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

Sonia Chiasson[1,2], **Alain Forget**[1,2], **Elizabeth Stobert**[3], **P.C. van Oorschot**[1], **Robert Biddle**[2]
[1]School of Computer Science, [2]Human Oriented Technology Lab, [3]Department of Psychology
Carleton University, Ottawa, Canada

{chiasson, aforget, paulv}@scs.carleton.ca
estobert@connect.carleton.ca, robert_biddle@carleton.ca

ABSTRACT
People have difficulty remembering multiple passwords. This results in reduced security as users reuse the same password for different systems or reveal other passwords as they try to log in. It can also lead to reduced privacy, as users may rely on centralized services to manage their passwords. In this paper, we report on a laboratory study comparing recall of multiple ordinary text passwords with recall of multiple click-based graphical passwords. 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. We suggest that this is due to memory cues offered by graphical passwords which help users to recall their passwords without resorting to insecure coping strategies.

Author Keywords
passwords, graphical passwords, multiple password interference, usable security, memory cueing

ACM Classification Keywords
H.5.2 [Interfaces and Representation]: User Interfaces Graphical user interfaces; K.6.5 [Computing Milieux]: Security and Protection Authentication.

INTRODUCTION

Special consideration is required to design usable, understandable, and manageable security features. At first glance, it seems like applying standard usability and HCI principles should suffice, but there are security constraints that 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 the usable security of 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 their 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 [13]. 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.

Our current work is motivated by recent proposals for alternative kinds of passwords, particularly click-based graphical passwords [4, 25]. 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. We wished to study whether this approach would have advantages over ordinary text passwords when multiple distinct passwords were necessary, since there is currently little research on this topic. We were particularly 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.

Our study was conducted in a laboratory setting where participants were assigned to use either textual or graphical passwords. They created distinct passwords for several different "accounts", and later had to recall the passwords for each account in a different order than they were created. In the case of graphical passwords, each account was associated with a different image, so users had one image for each password. We found that users had more difficulty recalling multiple text passwords than multiple graphical passwords. We further found that users in the text condition could more easily recall their passwords when they used insecure password practices such as choosing passwords that followed some common pattern or that were obviously as-

*version: September 25, 2008

sociated with the account name. These results constitute evidence for an important advantage inherent in click-based graphical passwords: built-in cueing that helps with memorability. We believe further study in field settings and deeper examination of the mechanisms involved are warranted.

The remainder of the paper is divided as follows. The background section provides background on the type of graphical password system used, memory cueing, and multiple password interference. We then outline the methodology of our study and present the results. Lastly, we offer some discussion and concluding remarks.

BACKGROUND

Security is rarely a user's primary task [23], 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 is also leveraged by 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. This particularly 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 remember or 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 [3, 13]. When trying to log in, they will cycle through their passwords until they find one that works. While this trial-and-error approach helps users to deal with password systems and multiple password interference, revealing all of one's passwords at every login can lead to security vulnerabilities.

Attackers trying to break into user accounts may target one specific user, in which case most systems have a security feature that will lock the account after some small number of incorrect entries. However, another strategy is to try and break into *any* account on the system, in which case the lock out feature does not deter the attacker.

One proposed solution to the password problem is to use password managers. With a password manager, users typically have one master password and the password manager creates, remembers, 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, it has been found that password managers have severe usability problems [9] that can leave users even more vulnerable than when they were managing their passwords themselves.

Figure 1. A PassPoints password consists of 5 ordered click-points on an image.

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 [20]. Proposed schemes include click-based graphical passwords such as PassPoints [25]. 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 [19]. 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 since it 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, 12, 15, 21, 25, 24, 26]. While these have found PassPoints to be generally usable, security concerns have been raised because users tend to select predictable passwords. Newer click-based graphical password schemes, such as Persuasive Cued Click-Points [7], address these concerns.

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. [25] 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 [16] 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.

To our knowledge however, the only evaluation of multiple password interference is as part of a field study of PassPoints [6]. In this study, students logged in to access their class notes for a semester. A sub-group of 30 students were in two of the participating classes and therefore had two passwords to remember. These users had lower login success rates than those with only one password. We are not aware of comparable studies for text passwords, so it is unknown how this performance decrease compares with text passwords.

STUDY DETAILS

We hypothesised that 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 and the PassPoints condition.

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 36 participants (14 males and 22 females). Each participant completed an individual one-hour session. This study used 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 click-based graphical passwords. Participants were primarily university students from various degree programs. All were regular computer and 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 and tolerance areas of 19x19 pixels. This configuration is consistent with previous studies [25, 6]. The images are identified as: Cars, Mural, Philadelphia, Pool, Statue, and Truck (Figures 2 to 7). These were selected from an earlier PassPoints lab study [6].

Figure 2. Cars image [5]

Figure 3. Mural image [26]

Figure 4. Philadelphia image [26]

Figure 5. Pool image [1]

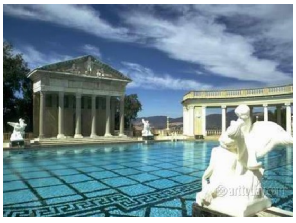

Figure 6. Statue image [2]

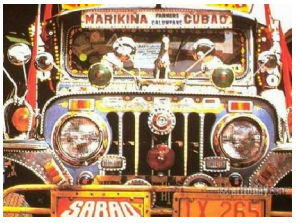

Figure 7. Truck image [2]

Methodology

The 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. As part of the instructions for the 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 "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). For the PassPoints condition, each account was associated with a distinct image, so users never had more than one password per image. In this phase, the accounts were presented to all users in the same order. In the

Table 1. Methodology

Phase	# of trials	Steps
Practice	2 trials This data was not used in the analysis	Create Confirm Answer Questions Distraction Login
Password Generation	6 trials Accounts were presented in the same order for all participants	Create Confirm Answer Questions Distraction Login
Retention	6 trials Account order was shuffled according to the Latin square	Recall

PassPoints condition, the accounts were consistently paired with the same images (although in a real implementation, a system would use different images for different users and offer a new image if a user reset their password). Participants were asked to pick realistic passwords that they could remember but that would be difficult for others to guess. They were further informed that they would need to remember these passwords later and reminded that each password was created for a specific account. In total, 216 account passwords were created: 108 in the Text condition and 108 in the PassPoints condition.

For each password trial, users followed a five-step process:

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).

Confirm: Participants confirmed their password. For text passwords, users entered their password, now echoed only as asterisks. PassPoints users 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 users could not remember their password, they could return to the Create step.

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

Perform Distraction Task: A 30-second distraction task was administered to simulate a longer passage of time. Users in the Text condition were asked to count backwards in 3s from a random 4-digit number, while users in the PassPoints condition completed a Mental Rotations Test (MRT) puzzle [17]. Different tasks were used to clear textual working memory and visual working memory, respectively.

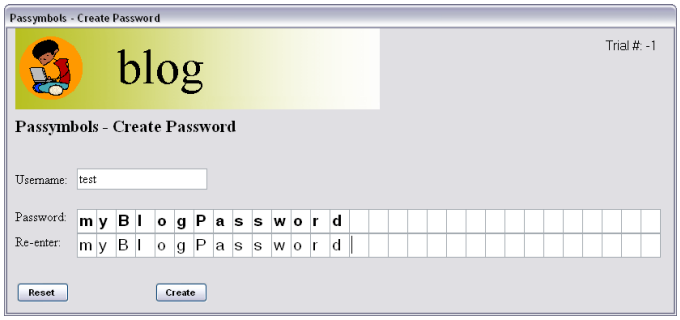

Figure 8. Password creation interface for the Text condition. The "blog" account is used in the Practice phase.

Login: Users 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.

Figure 9. Password creation interface for the PassPoints condition. The "blog" account is used in the Practice phase.

The Retention phase of the study tested whether users could recall each of their 6 passwords. The application prompted users 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 if they made a mistake, and an additional button was available: "I don't remember creating a password for this account". In fact, we never asked users to enter passwords for which they had no account, but they may have forgotten that they created a password for a particular account.

Table 2. Success rates for the Confirm, Login, and Recall steps.

	Password Generation				Retention	
	Confirm		Login		Recall	
	Text	PP	Text	PP	Text	PP
First Attempt	97%	94%	96%	100%	76%	94%
Multiple Attempts	100%	100%	99%	100%	89%	98%

In the Retention phase, the accounts were presented in shuffled order based on a Latin square, where each row represented a participant and the columns represented each account. One hour was sufficient for the generation and later recall of 6 passwords, so we used a 6x6 Latin square, and chose the number of participants to be a multiple of 6. 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 position effects.

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

RESULTS

Success Rates

We first examine success rates as a measure of users' performance. In the Password Generation phase, participants could confirm and login with both Text and PassPoints (PP) passwords equally well (as shown in Table 2). There is no statistically significant difference in success rates between conditions for this phase of the study. These results are similar to previous studies [6, 14] with the same methodology.

However, during the Retention phase, 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. If we consider only the first attempt for each password, users recalled their text password correctly only 76% of the time while those in the PassPoints condition had a 94% success rate (χ^2(1,216)=14.67, $p < .001$). Participants could try to recall their password as many times as they wished, until they either succeeded or gave up. Participants in the Text condition reached an 89% success rate with multiple recall attempts compared to 98% for those in the PassPoints condition. This means that for 11% of trials in the Text condition, participants eventually said they could not recall their password. Only 2% of trials in the PassPoints condition ended with such a failure to recall. The success rates for multiple recall attempts are significantly different (χ^2(1,216)=6.63, $p < .05$), further indicating that users in the PassPoints condition could more easily recall their passwords.

The very high success rates in many categories show that the participants memory was not strongly taxed. However, this makes the much lower success rate for recall in the Text condition all that more remarkable.

Table 3. Recall errors per condition in the 108 trials.

	Text	PP
Number of trials with errors	25	6
Total number of errors	66	14
Number of trials where users gave up	11	2
Number of participants who made errors	8/18	5/18

Table 4. Classification of recall errors for the Text condition.

Type of error	Number of occurrences
Wrong account	31
Wrong account variant	27
Misspelled, variant	8

Recall Errors

We will now focus on the Retention phase and examine the types of errors committed by users as they tried to recall their passwords. As shown in Table 3, users had more difficulty recalling text passwords than PassPoints passwords (t(145.46) = -2.838, $p < .01$). In total, users in the Text condition made 66 errors, while users in the PassPoints condition made 14 errors. For a recall trial, users who were unsuccessful could try to recall the password as many times as they wished, so there could be many more errors than trials.

Each group committed different types of errors. Users in the Text condition appeared to be more affected by interference from multiple passwords. In the Text condition, 8 out of 18 participants made recall errors. They often tried passwords from other accounts when asked to recall a password for a particular account. Many users 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, or some variant. 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. A few errors were due to misspelling or variants of the correct passwords such as entering "access!ble" when the password was "@ccessible". The types of errors for text passwords are summarized in Table 4.

As shown in Table 5, 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, 5 out of 18 participants made recall errors.

Table 5. Classification of recall errors for the PassPoints condition.

Type of error	Number of occurrences
Outside of tolerance area for 1 click-point	1
Incorrect click-point order	1
Forgotten click-points	8
Click-point pattern shifted downwards	4

Timings

Table 6 shows the times taken to complete each password-related step of the study and provides results of t-tests comparing the times from each condition. These times represent the total time spent during a given step. For example, time for login began when the login screen first appeared and continued until the user entered their username, password, and successfully logged in (including any errors committed).

As shown in Table 6, there is no consistent relationship between the two conditions with respect to the amount of time taken at each step. During Password Generation, participants were faster at creating PassPoints passwords than text passwords. The opposite was true for the Confirm phase, where participants could confirm their password more quickly in the Text condition. During Login, participants took approximately the same amount of time to enter their password in either condition.

In the Retention phase, the critical phase in our study, users were quicker at entering PassPoints passwords (t(152.87) = -2.368, $p < .05$). This aligns with the fact that users made fewer errors in the PassPoints condition and hence spent less time repeatedly entering their password. If we consider only recall attempts where users made no mistakes (correctly entered their password on the first attempt), then the difference in entry time is not significant.

Use of Mnemonics

When faced with the task of remembering multiple items, people naturally turn to memory aids, or mnemonics. In our study, users in the PassPoints condition had a built-in mnemonic since they could use the image as a memory aid. We gave no instructions to users 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 13 out of 18 (72%) 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 account. Others were somewhat related, such as "lovelove" for the online dating account. In total, 47% 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.

The use of account-related passwords apparently made it significantly easier for users to recall their passwords, presumably since seeing the account name and banner acted as mnemonics. Participants in the Text condition who used account-related text passwords had a 96% success rate for the Retention phase while those who did not had an 84% recall success rate (χ^2(1,216)=4.14, $p < .05$).

We found no apparent link between passwords created in the PassPoints condition and their associated accounts. Users in the PassPoints condition either did not need an additional mnemonic device since they could already use the password image, or they were unable to find a way to use the account characteristics as memory aids for this type of password. Since recall success rates and timings are better for the PassPoints condition than the Text condition for the Retention phase, this lack of a built-in mnemonic did not appear to hurt the usability of the system. This is advantageous for security because it removes one piece of information that attackers may use to determine likely passwords.

Patterns

We further evaluated whether users were more likely to use some coping strategy, such as selecting predictable passwords, when faced with the task of creating and remembering several passwords. We compared passwords collected in this study with passwords from previous studies that used identical methodology except that users only had to remember one password at a time [6, 14].

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 8 out of 18 (44%) of users in the Text condition created at least one pair of passwords that were similar to each other. In total, 28 out of 108 (26%) of passwords 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.

PassPoints Patterns

We examined whether the PassPoints passwords followed simple patterns. In previous work [8], the types of patterns created by the click-points of user-chosen passwords were classified. It was found that in PassPoints, users were likely to select click-points in simple patterns such as a straight line

Table 6. Timings for each step in seconds and results of t-tests comparing the timings for the two conditions.

Phase	Step	Mean Text	Mean PP	Median Text	Median PP	t-test
Password Generation	Create	34.9	29.2	27.8	24.7	(t(214) = -2.16, $p < .05$)
	Confirm	9.3	12.5	7.5	10.0	(t(214) = 3.30, $p < .01$)
	Login	11.9	12.9	9.5	10.8	not significant
Retention	Recall	28.3	15.8	14.0	10.4	(t(152.87) = -2.368, $p < .05$)
	Recall (correct on 1st attempt)	21.1	14.2	10.0	9.8	not significant

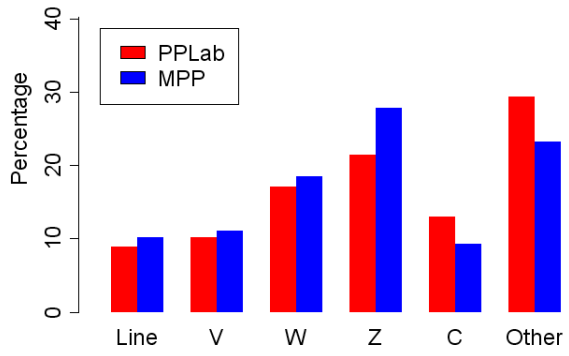

Figure 10. Patterns formed by the click-points of user passwords when connecting all 5 click-points of a password in sequence from the first to the last click-point. PPLab represents our earlier PassPoints lab study where users remembered only one password at a time. MPP represents the current study where users remembered multiple PassPoints passwords.

or C- shape. For a more detailed description of the shape classification scheme, see [8].

We tested the passwords from the current study for patterns, concerned that users 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 10. We found no statistical difference between the patterns found in the current study (where users had to create and remember multiple passwords) and our earlier PassPoints lab study (where users had to remember only one password at a time). So although we did see patterns, they were no more likely to occur than when users had only one graphical password to remember.

We also examined each user's 6 passwords to see if anyone consistently picked passwords in a given pattern. One user had 4 out of 6 passwords following a "Z" pattern, but no other participant used a specific pattern for the majority of their passwords.

Security Measures

Text Password Dictionary Attack

We tested whether the text passwords collected in this study were weaker, with respect to a dictionary attack using John the Ripper [11], than those created when users only had to remember one password at a time. 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 that we purchased from the John the Ripper web site.

The smaller dictionary cracked 10.2% (11 out of 108) of passwords while the larger dictionary identified 11.1% (12 out of 108) 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 [14], 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.

We suspect that there are a few reasons for the discrepancy between the two studies. First, users in the earlier study had to create more passwords and passwords were not associated with any account. These users may have run out of ideas for passwords and thus created simpler passwords. Secondly, users 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 be relatively simple, however, for an attacker to modify the dictionaries to target more specific types of accounts since the attackers would know which accounts they are targeting. For example, for an attack on bank passwords, attackers could modify the dictionary to include more financial type words or terms associated with the particular bank.

PassPoints Hotspot Formation

To evaluate the 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 more clustering of click-points was occurring across users. Clustering occurs when several users 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 users 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 [21]. In the current study, if users were compensating for having to remember multiple passwords, they may opt to select more "obvious" (and hence likely more memorable) click-points then they would otherwise. We might therefore expect to see more clustering of click-points in the current study than in the earlier PassPoints lab

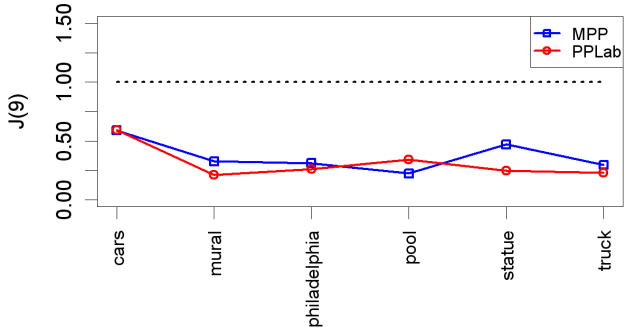

Figure 11. J-statistic at a radius of 9 pixels for the 6 images used in the PassPoints condition. Users were no more likely to select click-points that clustered in this study (MPP) than users who only had to recall one password at a time (PPLab).

study (where users only had to remember one password at a time, never revisiting previous images later in the study).

We used spatial statistics to evaluate whether users 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* [22] measures the level of clustering of points within a dataset. The J-function combines nearest-neighbour calculations with empty space measures for a given radius to measure the clustering of points within the dataset at that radius. We applied the J-function to the dataset of click-points selected by the 18 PassPoints participants for each image in this study (90 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 11 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 11 that PassPoints suffers from clustering click-points. This is a known issue with the original PassPoints scheme [12, 15, 21, 7].

However, the important result here is that users in the current study were no more likely to select click-points that fell into clusters than those users who had only one password to remember. We can see from Figure 11 that the results for the two datasets are very similar and a t-test reveals no significant difference between them.

User Opinion

Users completed questionnaires giving their opinion of the password systems. When we compared the 10-point Likert-scale responses, we found that users felt similarly in both the Text and PassPoints conditions. Table 7 includes the relevant subset of the categories we examined. None of these categories showed significant differences.

Table 7. Questionnaire responses.

	Mean		Median	
Category	Text	PP	Text	PP
Ease of password creation	8.2	6.9	9	6.5
Perceived memorability	5.9	4.4	5.5	5
Perceived security	5.7	6.7	5	6
Ease of login	7.4	8.3	8	8.5

During each trial, participants answered two Likert-scale questions immediately after confirming their password to get their immediate reaction as they created passwords. Participants in the Text condition felt that their passwords were easier to create (χ^2(9,216)=70.14, $p < .01$) and that they would more likely to remember them in a week (χ^2(9,216)=44.65, $p < .01$) than participants in the PassPoints condition. Interestingly, users' perceptions appear to contradict what their performance measures indicate. As users in the Text condition made more errors than those in the PassPoints condition, it seems likely that they would also have more difficulty recalling their password in a week. We suspect that this discrepancy is because users are much more familiar with text passwords.

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 supported.* Users in the PassPoints condition had significantly higher recall success rates when we considered only the first recall attempt and when we considered all recall attempts (as many tries as users wanted until they succeeded or gave up).

2. **Participants in the Text condition are more likely than PassPoints participants to use patterns across their passwords.** *Hypothesis partially supported.* We found that 44% of users in the Text condition created passwords that followed some common pattern across their accounts. In comparison, only one PassPoints participant used the same click-point pattern for the majority of their accounts. Although it appeared that Text passwords had more evidence of patterns, we could not perform a direct comparison between the two conditions 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. One factor influencing this result is that users made more errors, so spent more time re-entering their password in the Text condition. If we consider only successful first attempts, then there is no significant difference in password entry times between the Text and PassPoints conditions.

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 72% of users 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 users in the PassPoints condition created account-related passwords.

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 (66) than those in the PassPoints condition (14). In total, 8 users (44%) made errors in the Text condition as opposed to 5 users (28%) in the PassPoints condition.

DISCUSSION

Based on the number and types of errors seen, it appears that PassPoints passwords are easier to recall than text passwords when users have several passwords to remember. We suspect that having a cue helped PassPoints users remember which password was associated with each account. It appears that users in the Text condition created mnemonics in order to have an association between an account (or its banner) and its password. For example, users with text passwords included the word "email" in the password for their email account. We found this to be quite common and it appeared to be beneficial to users since success rates were considerably higher for those who used account-related passwords.

Using account-related mnemonics is problematic for security. Attackers may leverage this to guess an account-related password since the same mnemonic is also available to attackers. It can also be problematic for usability when users 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.

We suggest that cueing is an important characteristic of authentication schemes, especially if users are expected to remember more than one password. The cues provided by the system should be distinct, otherwise confusion and interference can become a problem. In our PassPoints system, each account had a distinct image. Our strong intuition is that users should not be forced to remember different passwords on the same image, as we suspect that interference problems would be highly likely in that case.

Cued-recall is a cognitively simpler task than uncued recall and many users will attempt to turn the task of remembering their password into a cued-recall task. This user strategy often means that their text password is weaker because it is likely based on some obvious and common cue such as the account or website name.

We emphasize that this is not necessarily an issue of advocating for graphical passwords instead of text passwords. It may be possible to add cueing to text passwords, which may be just as effective. One form of cueing that exists with text include personal verification questions [18], such as those asked when you forget your password and must reset it. However, many current implementations suffer from predictability problems since the set of likely answers is often sufficiently small that an attacker could cycle through them or find the information through some other source. When the questions are more obscure, often legitimate users have difficulty remembering their answers. Mnemonic phrases are also helpful in remembering text passwords [27]. For multiple passwords, we also need cueing because users must first recall their mnemonic phrase which in turn triggers the memory of their password.

Participants in the Text condition cycled through several passwords from their other accounts when they could not immediately recall the correct password for their current account. We have evidence [9] that this happens in previous password studies as well. While this may seem like a reasonable approach to such a situation, this behaviour could make users more vulnerable to attacks. This type of "dangerous success" [9] is problematic in computer security for several reasons. The computer may unknowingly have malware that records keystrokes and sends this information back to attackers, hence revealing a set of potential passwords. Secondly, users logging on to a website may have reached a phishing site that mimics a real site but has the sole purpose of stealing user information. By cycling through passwords, users have now revealed multiple passwords (and possibly usernames) to the malicious owners of the phishing site.

There are a few characteristics of click-based graphical passwords that can be viewed as beneficial from a security perspective but may be viewed as disadvantages for usability. The first is that password reuse is not possible since passwords are based on system-selected images. It is also more difficult (although not impossible) to write down graphical passwords for backup purposes. A system could allow users to select or upload personal images, which would allow for password reuse if users upload the same image for several accounts, but this comes at a security cost and may lead to greater memory interference if the exact same password is not selected for each account.

We acknowledge that this lab study created an artificial scenario. Users are unlikely to create 6 new passwords one after the other in quick succession in real life. However, it is unclear if the effects of interference are more pronounced in such a scenario than in a situation where users create passwords one at a time but then do not use them again for several weeks or months. 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.

Although users in the PassPoints condition performed better than users in the Text condition, PassPoints users created passwords that included patterns and formed hotspots across users. We believe that these could be addressed with alternative click-based graphical password systems. For example, Persuasive Cued Click-Points (PCCP) addresses both these issues in lab testing [7]. Future work includes testing PCCP for multiple password interference and conducting a field study to examine multiple password interference.

CONCLUSION

Results of our lab study indicate that graphical 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. In the current study, we show that users could more easily remember multiple graphical passwords than multiple text passwords. Participants in the graphical condition made significantly fewer recall errors and did not resort to additional coping strategies such as using common patterns across different accounts or cycling through all of their passwords when trying to recall their password. We believe the memory cueing provided by 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.

REFERENCES

1. PD photo website, accessed February 2007.
2. Free image website, accessed February 2008.
3. A. Adams and M. Sasse. Users are not the enemy. *Communication of the ACM*, 42(12):41–46, 1999.
4. G. Blonder. Graphical passwords. United States Patent 5,559,961, 1996.
5. I. Britton. Freefoto website, accessed February 2007.
6. S. Chiasson, R. Biddle, and P. 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. van Oorschot. Influencing users towards better passwords: Persuasive cued click-points. In *Human Computer Interaction (HCI), The British Computer Society*, September 2008.
8. S. Chiasson, A. Forget, R. Biddle, and P. van Oorschot. User interface design affects security: Patterns in click-based graphical passwords. Technical Report TR-08-14, Carleton University, 2008.
9. S. Chiasson, P. 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. S. Designer. John the ripper password cracker.
12. 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.
13. D. Florencio and C. Herley. A large-scale study of WWW password habits. In *16th ACM International World Wide Web Conference (WWW)*, May 2007.
14. A. Forget, S. Chiasson, P. van Oorschot, and R. Biddle. Improving text passwords through persuasion. In *Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS)*, July 2008.
15. K. Golofit. Click passwords under investigation. In *12th European Symposium On Research In Computer Security (ESORICS), LNCS 4734*, September 2007.
16. S. Komanduri and D. Hutchings. Order and entropy in picture passwords. In *Graphics Interface Conference (GI)*, May 2008.
17. M. Peters. Revised vandenberg & kuse mental rotations tests: forms mrt-a to mrt-d. Technical report, Department of Psychology, University of Guelph, 1995.
18. A. Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of facebook. In *4th Symposium on Usable Privacy and Security (SOUPS)*, July 2008.
19. K. Renaud. Evaluating authentication mechanisms. In L. Cranor and S. Garfinkel, editors, *Security and Usability*, chapter 6, pages 103–128. O'Reilly Media, 2005.
20. L. Standing, J. Conezio, and R. Haber. Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2):7374, 1970.
21. J. Thorpe and P. van Oorschot. Human-seeded attacks and exploiting hot-spots in graphical passwords. In *16th USENIX Security Symposium*, August 2007.
22. M. van Lieshout and A. Baddeley. A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, 50(3):344–361, 1996.
23. 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.
24. 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)*, July 2005.
25. 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.

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

27. J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password memorability and security: Empirical results. *IEEE Security & Privacy Magazine*, 2(5):25–31, 2004.