful password entry attempts divided by the total number of attempts, across all participants. We calculate the success rates specifically for the first attempt in a given step, after 3 attempts, and then consider all attempts. During Password Generation, participants could confirm and login with Text and PassPoints passwords equally well (see Table 2). There was no statistically significant difference in success rates between conditions for this phase of the study. These results are similar to previous studies [6, 16] with the same methodology.

Recall-1. During Recall-1, we found a significant difference in success rates between the Text and PassPoints conditions. Participants in the PassPoints condition were significantly more likely to successfully recall their password than those in the Text condition. As reported in Table 2, when we consider only the first attempt for each password, participants recalled their text password correctly only 68% of the time while those in the PassPoints condition had a 95% success rate. Participants could try recalling their password as many times as they wished, until they either succeeded or gave up. Participants in the Text condition reached an 88% success rate with multiple recall attempts, compared to 99% for PassPoints participants. This means that for 12% of trials in the Text condition, participants eventually said they could not recall their password. Only 1% of trials in the PassPoints condition ended with such a failure to recall. Very high success rates in many categories indicate that participants' memory was not strongly taxed during Session 1. However, this makes the much lower success rate for recall in the Text condition all that more remarkable.

Recall-2. During Recall-2, participants in both conditions had difficulty remembering their passwords (see Table 2), perhaps indicating that this was a more difficult memory task. Two weeks after creating their passwords, only 70% of Text participants and 57% of PassPoints participants were able to successfully recall their passwords. The differences in success rates between the two conditions were not statistically significant.

We took a closer look at the Recall-2 data to understand where the difficulties arose. We found that male participants were significantly more likely to successfully recall their PassPoints passwords

Table 2: Success rates for the Confirm, Login, Recall-1, and Recall-2 steps. MText indicates Multiple Text password and MPP indicates Multiple PassPoints passwords.

	Password Generation				Retention					
	Confirm		Login		Recall-1			Recall-2		
	MText	MPP	MText	MPP	MText	MPP	χ^2	MText	MPP	χ^2
First Attempt	98%	95%	96%	100%	68%	95%	$\chi^2(1,395) = 46.68, p < .001$	30%	38%	not significant
Within 3 Attempts	99%	99%	99%	100%	84%	99%	$\chi^2(1,395) = 27.96, p < .001$	59%	57%	not significant
Multiple Attempts	100%	99%	99%	100%	88%	99%	$\chi^2(1,395) = 18.43, p < .001$	70%	57%	not significant

than women. As shown in Table 3, 71% of males correctly entered their PassPoints password within 3 attempts as opposed to only 40% of female participants. This result aligns with psychology research which continues to show that males tend to perform better at visual-spatial tasks, while females generally have better performance with linguistic tasks [5,23]. Although this gender difference was not apparent in the success rates for Text passwords, males were more likely to employ a coping mechanism to help remember their text passwords (see Section 4.4).

4.2 Recall Errors

We now focus on the Retention phase and examine the types of errors committed by participants trying to recall their passwords.

Recall-1. As shown in Table 4, participants had more difficulty recalling text passwords than PassPoints passwords ($t(238.47) = 5.428, p < .001$) during Recall-1.[2] In total, participants in the Text condition made 173 errors, while participants in the PassPoints condition made 17 errors. Participants who were unsuccessful could re-try as many times as they wished, so there could be many more errors than trials.

Each group committed different types of errors. Participants in the Text condition appeared to be more affected by interference from multiple passwords during Session 1. In the Text condition, 20 out of 34 participants made recall errors. They often tried passwords from other accounts when asked to recall a password for a particular account. Many participants cycled through several of their passwords before reaching the correct one or giving up. In these cases, they either entered the exact password for another account, some variant of the exact password, or a variant of a password for another account. For example one user had "870103zx" as a password, but repeatedly entered "zx870103', although even if entered correctly this password would have been incorrect since it was for another account (we call these *wrong account variants*). A few errors were due to misspelling or variants of the correct passwords, such as entering "access!ble" when the password was "@ccessible" or entering "mybnakpwd" instead of "mybankpwd" (we call these *misspelled variants*). The types of errors for text passwords are summarized in Table 5.

As shown in Table 6, over half of the errors in the PassPoints condition were due to forgetting one or more click-points within the password. For one trial, the user remembered the pattern of the password (a straight horizontal line), but thought it was approximately 15 pixels lower, aligning to a lower linear feature on the Cars image (Figure 2), and tried to enter it 4 times in this shifted position. One user knew the correct password but one click-point was slightly outside of the tolerance region, while another user entered the correct click-points but in the wrong order. As expected, no user confused their password from one account for another, since all passwords were based on different images. In the PassPoints condition, 8 out of 32 participants made recall errors.

Recall-2. Participants in both conditions made considerably more errors after the two-week interval. In total, there were 216 Recall-2 errors in the Text condition, which is statistically greater than for Recall-1 ($t(122.39) = 4.61, p < .001$). Similarly, the 106 Recall-2 errors for PassPoints is significantly greater than for Recall-1 ($t(67.23) = 5.99, p < .001$).

Text participants were more likely to re-try entering their passwords, making on average 3.14 attempts per account, while PassPoints participants made 2.2 attempts on average ($t(150) = 2.575$, $p < .05$) before either successfully entering their password or giving up. As discussed later, this occurred because participants in the Text condition were more likely to cycle through their passwords or variations of their passwords when they did not know the password for a particular account. The distribution of errors in Tables 5 and 6 show that for Recall-2, participants in both conditions were more likely to enter passwords that had no relation to any of their 6 account passwords or that were variations of an existing password than during Recall-1. Interestingly, a common error in the PassPoints condition was to have some, or all, of the x-coordinates correct, but have incorrect y-coordinates. This occurred on several images (not only Cars, as in Recall-1).

4.3 Timings

Table 7 and Figure 10[3] show the times taken to complete each password-related step of the study and provide results of t-tests comparing the times from each condition. These times represent the total time spent during a given step. For example, the time to login began when the login screen first appeared and continued until the user entered their username and password, then successfully logged in (including any errors committed).

During Password Generation, there is no consistent relationship between the two conditions with respect to the amount of time taken at each step. Participants were faster at creating PassPoints passwords than text passwords. In contrast, participants were slower at confirming their password in the PassPoints condition than in the Text condition. During Login, participants took approximately the same amount of time to enter their password in both conditions.

Recall-1. During Recall-1, participants were quicker at entering PassPoints passwords ($t(262.64) = 3.93, p < .001$). This aligns with the fact that participants made fewer errors in the PassPoints condition and hence spent less time repeatedly entering their pass-

[2]In this and subsequent t-tests, if the value for degrees of freedom (i.e., the number in parentheses) includes a decimal point, it is because Levene's Test determined that equal variances could not be assumed and so a Welch Two Sample t-test was used.

[3]Notched box-and-whisker graph can be interpreted as follows. Each box indicates the Inter-Quartile Range (IQR - the interval between the 25th and 75th percentiles) while the dashed lines (whiskers) represent the first and fourth quartiles. The narrowest part of the box indicates the median time for each phase. The notches surrounding the median represent the confidence intervals. If the notches of two boxes do not overlap, then their medians are significantly different from each other at $p < .05$.

Table 3: Success rates of male and female participants for the Recall-2 step.

	MText			MPP		
	Male	Female	χ^2	Male	Female	χ^2
First Attempt	33%	28%	not significant	43%	33%	not significant
Within 3 Attempts	63%	57%	not significant	71%	40%	$\chi^2(1, 65) = 6.51, p < .05$
Multiple Attempts	80%	65%	not significant	71%	40%	$\chi^2(1, 65) = 6.51, p < .05$

Table 4: Recall errors per condition. Recall-1 includes 204 Text trials and 191 PP trials, while Recall-2 includes 90 Text trials and 65 PP trials. For the total number of errors, the percentage is calculated as the total number of errors divided by the total number of Recall attempts for that condition.

	Recall-1				Recall-2			
	MText		MPP		MText		MPP	
Number of trials with errors	65	(32%)	10	(5%)	62	(69%)	40	(62%)
Total number of errors	175	(49%)	18	(9%)	220	(78%)	106	(74%)
Number of trials where participants gave up	24	(12%)	2	(1%)	27	(30%)	28	(43%)
Number of participants who made errors	20/34	(59%)	8/32	(25%)	15/15	(100%)	10/11	(91%)

Table 5: Classification of recall errors for the Text condition (MText).

	Number of Occurrences	
Type of error	Recall-1	Recall-2
Wrong account	97	82
Wrong account variant	38	27
Misspelled variant	17	63
Unknown	21	44
Total number of errors	173	216

Table 6: Classification of recall errors for the PassPoints condition (MPP).

	Number of Occurrences	
Type of error	Recall-1	Recall-2
1 click-point outside of tolerance area	1	0
Incorrect click-point order	1	3
Forgotten click-points	9	60
Click-point pattern shifted	6	15
Partial click-point pattern shifted	0	28
Total number of errors	17	106

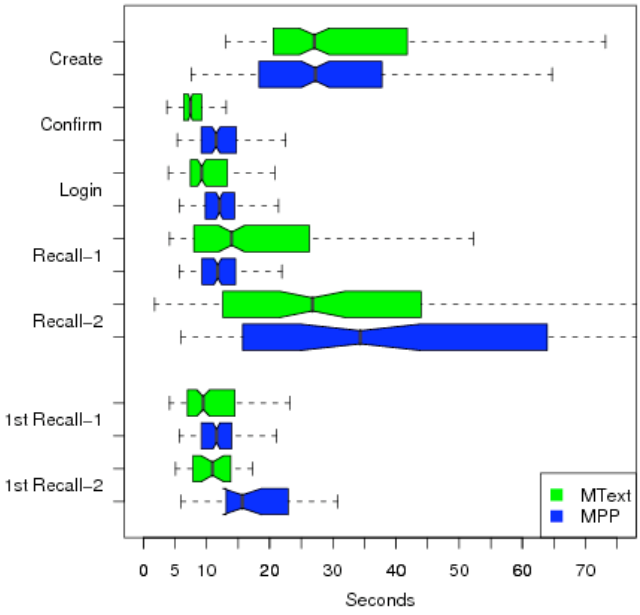

Figure 10: The total time for each step is shown in seconds. The upper whiskers for the Recall-2 phase end at 90 seconds for MText and 133 seconds for MPP. For the recall phases, the time taken for accounts where passwords were correctly entered on the first attempt are also shown as 1st Recall.

word. If we consider only recall attempts where participants made no mistakes (correctly entered their password on the first attempt), then the difference in entry time is not significant. In other words, during both the Login and Recall-1 steps of Session 1, participants entered PassPoints passwords at least as quickly, or more quickly, than text passwords.

Recall-2. There is no significant difference in the total amount of time taken by participants in the Text and PassPoints conditions in Recall-2. However, if we consider only those participants who were successful on their first attempt, then Text participants were quicker at entering these successful passwords.

4.4 Use of Mnemonics

When faced with the task of remembering multiple items, people naturally turn to memory aids, or mnemonics. In our study, participants in the PassPoints condition had a built-in mnemonic since they could use the image as a memory aid. We gave no instructions to participants in either condition as to what they could use as memory aids. No user tried to write down their passwords. The accounts were identified by banners just above the username and password entry fields (see Figures 8 and 9). We investigated whether various account characteristics, such as account names, types, or banners, were used as mnemonics.

We manually classified the passwords in the Text condition according to whether they were related to their account. We found that 23 out of 34 (68%) participants in the Text condition used the account as a cue for at least one of their passwords. Some passwords were directly linked with the account name. For example, one user entered “instantmsg” for the instant messenger

Table 7: Timings for each step in seconds and results of t-tests comparing the timings for the two conditions.

Phase	Step	Mean MText	Mean MPP	Median MText	Median MPP	t-test
Password Generation	Create	34.6	30.7	27.1	27.2	not significant
	Confirm	8.9	13.2	7.4	11.5	$(t(379) = 6.41, p < .001)$
	Login	11.3	13.4	9.2	12.0	$(t(393) = 2.87, p < .01)$
Retention	Recall-1	29.3	15.1	14.0	11.8	$(t(262.64) = 3.93, p < .001)$
	Recall-2	42.1	47.0	26.8	32.6	not significant
	Recall-1 (correct on 1st attempt)	16.6	13.9	9.5	11.6	not significant
	Recall-2 (correct on 1st attempt)	11.2	18.1	10.9	15.7	$(t(34.08) = 3.51, p < .01)$

account. Others were somewhat related, such as "lovelove" for the online dating account. In total, 40% of text passwords were related to their account. Other passwords appeared to be in languages we did not understand and may have corresponded to their accounts, but we did not count these in our totals. We observed gender differences in the use of mnemonics, with males being more likely to create passwords that were directly related to their accounts ($t(32) = 2.07, p < .05$).

We found no apparent link between passwords created in the PassPoints condition and their associated accounts. Possible explanations include that participants in the PassPoints condition either did not need an additional mnemonic device since they could already use the password image, were unable to find a way to use the account characteristics as memory aids for this type of password, or did so in a manner that was not apparent to us.

Recall-1. Since the use of mnemonics only applies to text passwords, we compared success rates for participants who used account-related passwords and those that did not for the Text condition. Participants classified as having used account-related text passwords had a 96% success rate for Recall-1 while those who did not had an 83% recall success rate (χ^2(1,204)=8.68, $p < .01$). The use of account-related passwords made it significantly easier for Text participants to recall their passwords during Session 1.

Recall-2. Similarly, we examined whether the use of account-related passwords affected the Recall-2 success rates for the Text condition. Those classified as having created account-related passwords had a 71% success rate for Recall-2, while those who did not had a 69% success rate. A Chi-square (χ^2) test shows that this difference was not statistically significant. Participants who created account-related passwords were no more likely to remember their passwords after two weeks. In our study, creating account-based text passwords helped with memorability in the short term, but it did not provide an advantage after two weeks.

4.5 Patterns

We further evaluated whether participants were more likely to use a coping strategy, such as selecting predictable passwords, when faced with the task of creating and remembering several passwords.

4.5.1 Text Password Patterns

We visually inspected all of the passwords created in the Text condition to see if a given user created similar passwords for all 6 of his or her accounts. Although they may help with memorability, patterns across accounts are a security vulnerability because an attacker who learns a user's password for one (perhaps weakly protected) account may be able to more easily guess passwords for the user's other (perhaps more important) accounts.

We found that 18 out of 34 participants (53%) in the Text condition created at least one pair of passwords that were similar to each other. In total, 71 out of 204 passwords (35%) were obviously related to other passwords created by the same user. An example of such a pair included: "ins901333" for the instant messenger account and "lib901333" for the library account. In this case, the passwords followed a pattern across passwords and were also directly related to the corresponding accounts. This particular user applied this strategy to all 6 passwords.

4.5.2 PassPoints Patterns

We examined whether the PassPoints passwords followed simple patterns, based on previous work on click-based graphical passwords [8] which classified the types of patterns created by the click-points of user-chosen passwords. The earlier study found that in PassPoints, participants were likely to select click-points in simple patterns such as a straight line or C- shape.

We tested the passwords from the current study for patterns, concerned that participants may be even more likely to resort to common patterns if they had several passwords to remember. A comparison of the types of patterns found in the current study and those from the previous PassPoints study [6] is provided in Figure 11. We found no statistical difference between the patterns found in the current study (where participants had to create and remember multiple passwords) and the earlier PassPoints lab study (where participants had to remember only one password at a time). So although we did see patterns, we have no evidence that they occur more frequently than when participants had only one graphical password to remember. It appears that contrary to text passwords, participants do not resort to these types of patterns as an additional coping strategy for handling multiple click-based graphical passwords.

We also examined each user's 6 passwords to see if anyone consistently picked passwords in a given pattern. Two participants had 4 out of 6 passwords following a "Z" pattern, but no other participants used a specific pattern for the majority of their passwords. It is possible that with additional experience with PassPoints passwords, users might develop different password selection strategies.

4.6 Other Security Measures

Rather than revisit known security issues (for example, see the PassPoints discussion in Section 2), in this section we are interested in exploring whether the requirement of remembering multiple passwords exacerbates these security issues.

4.6.1 Text Password Dictionary Attack

We tested whether the text passwords collected in this study were weaker than those created when participants only had to remember one password at a time, with respect to a dictionary attack using John the Ripper [12]. This open-source software tool uses a supplied dictionary to systematically try to guess passwords. We first tested passwords using the free dictionary of 4 million entries, followed by a second attack using a larger dictionary of 40 million entries purchased from the John the Ripper web site.

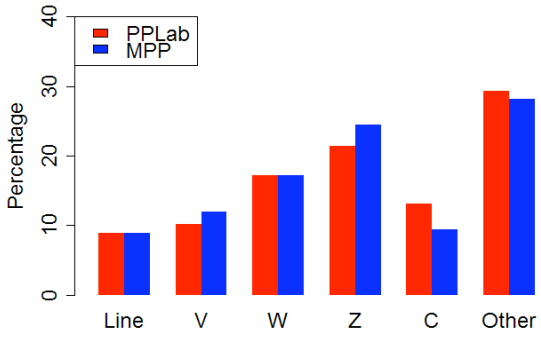

Figure 11: Patterns [8] formed by the click-points of user passwords when connecting all 5 click-points of a password in sequence from the first to the last click-point. PPLab represents an earlier PassPoints lab study [6,8] where participants remembered only one password at a time. MPP represents the current study where participants remembered multiple PassPoints passwords. The differences between the two pattern sets are not statistically significant.

The smaller dictionary cracked 9.8% (20 out of 204) of passwords, while the larger dictionary identified 15.2% (31 out of 204) of passwords. These are lower success rates than we expected since a visual inspection of the passwords revealed that many passwords were quite simple. Examples of passwords that were not cracked by John the Ripper include: "msnhotmail" for an email password, "instantmsg" for an instant messenger account, and "inlibrary" for a library account. In an earlier study of text passwords [16], 9.5% (18 out of 190) of passwords were cracked using John the Ripper with the same 4 million entry dictionary and 18.9% (36 out of 190) of passwords with the larger dictionary.

Participants in the current multiple password study often selected their passwords with some association to the account or with some pattern across passwords. The default John the Ripper dictionaries do not take into account these characteristics. It would seem relatively simple, however, for an attacker to modify the dictionaries to target specific types of accounts. For example, for an attack on bank passwords, attackers could modify the dictionary to include more financial terminology or terms associated with a particular bank.

4.6.2 *PassPoints Hotspot Formation*

To evaluate PassPoints passwords for predictability, we compared the distribution of click-points in the current study to those of an earlier PassPoints study on the same images [6]. We wanted to see whether there was increased clustering of click-points across participants. Clustering occurs when several participants select click-points in the same areas of the image. It is problematic because it signals that there are hotspots: areas of the image where participants are more likely to select click-points. Attackers can gather a small sample of passwords and use this information to predict hotspots and thus guess likely passwords [33]. In the current study, if participants were compensating for having to remember multiple passwords, they may opt to select more "obvious" (and hence likely more memorable) click-points than they would otherwise. We might therefore expect to see more clustering of click-points in the current study than in the earlier PassPoints lab study (where participants only had to remember one password at a time, never revisiting previous images later in the study).

We used spatial statistics to evaluate whether participants who had multiple PassPoints passwords were more likely to select click-points in common areas of the image. Spatial statistics are used in areas such as Earth Sciences to evaluate the spatial distribution of a dataset. The *J-function* [35] measures the level of clustering of points within a dataset. It combines nearest-neighbour calculations with empty space measures for a given radius to measure the clustering of points at that radius. We applied the J-function to the dataset of click-points selected by the 32 PassPoints participants for each image in this study (160 click-points per image). The earlier PassPoints datasets [6] contained between 155 to 220 click-points per image. We used a radius of 9 pixels to approximate the size of the tolerance area (19x19 pixels) used to determine whether a click-point was correct during password re-entry.

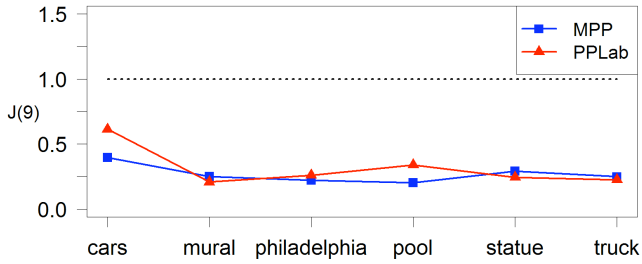

Figure 12: J-statistic at a radius of 9 pixels for the 6 images used in the PassPoints condition. Participants in this study (MPP) were no more likely to select click-points that clustered than participants who only had to recall one password at a time (PPLab).

Figure 12 shows the results of the J-function for all 6 images. It should be noted that these are discrete points and the lines were added only for readability. J approaching 0 indicates that clusters are occurring (J=0 indicates that all data points are in one cluster within radius r). J approaching 1 means that the points are randomly distributed, and J greater than 1 indicates that the points are uniformly distributed (evenly spaced). Ideally, we want our click-point distribution to be as close to J=1 as possible. We see from Figure 12 that PassPoints suffers from clustering click-points. This is a known issue with the original PassPoints scheme [7,13,20,33].

However, the important result here is that participants in the current study were no more likely to select click-points that fell into clusters than those participants who had only one password to remember. Figure 12 shows that the results for the two datasets are very similar.

5. VALIDATION OF HYPOTHESES

We now revisit our hypotheses on multiple password interference based on the results of our study.

1. **Participants will have lower recall success rates with text passwords than with PassPoints passwords.** *Hypothesis partially supported.* Users in the PassPoints condition had significantly higher recall success rates in the short-term. However, after two weeks, participants in both groups had similarly low success rates.

2. **Participants in the Text condition are more likely than PassPoints participants to use patterns across their passwords.** *Hypothesis partially supported.* We found that 53% of participants in the Text condition created passwords following a common pattern across their accounts. In comparison, only two PassPoints participants used the same click-point pattern for most of their accounts. Although it appeared

that Text passwords had more evidence of patterns, we could not perform a direct comparison since the measures were different for the two types of passwords.

3. **Participants will recall text passwords more slowly than PassPoints passwords.** *Hypothesis partially supported.* Overall, participants in the Text condition were slower to recall their passwords than PassPoints participants in the short-term. We believe this result will surprise many in the community. One factor influencing this result is that Text participants made more errors, so spent more time re-entering their password. When considering only successful first attempts, there is no significant difference in password entry times between the Text and PassPoints conditions. After two weeks, Text participants who were successful on the first attempt entered their password more quickly than PassPoints participants. There was no significant difference between the two conditions when the overall time for Recall-2 is considered.

4. **Participants in the Text condition are more likely than PassPoints participants to create passwords that are directly related to their corresponding accounts.** *Hypothesis supported.* We found that 68% of participants in the Text condition used the account name or type as a mnemonic to create a password that was obviously related to its account. We found no evidence that participants in the PassPoints condition created account-related passwords. In fact, it is not clear how one would create account-related passwords in such click-based graphical password schemes.

5. **Participants in the Text condition will make more recall errors than participants in the PassPoints condition.** *Hypothesis supported.* Users in the Text condition made significantly more recall errors than those in the PassPoints condition during both Recall-1 and Recall-2. Participants in the Text condition were more likely to make repeated attempts at entering their password, behaviour which is potentially dangerous since it may reveal other passwords in the presence of key loggers or if on phishing sites. PassPoints participants opted to indicate that they had forgotten their password rather than engaging in this potentially insecure behaviour. Having passwords based on different images dissuaded users from cycling through passwords.

6. DISCUSSION AND LIMITATIONS

Cued-recall is a cognitively simpler task than uncued recall and many users will attempt to turn remembering their password into a cued-recall task. This user strategy often means that text passwords are weaker because they are likely based on some obvious and common cues, such as account or website names.

In the short-term, PassPoints passwords are easier to recall than text passwords when participants have several passwords to remember. We suspect that having a cue helped PassPoints participants remember which password was associated with each account. It appears that participants in the Text condition created mnemonics in order to have an association between an account (or its banner) and its password. For example, many participants with text passwords included the word "email" in the password for their email account. This common strategy appeared to help participants recall their passwords since short-term success rates were considerably higher for those who used account-related passwords (see Section 4.4). However, our results indicate that this strategy was not effective when recalling passwords two weeks later.

Using account-related mnemonics is problematic for security. It may make it easier to guess account-related passwords since the mnemonic remains constant across all participants and is also available to attackers. It can also be problematic for usability when participants are required to change their passwords. A user who changes a password may still associate the account-related mnemonic with the old password and have more difficulty remembering the new password.

Prior to participating in our study, none of our PassPoints participants had previous experience with graphical passwords. However, all our Text participants had much past experience with text passwords. As such, Text participants a substantial advantage when recalling passwords two weeks later, since they had already developed behaviours to help them choose more memorable passwords (see Section 4.4). As there was no difference in two-week recall performance between the two conditions, the PassPoints participants also had an advantage: the image as a recall cue. Note however that the Text participants' use of mnemonics inherently leads to insecure passwords, while PassPoints' image cue does not. Thus, we believe that if a similar cueing mechanism could somehow be added to text password systems, their security could be improved without sacrificing memorability.

The role of the images as cues suggests that images provided by the system should be distinct, otherwise confusion and interference can become a problem. Our strong intuition is that users should not be asked to remember different passwords on the same image, as we suspect that interference problems would be highly likely in that case. Not only is using the same images across systems problematic for usability, it also decreases security since users are likely to re-use passwords. Similarly, a system could allow users to select or upload personal images. However, this would enable password reuse if users upload the same image for several accounts, which has a security cost and may lead to greater memory interference if the exact same password is not selected for each account. Additionally, Davis et al. [11] recommend against allowing user choice in graphical passwords. Click-based graphical passwords can offer a compromise by using system-assigned images and allowing user choice of click-points within the image.

Our intent in this study was to examine the effects of password interference on user behaviour and performance. We followed established psychological methods for clearing working (short-term) memory by administering a distraction task, and adopted a cognitively challenging design to avoid ceiling effects [22]. However, we acknowledge that this lab study does not mirror real-life usage. Users are unlikely to create 6 new passwords one after the other in quick succession in real life, or be asked to recall them all in quick succession after two weeks (without having used any of them in the intervening time). Albeit unusual, circumstances do sometimes require users to create multiple passwords in a short time period, for example, when starting a new job. For the purpose of our comparison, both the text passwords and the PassPoints passwords were created and recalled under similar conditions. Furthermore, although no one in our study wrote down their password, users often do so with their real passwords. This acts as a useful memory aid, but has security risks if anyone gains access to a user's written list of passwords. And lastly, users were told in the instructions to the study that their passwords were for 6 specific accounts. Although this reflects the implicit reality in practice, mentioning this explicitly may have primed users to select more account-specific passwords. Despite these limitations, examining the issue of multiple password interference in a controlled laboratory setting is an important step in understanding the effects of increased memory load and the coping behaviours exhibited by users.

Many passwords created in both the Text and PassPoints conditions were weak. For example, Text participants used account-based passwords and used similar passwords across accounts, while PassPoints participants created passwords that included patterns and formed hotspots across participants. We believe that while PassPoints passwords are often weak, this does not diminish the results of our study, and may be addressed with alternative click-based graphical password systems (see related discussion in Section 2). Future work includes testing for multiple password interference with these alternative schemes and conducting a field study to further examine the effect of multiple password interference in a more ecologically valid setting. To provide more accurate comparison between graphical passwords and text passwords, it may also be necessary to help users become more familiar with graphical passwords beforehand so that their behaviour is more natural and to avoid novelty effects.

7. CONCLUSION

Results of our lab study indicate that in the short-term, PassPoints passwords are more robust than text passwords against multiple password interference (assuming distinct background images). Often, the usability of a system is tested in isolation but in the case of passwords this is especially problematic because user behaviour may change as users accumulate passwords. We show that in the short-term, participants could more easily remember multiple click-based graphical passwords than multiple text passwords. Participants in the Text condition made significantly more recall errors and resorted to additional coping strategies such as using account-based passwords or cycling through all of their passwords when trying to recall their password. We believe the memory cueing provided by click-based graphical passwords is at least part of the reason for better user performance and that cueing should be part of any recall-based authentication scheme, where possible.

After two weeks, recall of the passwords in the text and graphical conditions was not statistically different from each other. Given that users were much more familiar with memorizing and recalling text passwords, they were better positioned to remember these passwords. However, this advantage was offset by the graphical passwords' built-in memory cue, which is a more secure memory aid than users' typical text password coping mechanisms. Our results raise an interesting research question for text passwords: can cueing mechanisms be (safely) added to text passwords in order to achieve the same memorability advantages seen in click-based graphical passwords?

ACKNOWLEDGEMENTS

The authors thank anonymous referees whose comments have helped improve the clarity of this paper. The fourth author is Canada Research Chair in Network and Software Security, and acknowledges NSERC funding of this chair, a Discovery Grant, and a Discovery Accelerator Supplement. The fifth author acknowledges funding of a Discovery Grant through the Natural Sciences and Engineering Research Council of Canada (NSERC). Partial funding from the NSERC Internetworked Systems Security Network (ISSNet) is also acknowledged.

REFERENCES

[1] A. Adams and M. Sasse. Users are not the enemy. *Communication of the ACM*, 42(12):41–46, 1999.

[2] M. Anderson and J. Neely. *Memory. Handbook of Perception and Cognition*, chapter 8: Interference and inhibition in memory retrieval, pages 237–313. Academic Press, 2nd edition, 1996.

[3] G. Blonder. Graphical passwords. United States Patent 5,559,961, 1996.

[4] I. Britton. Freefoto website. `http://www.freefoto`, accessed February 2007.

[5] B. Burstein, L. Bank, and L. Jarvik. Sex differences in cognitive functioning: Evidence, determinants, implications. *Human Development*, 23:289–313, 1980.

[6] S. Chiasson, R. Biddle, and P. C. van Oorschot. A second look at the usability of click-based graphical passwords. In *3rd Symposium on Usable Privacy and Security (SOUPS)*, July 2007.

[7] S. Chiasson, A. Forget, R. Biddle, and P. C. van Oorschot. Influencing users towards better passwords: Persuasive Cued Click-Points. In *BCS-HCI '08: Proceedings of the 22nd British HCI Group Annual Conference on HCI*. British Computer Society, September 2008.

[8] S. Chiasson, A. Forget, R. Biddle, and P. C. van Oorschot. User interface design affects security: Patterns in click-based graphical passwords. *International Journal of Information Security*, 8(5), 2009.

[9] S. Chiasson, P. C. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *15th USENIX Security Symposium*, August 2006.

[10] L. Cranor and S. Garfinkel. *Security and Usability: Designing Systems that People Can Use*. O'Reilly Media, edited collection edition, 2005.

[11] D. Davis, F. Monrose, and M. Reiter. On user choice in graphical password schemes. In *13th USENIX Security Symposium*, August 2004.

[12] S. Designer. John the Ripper password cracker. `http://www.openwall.com/john/`.

[13] A. Dirik, N. Menon, and J. Birget. Modeling user choice in the Passpoints graphical password scheme. In *3rd ACM Conference on Symposium on Usable Privacy and Security (SOUPS)*, July 2007.

[14] K. Everitt, T. Bragin, J. Fogarty, and T. Kohno. A comprehensive study of frequency, interference, and training of multiple graphical passwords. In *ACM Conference on Human Factors in Computing Systems (CHI)*, April 2009.

[15] D. Florencio and C. Herley. A large-scale study of WWW password habits. In *16th ACM International World Wide Web Conference (WWW)*, May 2007.

[16] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle. Improving text passwords through persuasion. In *4th Symposium on Usable Privacy and Security (SOUPS)*, July 2008.

[17] Free Images.com. Free Image website. `http://www.freeimages.com`, accessed February 2008.

[18] S. Gaw and E. Felten. Password management strategies for online accounts. In *2nd Symposium On Usable Privacy and Security (SOUPS)*, July 2006.

[19] E. Goldstein. *Cognitive Psychology*. Wadsworth Publishing, 2006.

[20] K. Golofit. Click passwords under investigation. In *12th European Symposium On Research In Computer Security (ESORICS), Springer LNCS 4734*, September 2007.

[21] S. Komanduri and D. Hutchings. Order and entropy in Picture Passwords. In *Graphics Interface Conference (GI)*, May 2008.

[22] R. S. Lockhart. *The Oxford Handbook of Memory*, chapter 3: Methods of Memory Research, pages 45 – 57. Oxford University Press: New York, NY, 2000.

[23] P. A. Lowe, J. W. Mayfield, and C. R. Reynolds. Gender differences in memory test performance among children and adolescents. *Archives of Clinical Neuropsychology*, 18:865–878, 2003.

[24] S. Madigan. Chapter 3: Picture memory. In J. Yuille, editor, *Imagery, Memory, and Cognition: Essays in Honor of Allan Paivio*, chapter 3. Picture Memory, pages 65–89. Lawrence Erlbaum Associates, 1983.

[25] W. Moncur and G. Leplatre. Pictures at the ATM: Exploring the usability of multiple graphical passwords. In *ACM Conference on Human Factors in Computing Systems (CHI)*, April 2007.

[26] F. Monrose and M. Reiter. Graphical passwords. In L. Cranor and S. Garfinkel, editors, *Security and Usability: Designing Secure Systems That People Can Use*, chapter 9, pages 157–174. O'Reilly, 2005.

[27] PD Photo. PD Photo website. `http://pdphoto.org`, accessed February 2007.

[28] M. Peters. Revised Vandenberg & Kuse Mental Rotations Tests: forms MRT-A to MRT-D. Technical report, Department of Psychology, University of Guelph, 1995.

[29] K. Renaud. Evaluating authentication mechanisms. In L. Cranor and S. Garfinkel, editors, *Security and Usability: Designing Secure Systems That People Can Use*, chapter 6, pages 103–128. O'Reilly Media, 2005.

[30] A. Salehi-Abari, J. Thorpe, and P. C. van Oorschot. On purely automated attacks and click-based graphical passwords. In *24th Annual Computer Security Applications Conference (ACSAC)*, 2008.

[31] L. Standing, J. Conezio, and R. Haber. Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2):73–74, 1970.

[32] X. Suo, Y. Zhu, and G. Owen. Graphical passwords: A survey. In *Annual Computer Security Applications Conference (ACSAC)*, December 2005.

[33] J. Thorpe and P. C. van Oorschot. Human-seeded attacks and exploiting hot-spots in graphical passwords. In *16th USENIX Security Symposium*, August 2007.

[34] E. Tulving and Z. Pearlstone. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5:381–391, 1966.

[35] M. van Lieshout and A. Baddeley. A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, 50(3):344–361, 1996.

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

[37] A. Whitten and J. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *8th USENIX Security Symposium*, Washington, D.C., August 1999.

[38] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. Authentication using graphical passwords: Basic results. In *11th International Conference on Human-Computer Interaction (HCI International)*, 2005.

[39] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon. PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, 63(1-2):102–127, 2005.

[40] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon. Authentication using graphical passwords: Effects of tolerance and image choice. In *1st Symposium on Usable Privacy and Security (SOUPS)*, July 2005.