Do You See Your Password?
Applying Recognition to Textual Passwords

Nicholas Wright, Andrew S. Patrick, Robert Biddle
Carleton University
Ottawa, Canada
Email: nick.m.wright@gmail.com

ABSTRACT

Text-based password systems are the authentication mechanism most commonly used on computer systems. Graphical passwords have recently been proposed because the pictorial-superiority effect suggests that people have better memory for images. The most widely advocated graphical password systems are based on recognition rather than recall. This approach is favored because recognition is a more effective manner of retrieval than recall, exhibiting greater accuracy and longevity of material. However, schemes such as these combine *both* the use of graphical images and the use of recognition as a retrieval mechanism. This paper reports on a study that sought to address this confound by exploring the recognition of *text* as a novel means of authentication. We hypothesized that there would be significant differences between text recognition and text recall conditions. Our study, however, showed that the conditions were comparable; we found no significant difference in memorability. Furthermore, text recognition required more time to authenticate successfully.

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, usable security

1. INTRODUCTION

Authentication, in the context of computer security, is the practice of identifying oneself in order to acquire access to information or resources. The vast majority of user authentication is accomplished using text password mechanisms [12] . In text password systems, the user is required to submit a secret password, which only they should know, in order to verify their identity to a computing system. Ideal passwords would be those that are easy for users to remember, simplifying the process of authentication, but difficult for attackers to guess, rendering the system secure [44].

In striving for ideal passwords, we are introduced to the security / usability tradeoff. Strong (secure) passwords are difficult to remember, and passwords that are easy to remember are typically weak.

Systems requiring passwords that are too strong result in frequently forgotten, reset and disclosed passwords, and these systems inspire password reuse and the creation of passwords that are minimally secure. Weak passwords, while easy to remember, are vulnerable to a wide variety of attacks ranging from shoulder surfing to brute force and dictionary attacks, as we describe in the sections below. The intent of this study is to explore potential improvements to text passwords for authentication.

Many of today's current practices related to text password systems are based on recommendations made by credible sources, and industry best practices have evolved based on them [4][18]. These best practices have evolved with little emphasis on human factors research, focusing instead on security. For example, password-based systems often insist that new passwords be composed of at least eight characters, use both upper and lowercase letters, at least one number and one special character, that they be changed frequently, and that they not resemble previously used passwords[16]. Rules such as these are often too difficult for users to observe without disclosing their passwords. Human factors practitioners suggest that security mechanisms be designed to account for human characteristics in an effort to enhance security [36]. Human factors research may be able to inform designs that are sensitive to the limitations of human cognitive ability, while simultaneously acting to increase the level of security.

A large body of work intended to improve the use of knowledge-based authentication systems involves "graphical passwords" [38][11][2]. These are systems where the "password" or secret is not a word at all, but rather a picture, set of pictures, or picture features. Graphical authentication mechanisms are potentially even more secure than alphanumeric text based password systems, while capitalizing on humans' enhanced memory for images.

The proposals for graphical password systems, however, introduce design features that go beyond the use of pictures instead of text. For example, some systems involve cued recall and others involve recognition. By contrast, text-based systems typically involve pure recall alone, with the user entering text in a traditional blank input box. Some text authentication schemes, such as "challenge questions" involve cued recall [28]. Previous studies have shown support for the usability of graphical password systems by comparing cued-recall and recognition-based graphical password systems with the traditional recall-based text password systems [3]. These comparisons involve a confound, however, because the proposed authentication mechanisms benefit from both the pictorial superiority effect and recognition-based retrieval. Text password systems could potentially benefit from the work that has gone into the development of the new recognition-based methods, without resorting to the use of graphical materials. This paper reports on a

study to resolve this confound. We first review related work and underlying theory, then report on two experiments and their results, and present our interpretations and conclusions. Both of the experiments performed as part of this research received approval from our university's Ethics Committee for Psychological Research.

2. BACKGROUND

2.1 Graphical Password Systems

A great deal of research in the area of usable security has focused on the design and implementation of graphical password systems. Surveys of the proposed schemes have been provided by Suo & Zhu [38], Davis, Monrose & Reiter [11] and Biddle, Chiasson, & Van Oorschot [2]. Graphical password systems can be categorized as a drawmetric, locimetric or cognometric mechanism [34].

Drawmetric systems, such as the Draw-a-Secret system [25], require users to compose an initial drawing during the set-up phase, which must then be redrawn later in order to authenticate with a given system. This is therefore a pure recall system, similar in nature to text passwords. Moreover, users may experience difficulty with this graphical password system because their drawing must be reproduced with sufficient accuracy for the mechanism to recognize it as being correct and granting the associated permissions. Further analysis of many users' drawings suggested that people tended to compose symmetric drawings as their secret [39]. In response, Background Draw-a-Secret systems require that users compose and redraw their secret pattern over a background image[15], rendering it increasingly similar to locimetric mechanisms.

Locimetric graphical password schemes are a cued-recall-based method of authentication whereby the system presents users with an image, and locations on the image are selected and recorded for authentication. The initialization phase requests that users select several points on the image, the set of which becomes that user's "password" which must be recalled for future authentication.

Wiedenbeck, Waters, Birget, Brodskiy & Memon [43] developed the most widely known locimetric graphical password system, *PassPoints*. The usability of this system was established as sufficient for practical deployment [5]. However, Thorpe & Van Oorschot [39], discovered that the distribution of chosen click-points for a particular image was far from random, instead centering on "hotspots" – places of increased likelihood of selection. Persuasive Cued Click-Points (PCCP) has since been proposed as a modification of the original PassPoints scheme, forcing users to select their click-points from smaller random areas of the image [6]. This adaptation appears to have remedied the hotspot issue while preserving previous login success rates.

Cognometric graphical password systems function by presenting a series of panels of images to the person requesting access, asking them to choose the single correct image on each panel. These recognition-based schemes involve the selection of specific images for password entry. These images are learned upon registration with the system and are plainly presented to the user for recognition at login time. Among existing cognometric schemes, *PassFaces* [32] is the most commercialized and studied example. All of the images used in *PassFaces* are of peoples' faces. In the original system, users selected their chosen faces, displayed along with 8 distractor images per panel, over four panels.

These systems have been lauded as highly usable [14]. User studies have shown that users can remember these types of passwords well, and for long periods of time [3][31]. However, the PassFaces example demonstrates relatively weak security, comparable to that of a four-digit bank card PIN (see section 1.2.2 for a discussion of password strength).

As was found with some locimetric schemes, Davis et al. [11] identified a weakness regarding the images chosen by users to compose their password in PassFaces. There was a strong tendency to select attractive faces, especially those from one's own race, and this increases the likelihood that an attacker could guess a user's password. The bias inherent in user selection of password images can be guarded against, by assigning users' password images. This is the approach now taken by the commercial PassFaces system. Assigning the faces that comprise the password raises the possibility of reduced memorability, but a recent study suggests the approach is still usable [23].

Cognometric graphical passwords have been suggested as the most promising innovation in knowledge-based authentication because they leverage recognition rather than recall [33]. Since human memory is better able to recognize previously encountered information than it is at recalling material without cues [29], we speculate that this principle can be used in the design of improved text-based authentication mechanisms.

2.2 Recognition Versus Recall

There are three principle mechanisms for accessing information previously acquired. In *recognition*, information is presented to the individual, who then must make a judgment about whether or not the information is familiar or not. When material is not present to be recognized, it must be recalled from memory. *Recall* can take place with or without the presence of cues. *Cued recall* is a method of retrieving information from memory with the help of a cue, which acts as a hint, aiding the search through memory. Pure, free, or uncued recall is the retrieval of items from memory without any help from the surrounding environment [10]. Providing a cue to an item in memory increases the likelihood of successful recall, and speeds the rate of recall. Uncued recall is the most difficult manner in which known material is retrieved for use [40].

Tulving's encoding specificity principle stated that retrieval was dependent upon the combination of material stored in memory and certain cues that would facilitate its availability. Cues could be very general, such as "enter your password," or more specific, as is the case with "what is your mother's maiden name?" In either case, successful retrieval relies on the extent to which the cue reinstates the manner in which the information was stored.

There were two early hypotheses on the relation between recognition and recall. First came the threshold-sensitivity hypothesis, which stated that recognition was much like recall, only easier (requiring a lower threshold). Research later discredited this hypothesis because it implied a constant positive correlation between recognition and recall, with recognition always easier then recall.

The second main hypothesis, known as the generate-plus-recognize hypothesis, posited that recall is similar to recognition, but with an extra step. Recall was said to require individuals to generate a set of items that could possibly contain the one sought after and then a recognition decision would be made among the items in the set, whereas in recognition the generation phase could

be skipped because the item was presented to the individual. According to the Generate-Recognize model, which developed upon this hypothesis [42], the cue restricts the set of possibilities through which someone would have to search to determine the answer. This model is strongly aligned with the theory of encoding specificity.

More recent work [20][35][37] has determined differences between declarative, explicit or conscious memory and non-declarative, implicit or unconscious memory. This distinction is strongly supported by studies involving amnesic participants, who have impairments in their abilities to recognize, recall and learn new material, but who are perfectly capable in tests of priming, conditioning and skill learning.

The processing of cues can benefit from a phenomenon known as perceptual priming, a process through which detecting and identifying material is facilitated by recent encounters with that same material [41]. Thus the ability to recognize words or objects depends not only on a conscious assessment of the material, but also on "increased perceptual fluency" or priming [19][24][26]. Therefore, recognition capitalizes on encoding in both declarative and non-declarative memory, while recall is limited to declarative memory alone.

Knowledge of these models of memory and retrieval processes leads us to an understanding that recognition and recall situations are handled in different manners by people asked to remember material. The majority of tests performed on memory indicate that our faculty for recognition consistently produces more effective and persistent results than those of recall. Therefore, it stands to reason that authentication methods capable of taking advantage of our enhanced ability to recognize information are more memorable, and thus more usable, than traditional text passwords relying solely on pure recall.

2.3 Password Strength

There are several manners of attack that motivated parties can use in an attempt to thwart password authentication mechanisms and gain access to restricted information or services. Excluding attackers' manipulation of software vulnerabilities to circumvent the authentication phase altogether, attacks are generally classified as capture or guessing-based attempts to determine actual passwords.

Capture related attacks are those that require interception of the password during entry, or deceiving a user into divulging the secret under false pretense. These include shoulder surfing, reconstruction, malware, phishing, and social engineering methods. Guessing-based attacks involve performing numerous attempts of potentially educated guesses at generating a password to gain access to the protected system. These attacks may be performed systematically, guessing every possible password in what are known as "brute-force" attacks. More refined guessing attacks limit the attempts to "words" found on discrete lists, or "dictionaries". Dictionary attacks can be optimized, for example, guessing more likely words first, and potentially omitting words that are unlikely to be used.

This study does not address capture attacks. Moreover this study does not address ordered dictionary attacks, because the passwords to be used are random and assigned, which ensures maximum entropy and resistance to these attacks, as all possibilities are equally likely. Previous studies of graphical recognition based passwords have shown that user-chosen graphical passwords are so vulnerable to dictionary attacks that user choice is expressly advised against [11]. All password-based schemes are still subject to brute force attacks. Because of this, care must be taken to ensure that each authentication scheme in a study is equally resistant to them, designed with equal strength.

The strength of any password system lies with its associated password space. The larger the potential variety of password combinations (or password space) available for use, the more guesses will have to be attempted before there is success, and thus the more secure the system can be.

The theoretical password space of a system is the set of all possible unique combinations allowed by that system. Theoretical password space can be calculated, and this is elaborated on and demonstrated in the discussion that follows. However, a more meaningful evaluation of the password space associated with the security mechanism in question will reference its *effective* password space. The effective password space is the set of all possible password combinations that people may actually use, within the theoretical password space. For example, in text password systems people are almost certainly not going to choose "XzalCH49fQi5" due to its complexity. As previously mentioned users of the original PassFaces system tended to choose attractive faces of people sharing their race, and users of the Draw-a-Secret system tended to draw symmetric patterns to use during authentication. Weaknesses such as these allow attackers to narrow their password dictionaries in guessing-based attack methods, rendering their attacks increasingly successful. Following the change made to PassFaces, we suggest that recognition-based password systems should not allow user choice, but rather work with random and assigned passwords. In this way, the effective password space will be the same as the theoretical password space, and ensure consistent strength against brute force attacks. In designing our study, we must therefore ensure that the password space will be the same in any conditions to be compared.

3. STUDY DESIGN

We wished to investigate the documented disparity between the effectiveness of recognition and recall, in the context of text-based password systems. By comparing participants' abilities to recall text-based passwords and their ability to recognize words as passwords, we were able to assess practical limitations of memory and the implications on user authentication. Furthermore, preserving a constant password space across all three conditions increased the validity of our observations.

When people have the opportunity to choose their passwords, they tend to create passwords that are as simple as possible [17], since those are most easily remembered. It is not clear if people are able to effectively remember passwords that are assigned to them, especially when the memory task is recognition. To summarize, the research question is: Can a recognition-based text password system facilitate authentication to a greater extent than traditional text passwords?

In considering password schemes to compare, we first identified as a control condition a scheme where users were assigned random passwords of six lower-case letters. The space for this condition will therefore be $\log_2(26^6)$ or 28 bits.

To explore a text-recognition scheme, we considered recognition of words in the same way PassFaces uses recognition of faces. The user would have to recognize a set of several words, each

word being displayed on screen amongst a set of distracter words. To match the password space of our control condition, users would be asked to recognize one word from a display of 26 total words, and to do this 6 times in a row. The space for this condition will therefore also be $\log_2(26^6)$ or 28 bits.

To explore another alternative, we also decided to consider recall of whole words. In this scheme, users would remember a list of 4 whole words, which will serve as one password. These 4 words are taken from a set of 156 possible words, and so its password space is calculated as $\log_2(156^4)$, or 29 bits, similar to the other conditions. We acknowledge an issue related to the password space associated with this condition, which is outlined in greater detail below.

4. EXPERIMENT 1

Our first experiment used a within-subjects design consisting of three experimental conditions. Using a within-subjects design controls for individual differences, and permitted the use of statistically stronger hypothesis tests. Each condition required participants to employ a different text-based authentication mechanism to log in and interact with web sites set up specifically for use in this experiment. The first password type was a traditional six-character random text password, composed of lowercase alphabetical characters (the **letter recall** condition). The second consisted of four assigned whole words, which when entered at the prompt served as one password (the **word recall** condition). The final condition was a cognometric graphical password system, except that rather than displaying a set of pictures for participants to select from in order to authenticate, they were shown panels of whole words which could be clicked in series to authenticate (the **word recognition** condition). Each of these conditions was implemented using 28-bit strength.

The word and letter recall conditions both represented authentication conditions involving pure recall, and the word recognition condition presented an opportunity to capitalize on recognition as a retrieval mechanism. Both the word recall and word recognition conditions possessed the potential advantage of using whole words in passwords. This should allow users to process their passwords more deeply when attempting to memorize them because of the semantic meanings associated with words, which the letter recall condition does not allow [9].

All passwords assigned in this study were randomly generated for each participant. The type of authentication used served as the independent variable (IV) in each of the planned analyses. To evaluate the hypotheses stated below, we needed to measure the length of time each type of password was remembered by each participant as the dependent variable (DV, Maximum Memory Time) by recording the amount of time between password creation (or reset) and the last successful login. If there were no resets, then the memory time would be the time between the beginning and end of participation, which was one week. We also measured the number of resets requested per participant per condition (Resets). Login efficiency was also measured as a DV, which was recorded as the time taken for each participant to authenticate successfully (Login Time). Lastly, the number of passwords that persist in memory for the duration of the study was recorded as a DV (Remembered Passwords).

4.1 Hypotheses

After having reviewed the theory behind the effectiveness of recognition and recall, and making some interpretation based on the data supporting graphical passwords, we were prepared to identify some hypotheses regarding the outcome of this experiment. Because recognition judgments have been described as more effective over lengthy periods of time, it is believed that the word recognition condition will result in significantly more memorable passwords than the two conditions relying on the process of pure recall, as stated in hypothesis one:

$H1_1$: There will be significantly greater memorability in the word recognition authentication system when compared to the recall-based (text entry required) authentication mechanisms, measured according to maximum memory time.

For our second and third hypotheses, there was less certainty associated with each of the authentication methods. The three authentication methods are very different, creating a situation where novelty may play a role and the time required to type (or identify) the passwords will cause an unknown influence. Because of this, the second and third hypotheses are non-directional.

$H2_1$: There will be a significant difference in the number of password resets initiated for each type of authentication.

$H3_1$: There will be a significant difference in time required to log in across authentication types.

In an attempt to measure the simple effectiveness of each authentication mechanism, we compared the number of passwords that are remembered correctly at the end of their participation. As with hypothesis one, it was expected that recognition would be superior to recall, as outlined below:

$H4_1$: There will be a significantly greater number of passwords remembered for the duration of the study in the Word Recognition condition than in either of the two recall related conditions.

4.2 Method

4.2.1 Participants

Participants recruited for this experiment were individuals who made regular use of the Internet and web sites that require authentication. Participation was restricted to those who do not have any serious visual or memory related impairments that may have affected the outcome of this investigation. The participants for this study consisted mainly of university students, and young adults who were compensated for their time either in the form of twenty dollars, or two bonus percentage points in the undergraduate Psychology course in which they were enrolled.

4.2.2 Materials

To administer this experiment we created three websites that required authentication in order to view and contribute content. Automated reminders were sent to our participants at regular intervals, asking them to log in by entering their three assigned passwords at the prompts as visible in Figures 1 and 2 below.

The set of words used to generate the passwords we assigned in the word recognition and word recall conditions were selected from Ogden's Basic and International word lists [1]. Our words were chosen from Ogden's lists in order to create a selection of words that is representative of daily language.

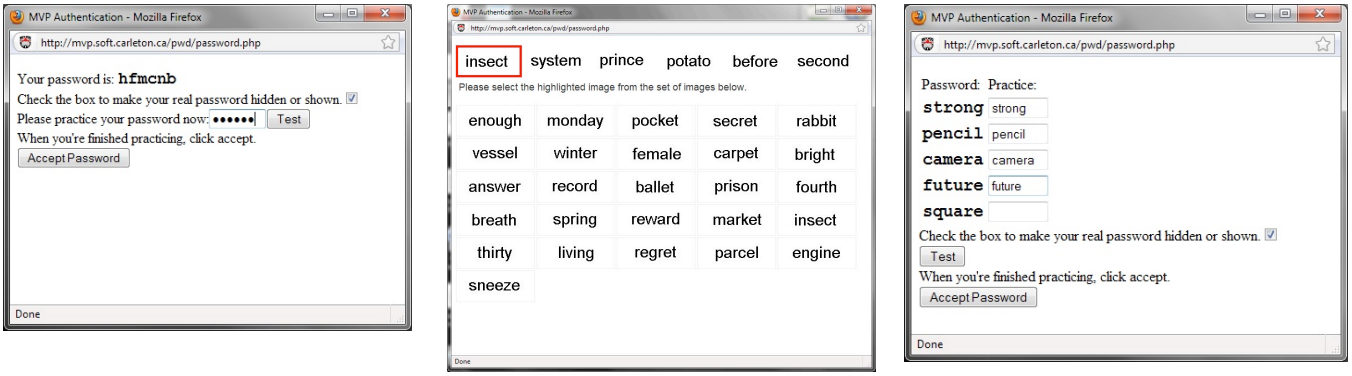

Figure 1a: Examples of the *registration* screens for the Letter Recall, Recognition and Word Recall conditions, respectively.

Figure 1b: Examples of the *login* screens for Letter Recall, Recognition and Word Recall conditions.

By creating passwords from words widely recognized as central to an understanding of the English language, we hoped to ensure that all participants were familiar with them, further controlling for individual differences among participants. The list we chose was 156 words long, so resulting in a space of $\log_2(156^4)$ or 29 bits. We acknowledge there is a limitation to this approach because people know many more words, but we considered this condition an exploratory addition to the study.

4.2.3 Apparatus

Participants used personal computers (PCs) with Internet access in our lab and at home to authenticate at each of the three sites created for use in this study.

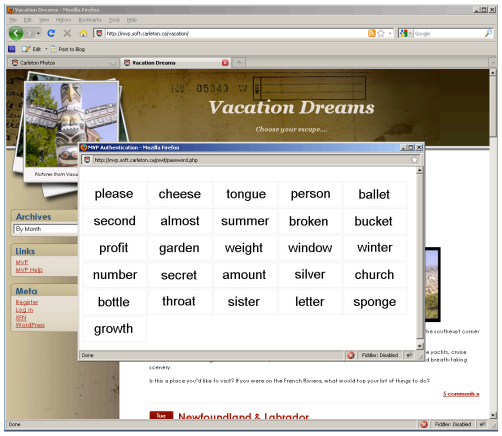

Figure 2: Example of one of the websites, showing the *login* screen for the recognition condition.

The three websites had each been outfitted with a password protection scheme according to our conditions of interest. In our lab they used PCs operating with Microsoft Windows 7 and browsed the Internet using Internet Explorer. Between lab sessions, the participants were free to use whatever combination of computer and Internet browser they preferred. The MVP authentication framework was used to facilitate account management and automated participant reminders [8]. In this framework, users access realistic and distinct websites and complete tasks that require authentication; only the authentication schemes differ. Figure 2 shows a sample site during the login process.

4.2.4 Procedure

Phase 1: Participants arrived at the lab at a time previously agreed upon. At arrival, participants were given an explanation of the experiment, and told that they would be able to withdraw from the experiment at any time without penalty. A consent form was then provided for them to read and sign, if they agreed, before the experiment commenced.

After providing their consent, the participants were shown to a computer and given a simple introductory questionnaire, which gathered demographic information. The participants were then given their first password and asked to sign-in to a web site, and then sign out. They were then shown to the other two sites, given those passwords and asked to learn them and authenticate. They were then asked to authenticate to each of those same sites again, to demonstrate that the password had been memorized. If they were unable to reproduce their password and log in successfully, they were shown their password again and encouraged to login until they could do so without requiring any help, to ensure that

they had memorized their passwords before the lab session was concluded. Participants were asked not to write down their passwords.

Note that each password was of a different type, and each site was distinct. Counter balancing was employed in this study to control for order effects so that the participants did not all see the same schemes in the same order. This served to protect against the possibility of order effects influencing the data across each authentication condition.

Bringing participants into the lab also allowed us to gather additional qualitative observations, and to ensure a reasonable amount of care was taken in learning the different passwords. Once all three passwords had been memorized, an appointment was scheduled for the second lab session approximately one week later. They were then encouraged to login from home and told to expect two notification e-mails, and the session was then concluded.

Phase 2: Over the period of one week, while at home, participants received two reminder e-mails asking that they log into each of the three sites for which they were assigned passwords and to contribute to the content of the site. The participants were free to do this on any computer that they could access. All actions taken with regards to authentication were logged on the web servers, so the time required to log in, number of successful as well as failed attempts, and the time and number of password resets were recorded for later analysis.

Phase 3: At the second scheduled appointment, which took place at an agreed upon time approximately one week after the first, the participants arrived at the lab and were greeted by a researcher. They were then shown to a computer, asked to log into each of the sites and add a written entry to each of the web sites one last time. When they had finished, they were given a questionnaire related to the password schemes and their experiences throughout the study. Upon submitting the questionnaire, they were handed the debriefing form and encouraged to ask any questions or voice any concerns they may have had. They were then compensated and thanked for their time, and their participation in the experiment was then concluded.

4.3 Results

4.3.1 Participants

We included 36 participants in this study. The results generated from 8 and 17 were omitted as they never successfully completed phase 1, and participants 37 and 38 were recruited in their place. Participant 1 was unable to complete the initial survey due to a power outage on the date of their appointment, but all other data originating from this participant was valid, and included in our results, so our dataset consists of a full week of observations on 36 participants.

The 36 participants comprised 15 males and 21 females, with a mean of 29.8 years of age. Twenty-five of them had a social science related background and seven reported a natural science or engineering related background, with the remaining four participants choosing not to disclose. Participants showed an average of 3.87 years of post-secondary education. Twenty-nine (or 80.6%) of respondents had English as a first language, and the other seven (or 19.4%) spoke English as a second language. When asked to rate their computer skills on a scale from 1 to 10, where one meant "novice" and ten meant "expert", this sample's mean was 7.08, the median response was 8, and one person rated themself a 3, which was the lowest response. The vast majority of participants (91.7%) use the Internet daily, and the others all reported using it several times per week. Lastly, while everyone participated in the two lab sessions, participation dropped by about one third for the first attempt from home, and roughly half of the participants made an attempt after receiving the second e-mail notification.

4.3.2 Hypothesis One

$H1_1$: There will be a significantly greater memorability in the word recognition authentication system when compared to the recall-based (text entry required) authentication mechanisms, measured according to maximum memory time.

To test this hypothesis, the authentication mechanism served as the independent variable, and maximum memory time was the dependent variable. The maximum possible value for memory time is about 200 hours, because participants were enlisted for a period ranging from six to eight days. The data is shown as boxplots in Figure 3., where the dark horizontal lines indicate the medians, the box indicates the central quartiles, and the whiskers the outer quartiles; circles identify outliers.

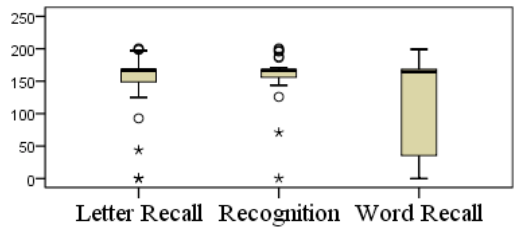

Figure 3: Boxplots of memory persistence (hours) in each authentication condition

We then assessed normality: skewness measurements of -1.93, -2.37, and -0.83, all with S.E of 0.393; kurtosis measurements of 4.53, 10.20 and -1.09, with S.E. of 0.768, across the letter recall, recognition and word recall conditions respectively. The skewness confirmed non-normality. The non-parametric repeated-measured Friedman's test was therefore used, rather than repeated-measures ANOVA.

Friedman's test showed no significant differences between the conditions in this study ($\chi 2 = 1.167$, $p = 0.558$). Therefore recognition did not prove significantly more memorable in this case. Because there was no significant difference identified, no further tests were conducted. Therefore no support for hypothesis one was found. As a non-parametric test based on ordinality Friedman's test does not address skewness. The obvious skewness of the distribution of the word recall condition is quite striking in this case, and indicates that a larger number of people had difficulty remembering their passwords for a comparable period of time.

4.3.3 Hypothesis Two

$H2_1$: There will be a significant difference in the number of password resets initiated for each type of authentication.

Again, authentication mechanism served as the independent variable, and the mean number of password resets served as the dependent variable.

The observed results for password resets did not meet the assumptions of normality (skewness measurements of 3.15, 6, and 1.69, all with S.E of 0.393; kurtosis measurements of 8.37, 36 and 2.16, with S.E. of 0.768, across the letter recall, recognition and word recall conditions respectively. Participants were allowed to reset their passwords as many times as they chose throughout their participation, however, nobody reset their passwords more than twice per authentication condition. Figure 4 displays the distributions of reset attempts by condition.

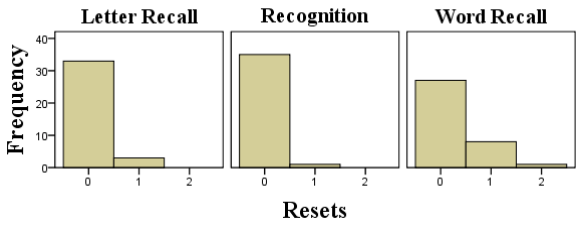

Figure 4: Histograms of reset frequency per authentication condition

Friedman's test verified the presence of a significant difference in number of password resets requested between the authentication conditions ($\chi 2 = 9.455$, $p = 0.009$). We then conducted post hoc analysis using Wilcoxon paired tests. After applying Bonferroni corrections, the recognition condition had significantly fewer resets than the word recall condition ($Z = 2.496$, $p = 0.039$), however the difference between the letter recall and word recall conditions was not significant ($Z = 2.111$, $p = 0.105$), and the difference between the recognition and letter recall conditions also showed no significance ($Z = -1.000$, $p = 0.951$). Hypothesis two has been supported by these findings.

4.3.4 Hypothesis Three

$H3_1$: There will be a significant difference in time required to log in across authentication types.

Authentication mechanism was used as the independent variable and the average login time was the dependent variable in this analysis.

The data is shown in Figure 5, and again the skewness and kurtosis measurements revealed that the distributions were not normal.

Friedman's test was significant in this case and deserving of further investigation ($\chi 2 = 60.743$, $p < 0.001$). We continued to post-hoc analysis using three Wilcoxon paired tests. All differences were significant in this evaluation. After applying Bonferroni corrections, the recognition condition was significantly different from the letter recall condition ($Z = -5.160$, $p < 0.001$) and the word recall condition ($Z = -4.769$, $p < 0.001$), and there was also a significant difference between the letter and word recall conditions ($Z = -4.845$, $p < 0.001$).

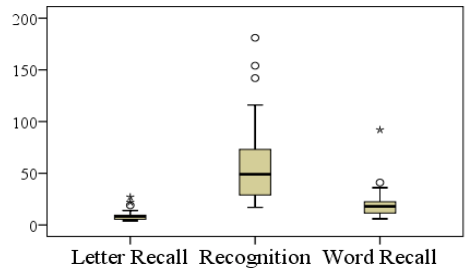

Figure 5: Boxplots of performance: login time (seconds) per authentication condition.

Therefore the letter recall authentication mechanism handily outperformed the other two, in terms of time required to login successfully. These differences support hypothesis three.

4.3.5 Hypothesis Four

$H4_1$: There will be a significant difference in the number of passwords remembered for the duration of the study, across each of the experimental conditions.

To test for a significant difference in number of persistent passwords amongst the three conditions, a Chi-squared analysis was performed. The authentication mechanism used served as the independent variable and the number of passwords used successfully throughout the experiment served as the dependent variable in this case. Figure 6 displays the number of participants who remembered their passwords for each condition throughout the duration of the study, and as well as counts of the passwords that were not remembered throughout.

The chi-squared test indicated a significant difference in participants' abilities to remember their passwords for the duration of the experiment ($\chi 2 = 10.717$, $p = 0.005$). Next, three chi-square tests were performed to explore the differences between each pair of conditions.

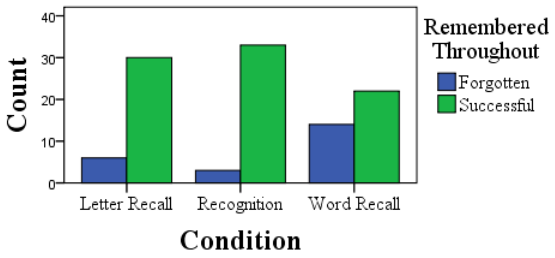

Figure 6: Bar graph of passwords remembered throughout the study per condition

Having performed the three chi-squared post-hoc tests and applying Bonferroni corrections, we found no significant difference between the letter and word recall conditions ($\chi 2 = 4.431$, $p = 0.105$), nor the recognition and letter recall conditions ($\chi 2 = 1.143$, $p = 0.885$), though the recognition and word recall conditions showed a significant difference ($\chi 2 = 9.318$, $p = 0.006$). Hypothesis four is supported by these results.

4.3.6 Summary of Hypothesis Tests

Reflecting upon the results of these hypothesis tests, there is a clearer picture of the influence of recognition on password

usability. The outcome of hypothesis one revealed that there were no significant differences between authentication conditions in terms of maximum memory time.

When considering the number of password resets that occurred in each condition (per hypothesis 2), the recognition condition performed significantly better than the word recall condition, though not significantly better than the letter recall condition. The number of passwords that were remembered for the duration of the study is another important usability metric for which we could track the results. The fourth hypothesis revealed findings very similar to those of the second hypothesis, where the recognition condition enhances the ability of people to remember the passwords they were assigned. The recognition condition produced significantly more passwords remembered for the study's duration than the word recall condition, though not significantly different from the letter recall condition.

The third hypothesis test investigated a usability metric regarding performance time. Among the three authentication conditions, the recognition condition performed most poorly, requiring by far the greatest amount of time for participants to login successfully. Overall differences in this test were significant, with the letter recall condition permitting faster login times than the other two conditions.

The three memorability related metrics tested revealed either no significant difference, or simply a difference between recognition and word recall, but not letter recall. The time required to log in was significantly different for all pairings, and notably worst in the recognition condition. Further differences were investigated and reported in the analyses of interest or questionnaire results sections, below.

4.3.7 Influence of demographic factors

During the initial lab session, immediately following the informed consent form, we administered a participant information survey that captured various demographic factors. Pursuant to the analyses of interest mentioned earlier, we looked for distinctions between participant categories, including age, gender, first language (English vs. other), field of study (social science related, or natural science and engineering), and years of post secondary education, with outcome variables, namely: resets, time to login, maximum memory time, and passwords remembered for the duration of the study.

These calculations revealed a significant correlation between age and login time ($r = 0.362$, $p < 0.001$). Further investigation then revealed that age was moderately correlated with login time in the context of the letter recall condition ($r = 0.494$, $p = 0.002$), only weakly correlated in terms of the recognition condition ($r = 0.289$, $p = 0.002$) and not significantly with the word recall condition ($r = 0.253$, $p = 0.143$) This bodes well for the recognition condition, hinting that it may be less susceptible to the effects of aging than a mechanism that capitalizes on recall alone. It is however important to recall that the recognition condition required the greatest amount of time to authenticate successfully, by far.

That said, there were no other significant distinctions between the remaining demographic variables (gender, first language, field of study or years of post-secondary education) and any of the outcome variables (password resets, login time, maximum memory time and passwords remembered for the duration of the study).

5. EXPERIMENT 2

Interference between passwords may play a role in the ability of participants to successfully retrieve their passwords from memory and submit them to the intended web sites in order to authenticate. Because interference may represent an important aspect of a password system's usability, a secondary investigation was conducted. This study involved assigning each participant three of the same type of password (either letter recall or recognition types), for use in the same three websites as those in the main study, and monitoring their authentication attempts for potential instances of interference. Our focus was on the potential of the novel recognition condition and that of the letter recall condition, which is most similar to present-day passwords, and word recall type passwords were omitted from this analysis.

This investigation used a between-subjects design consisting of two experimental conditions. Participants in one condition were assigned three letter recall type passwords, and participants in the other were assigned three recognition type passwords. Both conditions were used in the same three websites that were employed in experiment 1. Assigning participants three of the same types of passwords allowed us to observe evidence of password interference across the participants' three web sites.

The authentication condition again served as the independent variable (IV) in this evaluation. To investigate our hypothesis we needed to review the access logs of each web site and count the instances of interference, which served as the dependent variable (DV). For the purposes of this study we defined an instance of password interference as an event where at least half of the contents of a password for one website are submitted as credentials in another (including passwords that may have been reset, and were previously valid). Instances of password interference were counted on each of the three websites, and mean interference was calculated for each participant.

5.1 Research Hypothesis

Recognition judgments have been described as more effective over lengthy periods of time than those based solely on recall. The recognition condition also presents users with *cues* to the content of their passwords, and limits the users' password selections to a small subset of potential words. It is believed that the word recognition condition would result in significantly fewer instances of password interference than the letter recall condition, which relies on the process of pure recall, as stated in hypothesis one:

$H1_1$: There will be significantly fewer instances of password interference witnessed in the word recognition authentication system when compared to the letter-recall based authentication mechanism.

5.2 Method

The investigation for this iteration of our study was conducted very similarly to the manner in which experiment 1 was carried out, albeit with a few key differences that are highlighted in the paragraphs that follow.

5.2.1 Participants

Participants for this experiment were recruited in the same manner as those for the initial experiment. As in experiment 1, we recruited people without significant visual impairment or other conditions, which could have impeded performance and hence comparability, in this study.

5.2.2 Materials

To administer this experiment we made use of the same three websites as in the earlier investigation. Automated reminders were sent to our participants at the same intervals of two and four days following recruitment, which asked them to visit each of the three websites and authenticate using the passwords they were assigned.

5.2.3 Apparatus

This experiment made use of the same apparatus as was used and outlined for experiment 1. Participants used personal computers with Microsoft Windows and Internet Explorer when in the lab to learn their passwords and fill out the questionnaires. When browsing from outside the lab setting, participants were free to use any computer with an Internet connection. Password assignment, training and logs were managed via the MVP password framework [8].

5.2.4 Procedure

Phase 1: In this experiment, the procedure was largely identical to that of experiment 1. Participants arrived at the lab, were given an explanation of the study, an informed consent form, and a participant information survey, which asked for demographic related data. The participants were then assigned their passwords, shown to the sites and given opportunity to practice. The only difference in this phase was that participants in this study were not asked to fill out a pre-test questionnaire.

Phase 2: As in experiment 1, participants were asked to log into the sites between lab sessions and were sent two notification e-mails to this end at intervals of two and four days from their initial lab session. The web sites kept logs of all authentication related activity during this time, as well as the two lab sessions, for later analysis.

Phase 3: Upon arriving at the lab, participants were greeted and asked to log into each site one last time. If they failed to do so, they were asked to make at least two attempts, without resetting their passwords. Participants were then debriefed and compensated, but were not asked to complete a post-task questionnaire.

5.3 Results

5.3.1 Participants

We recruited 20 participants for the investigation on password interference, and all of them completed the first session. In this secondary investigation of password interference, participants 1 and 20 were omitted due to corrupt data, and participant 6 was eliminated for a complete lack of participation. Hence eight people were assigned to the letter recall condition, and nine to the word recognition condition. No additional participants were recruited for this study.

Among the participants in this investigation, the average age was slightly older than 25 years, ten of whom were male and seven female. Eleven of them reported social science related backgrounds, while six declared natural science or engineering related fields of study, and there was an average of 3.88 years of post secondary studies in this sample. Twelve people spoke English as a first language, and for five it was a second language. When asked to rate their computer skills on a scale from 1 to 10 where 1 meant "novice" and 10 meant "expert", the average response was 7.53, and 8 was the median answer; Nobody rated their skills less than 5, and all of them reported using the Internet daily.

5.3.2 Hypothesis One

Upon completion of experiment 1, it became clear that the resulting observations did not produce a normal distribution, as seen in Figure 7. This violation of the assumption of normality dictated that we investigate any significant difference among the groups using a Mann-Whitney U test.

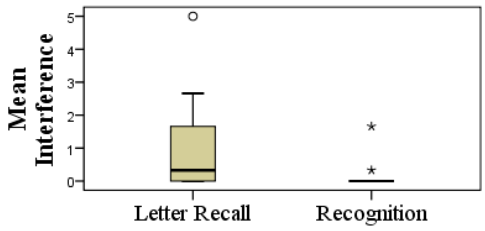

Figure 7: Boxplots of mean password interference

While Figure 7 appeared to suggest there was a difference in means between the two authentication conditions in terms of password interference, it was not found to be significant ($U = 20.500$, $p = 0.094$). This seems surprising because the recognition-based authentication condition is in fact more resistant to interference by design. In all of the password schemes in this study, passwords are assigned and must be entered in the correct order. In either of the recall mechanisms participants are free to enter whatever text they choose into any of the text boxes displayed to them, however, in the recognition condition participants are forced to choose one word at a time from a set of 26 shown per panel, making the possibility for complete password interference extremely unlikely. This experiment showed no evidence for reduced password interference in the recognition condition, however this issue may be deserve further study.

6. DISCUSSION

The three password systems used in this study each made use of text-only passwords. Passwords of the first type were composed of six randomly generated letters that were to be typed in succession in order to authenticate. Among the three conditions used in this study, this one most closely approximated passwords in use today, and involves a pure recall situation. The second password type consisted of four randomly assigned whole words, which the user had to type out to submit. This also presented the user with a pure recall situation, but with greater potential for semantic meaning in their content. The third password scheme was a cognometric password mechanism, except that where graphical images would normally be used in sets of tiles displayed in several panels, pictures of words were displayed instead. This preserved the recognition aspect of a cognometric password system, while eschewing the potentially confounding use of images.

This study was able to focus on the usability of the password mechanisms themselves because the security level was held constant across the three conditions. That is to say that in each case, the complete set of potential passwords was the same size. The password space was 28 bits in strength, which is acceptable in some text-based password systems in the real world. Recently, passwords of just 21-bit strength were suggested to be sufficient for protection against the majority of attacks online [18]. Many

text password systems in use today have theoretical password spaces that are much larger than the space allotted in this study, however, they allow for user chosen passwords, and these same users will often create passwords that only meet a system's minimum requirements, resulting in passwords of comparable strength to the security level used in this study or weaker.

When considering participant behaviour, it appeared that participants exerted comparable amounts of effort in each authentication condition. There were no significant differences in login attempts overall, or successful attempts across conditions. Practice didn't make perfect in this study either, as the number of login attempts per successful authentication was not significantly different between conditions, or over time.

Upon concluding their involvement in the study, participants were asked if they had written their passwords down. Four of them acknowledged that they had, although it seemed that the majority had only done so as a form of practice during the learning phase. Additionally, it appeared that writing down passwords was more commonly used as a strategy for this in the recognition and word recall conditions, which used whole words.

6.1 On Word Recall

Perhaps the most surprising finding from among the memorability results is just how poorly participants fared in the word recall authentication condition. The letter recall condition served as a pure recall form of authentication and the recognition condition allowed a comparison of recognition and pure recall. The word recall scheme, also a pure recall scheme, was included in the study because it potentially represented an enhanced form of recall over the letter recall condition.

Requiring participants to recall whole words as a password, we anticipated would allow for increased semantic meaning of the password content and thus more thorough encoding process, but the significantly higher mean number of resets and fewer number of passwords remembered for the duration of the study highlights the fact that this was not the case. The notable skew in the maximum memory time distribution of the word recall condition indicates that while many were able to remember that type of password for a considerable period of time, many people had a great deal of trouble remembering that password for a meaningful duration, and this stood in contrast to the distributions associated with the other two conditions for the test of that hypothesis.

This condition presented an additional source of error in the misspelling of words; however, we observed that spelling mistakes were not a common source of error.

6.2 The Effect of Recognition

The recognition condition required significantly longer for its users to login than that of either recall based condition. The time to login in the word recall condition was roughly twice that of the letter recall condition, which seems appropriate, given that participants had to type four times as many characters, however the time to submit a recognition-based password was more than double that of word recall. When the sizeable differences in performance times across conditions was revealed, the login times were investigated using only data from the second lab session. This would allow for the maximum amount of repetition and less chance for distraction, but these same results were mirrored in that analysis.

Considering memorability, the recognition and letter recall conditions produced similar results, and both of them outperformed the word recall condition. There was no significant difference in maximum memory time between the three conditions. There was a significant difference between the recognition and the word recall conditions in terms of password resets, and there was also a significant difference between recognition and word recall when considering the number of passwords remembered for the duration of the study.

It was surprising that recognition as retrieval mechanism showed no greater memorability than the recall-based random letter scheme. The two factors that bolster the idea of using cognometric graphical passwords as a successor to text-based passwords were the pictorial-superiority effect, and recognition rather than recall as retrieval mechanism. Our study eliminated the possible confound by omitting the use of pictures from the experimental conditions.

Recognition failed to produce any enhancement in the memorability of text-based passwords in this experiment. Participation in this study was limited to eight days, which may have caused the observed ceiling effects in the recognition and letter-recall conditions. However, due to the similarity in performance of these two conditions, it is also possible that no difference in memorability would materialize if a longer period of study were permitted. The comparable results produced by the Recognition and Letter-Recall conditions across each test suggests that perhaps it is the use of pictures alone that may improve the usability of password systems.

6.3 Means of Improvement

The first possible explanation rests with the distinctiveness of the words chosen. A similar study was conducted last year [23] comparing the usability of different kinds of images in cognometric graphical password schemes. In that research, the distinctiveness of each image in the password set was suggested to be the principal factor in the learning and retrieval of passwords. The concept of distinctiveness implies that unique perceptual properties enhance memorability [22]. Furthermore, in the context of words, a distinctive orthographic to phonologic mapping has also been demonstrated to enhance memorability [21]. Capitalizing on distinctiveness could enhance either of the whole-word based conditions in this study.

The influence of distraction available in this scenario is one source of concern. Participants may have seen a word with a unique meaning to them and begun to think about that for a moment. Alternatively, the volume of words in front of them may have been too great, causing the participant to attempt to recall their password instead, and then search laboriously for the word they had recalled.

Reducing the number of distractor words per panel, and increasing the number of panels per password is worthy of some attention in order to maximize the speed per panel, while preserving a desirable level of security. For example, presenting two words per panel would result in extremely fast processing of the panels, but would likely require processing a prohibitive number of panels in order to authenticate.

The great length of time required to login in the recognition condition seems to suggest that although recognition decisions can in general be made more quickly than attempts at recall, the user is making a sufficient number of recognition decisions to make

the process slower than pure recall, hindering the system's usability.

We had anticipated that the recognition condition would prove significantly more memorable than the two recall based conditions, but that difference did not materialize. In the test of maximum memory time, a ceiling effect was witnessed among the distributions. Quite a few participants remembered their passwords for more than six days. It is possible that a longer duration of participation time in this study would have allowed for greater differentiation between the conditions and results more closely resembling a normal distribution, but this is unclear. Normally distributed results would have facilitated the use of statistically stronger tests, however, a greater period of time between lab sessions may not have differentiated the letter recall from the recognition condition, because they performed so similarly in this experiment.

The words used in our recognition condition were shown in different positions every time an authentication attempt was made. The same words were always used on the same panels, but those words were always shuffled. Assigning passwords maximized entropy in order to combat guessing or brute-force attacks, and shuffling is implemented in cognometric password schemes as a preventative measure to counter capture-based attacks such as shoulder surfing. In this investigation, shuffling also ensured a pure recognition scenario. However, if we were less concerned with shoulder surfing or more sophisticated capture attacks, we could do away with the shuffling of the paneled words.

This shuffling may indeed have been the element that caused the recognition condition to demand additional effort, resulting in its burdensome login times. Eliminating the shuffling may have enhanced the memorability of the passwords, allowing users to recognize the whole panel of words, and first *recall* the general position of their words, and then allowing recognition of the sought after words themselves. Over time people could have even developed a spatial memory regarding the positions of their words, or muscle memory for the motion of their hands positioning the cursor. This may have improved upon the terribly poor login times associated with this condition. The authentication condition would no longer be an example of pure recognition, but this would not be of concern to the user. The added ability to recall information about the panels would combine the two retrieval mechanisms and possibly result in the creation of a more usable authentication mechanism.

6.4 Password Interference

In the context of passwords, memorability is a large part of their usability. As the number of passwords we use grows, individual passwords may become less memorable. Novel password schemes may present themselves as more memorable because of an artificial lack of any similar passwords in memory. The mistaken submission of a valid password from one application into another system would result in an unsuccessful login attempt. This is referred to as password interference, and is a failure to retrieve the correct password from memory.

A great deal of research has been conducted on the role of interference in memory, and some recent work in the specific context of authentication [7]. While text based passwords have existed for decades, the present study is the first to substitute words into a cognometric password mechanism. Since interference can severely hamper the usability of text based password schemes it is important to investigate the potential for interference in the proposed recognition-based password system.

Experiment 2 was conducted specifically to address the possibility of password interference in the novel recognition condition by comparing interference observations in that condition to observations in the letter recall condition. The data revealed no significant difference between conditions. However, The low number of participants in this study may have been a key limitation, and this result may merit further investigation. Participant observations indicated there were a greater number of occurrences of interference in the letter recall condition, however, complete failure to remember the appropriate words or letters was the bigger issue for either condition. When learning their passwords, upon realizing that a word or two in their new password was also used in one of their other passwords, some seemed relieved, as though it would simplify the learning process, and others immediately realized that this would add to the difficulty. An interference evaluation of the word recall condition was not done, which would have been interesting, however given these findings it is unlikely there would have been a significant difference there.

6.5 Participant Observations

Individual differences among the participants were controlled for in the within-subjects design of the first experiment. We had anticipated that the recognition condition would fare best in the memorability related tests. However, we acknowledge that while some people may prefer a recognition scenario, others may perform best in a situation involving recall. Since the recognition scheme did not improve usability uniformly, perhaps the process of authenticating could be improved by accommodating the preferences of each user. This might possibly be accomplished by allowing users to choose which type of authentication they would like to use on each site or service they use.

Participants were observed using several strategies to aid in the memorization and retrieval of their passwords. In the letter recall condition, participants tried to pronounce their random letter passwords as whole words, and commented that these passwords were easiest to remember when there were vowels present, making them easier to sound-out. Jung [27] found that meaningful syllables were easier to remember than non-meaningful syllables. Participants were thus attempting to ascribe meaning to their meaningless random letter passwords.

People like to use whole words in their passwords, and so it was expected that the word recall authentication mechanism would produce better results than were observed. Using whole words allows people to group the letters composing their passwords and capitalize on the phenomenon known as memory chunking [30]. While chunking is put to use to group the letters of a known word, participants must not have been able to treat their sets of four words as groups or chunks of words. For example they may have created sentences composed of their four words. The significant difference in password resets, however, leads us to believe that chunking did not help participants learn or remember their word recall based passwords. Instead, it was observed that some people tried to remember the first letters of each of their words, which offered little help in any scenario. Indeed, many of the mistakes made included words with the same first letter. We consider the deeper issue to be that recall was from the large set of words people know, rather than the limited set of basic words we used. Of course, showing those words would transform the task into

recognition, but an alternative approach might be to limit the words to a well known but limited set, such as names of the months: but finding a suitably sized and widely known set might prove challenging.

In the word recall and recognition conditions some participants tried to compose sentences or stories from their password set. Participants commented that when their recognition password sets included verbs they were easier to remember, and likewise when words "belonged together" (The words "tongue" and "throat" were given as one example). So, similar semantic meaning of words therefore played a role in the memorability of these passwords, which is consistent with previous findings [13]. Interestingly, the ratio of verb words to the set of all words in the selection set was much lower than the ratio of vowels to the size of the alphabet, which may have made the whole word passwords more difficult to group. It therefore may be possible to better facilitate this strategy for password memorization.

7. CONCLUSION

Text-based passwords can be difficult to remember and use, especially given the increasing number of passwords we are expected to initialize and use regularly, and the wide variety of security policies to which we must adhere. Cognometric graphical password systems have been lauded because they may capitalize on both the pictorial superiority effect, and recognition as a retrieval process, which is regarded as superior to recall. While the former cannot be applied to text-based passwords, we sought out to discover whether or not recognition alone could enhance the usability of text-based password mechanisms.

This study sought to determine if the use of recognition in text-based authentication systems could improve their usability. The experiment involved assigning three different types of passwords to participants to use on three websites that we could monitor, for a period of one week. One form of password consisted of 6 randomly generated lower case letters, and another type consisted of four randomly generated whole words. Both of these mechanisms use recall as retrieval mechanism, with the difference being that the whole word condition would involve a great deal more semantic information, ideally simplifying retrieval of the password. The third password mechanism closely resembled a cognometric graphical password system with 26 images per panel, except that in our case the images were simply pictures of words. With the password space held constant across conditions, we were then able to compare the system based on several usability metrics, and through our participants' feedback.

No significant differences were observed in maximum memory time across conditions. The recognition condition produced significantly fewer password resets than the word recall condition, as did the letter recall condition. In terms of the number of passwords that were remembered for the duration of the study, the recognition condition performed comparably to the letter recall condition, and significantly outperformed the word recall condition, though the letter recall condition did not. The surprising weakness of the recognition condition was the time required to login. All differences were significant in this test, where the letter recall condition performed best, followed by the word recall condition and then the recognition condition. Even if implementing recognition resulted in more memorable passwords, in this present form it would also make them more time-consuming to submit.

We also conducted a secondary study to investigate the potential problem of password interference in this recognition-based text password mechanism. Participants were either assigned three letter-recall type passwords, which most closely resemble passwords in use today, or three of the recognition-based passwords. No significant difference in the occurrence of password interference was observed.

The impetus for this study was the confound created in the comparison of the usability of cognometric graphical password systems to that of traditional text-based password systems. Previous studies which incorporated this confound have demonstrated increased usability in graphical password mechanisms relative to text-based authentication. In light of this, and considering that we failed to support the influence of recognition rather than recall as retrieval mechanism as the factor enhancing usability, perhaps it may be the use of pictures rather than text that potentially renders graphical password mechanisms more effective than the traditional text password system. Further study of this issue is necessary.

This study was subject to two obvious limitations. While we strove for ecological validity in the design of this experiment, the period of time for which we could monitor password use was limited to one week, and the motivation to remember the passwords or reset them when forgotten was not critical. For greater ecological validity, a longer-term study would be desirable. The first experiment involved 36 participants and the second involved 17. Sample size is thus a potential factor in the lack of significance in some of the hypothesis tests, and this limitation could be addressed by conducting studies with greater numbers of participants in future.

All passwords used in this study were randomly generated and assigned to the participants; This prevents the use and disclosure of previously known passwords and ensures equal password spaces in each condition, however, it may also limit the applicability of these findings to password systems that allow user choice.

An alternative interpretation may relate to the distinctiveness of words comprising a password set, as distinctiveness of words may lead to enhanced memorability. The shuffling of the paneled words in our recognition condition ensured that this was a pure recognition scheme, however it seemed to add undue difficulty to the mechanism. A scheme allowing users to capitalize on all manners of memory retrieval may in fact represent an optimally usable authentication mechanism.

Having completed this study, there are a few investigations that could naturally follow. The issues of study duration and sample size have been addressed. A direct comparison between the recognition condition used in this study with one using a graphical cognometric scheme of the same password space, to contrast the use of words with pictures would also be very interesting. This study's recognition condition suffered from terribly long login times. Some investigation adjusting the number of images per panel, and the number of panels per password seems like another investigation of our cognitive abilities that would be appropriate for password usability. Finally, in order to create a satisfactory interface for users, recognition might be implemented in such a way that it would not hinder people from logging in as fast as they can with their current passwords. It may be that an approach allowing for both recognition and recall would be ideal.

8. REFERENCES

[1] Bauer, J. L. 2008. Ogden's Basic English. Retrieved November, 2010, from http://ogden.basic-english.org/

[2] Biddle, R., Chiasson, S., & van Oorschot, P. C. in press. Graphical Passwords: Learning from the First Twelve Years. ACM Computing Surveys.

[3] Brostoff, S. & Sasse, M. A. 2000. Are PassFaces More Usable Than Passwords? A Field Trial Investigation. British Human-Computer Interaction Conference (HCI), September 2000.

[4] Burr, W. E., Dodson, D. F., Polk, W. T., Evans, D. L. 2004. Electronic Authentication Guideline, in NIST Special Publication 800-63.

[5] Chiasson, S., Biddle, R. & van Oorschot, P. C. 2007. A Second Look at the Usability of Click-Based Graphical Passwords. Symposium on Usable Privacy and Security (SOUPS), Pittsburgh, PA, U.S.A.

[6] Chiasson, S., Forget, A., Biddle, R., & Van Oorschot, P. C. 2008. Influencing users toward better passwords: Persuasive Cued Click-Points. Human Computer Interaction (HCI), the British Computer Society, September 2008.

[7] Chiasson, S., Forget, A., Stobert, E., Biddle, R., & Van Oorschot, P. C. 2009. Multiple password interference in text and click-based graphical passwords. ACM Computer and Communications Security (CCS), Chicago, USA, Nov. 2009.

[8] Chiasson, S., Deschamps, C., Stobert, E., Hlywa, M., Freitas Machado, B., Forget, A., Wright, N., Chan, G., & Biddle, R. 2012 The MVP Web-based Authentication Framework, Financial Cryptography and Data Security, Springer.

[9] Craik, F.I.M. & Lockhart, R.S. 1972. Levels of processing. A framework for memory research. Journal of Verbal Learning and Verbal Behaviour (11), 671 – 684.

[10] Crowder, R. G. 1976. Principles of Learning and Memory. New Jersey: Lawrence Erlbaum Associates.

[11] Davis, D., Monrose, F., & Reiter, M. K. 2004. On User Choice in Graphical Password Schemes. Proceedings of the 13th USENIX Security Symposium, 151-164.

[12] De Angeli, A., Coventry, L., Johnson, G. & Renaud, K. 2005. Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems. International Journal of Human-Computer Studies (63), 128 – 152.

[13] Deese, J. 1959. Influence of inter-item associative strength upon immediate free recall. Psychological Reports (5), 305-312.

[14] Dhamija, R., & Perrig, A. 2000. Déjà vu: A user study using Images for Authentication. Proceedings of the 9th Conference on USENIX Security Symposium, 9, 4-4.

[15] Dunphy, P. & Yan, J. 2007. Do Background Images improve "Draw a Secret" Graphical Passwords? Proceedings of the ACM conference on computer and communications security. pp. 36 – 47.

[16] Federal Information Processing Standards Publication (FIPS) 1985. FIPS 112: Password Usage, National Institute of Standards and Technology, http://www.itl.nist.gov/fipspubs/fip112.htm. Accessed Jan, 2011.

[17] Florencio, D. & Herley, C. 2007. A Large-Scale Study of Web Password Habits. WWW: Banff, AB. Canada.

[18] Florencio, D. & Herley, C. 2010. Where do security policies come from? Proceedings of the Sixth Symposium on Usable Privacy and Security. ACM: Washington.

[19] Gardiner, J.M. 1988. Functional aspects of recollective experience. Memory and Cognition. 16(4), 309 – 313.

[20] Hintzman, D.L. 1990 Human learning and memory: Connections and Dissociations. Annual Review of Psychology. 41, 109–139.

[21] Hirshman, E., & Jackson, E. 1997. Distinctive perceptual processing and memory. Journal of Memory and Language, 36(1), 2-12.

[22] Hunt, R. R, & Elliot, J. M. 1980. The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. Journal of Experimental Psychology: General, 109(1), 49-74.

[23] Hlywa, M.A.X., Patrick, A.S, Biddle, R. 2011. Do houses have faces? The effect of image type in recognition-based graphical passwords. Annual Computer Security Applications Conference (ACSAC 2012).

[24] Jacoby, L. L. 1983. Perceptual enhancement: Persistent effects of an experience. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 21-3.

[25] Jermyn, I., Mayer, A., Monrose, F., Reiter, M. K., & Rubin, A. D. 1999. The design and analysis of graphical passwords. Proceedings of the 8th conference on USENIX Security Symposium, p.1-1, Washington, D.C.

[26] Johnston, W.A., Dark, V.J., & Jacoby, L.L. 1985. Perceptual fluency and recognition judgments. Journal of Experimental Psychology: Learning, Memory and Cognition. 11(1), 3 – 11.

[27] Jung, J. 1968. Verbal Learning. New York: Holt, Rinehart & Winston.

[28] Just, M. & Aspinall, D. 2009 Personal Choice and Challenge Questions: A Security and Usability Assessment, in Proceedings of the 5th ACM Symposium on Usable Privacy and Security (SOUPS).

[29] Kausler, D. H. 1974. Psychology of Verbal Learning and Memory. New York: Academic Press.

[30] Miller, G.A. 1956, The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review, 63, 81-97.

[31] Moncur, W. & LePlâtre, G. 2007. Pictures at the ATM: Exploring the usability of multiple graphical passwords. Human Factors in Computing Systems (CHI). San Jose, California, USA.

[32] Passfaces Corporation, "The science behind PassFaces," http://www.passfaces.com/enterprise/resources/white_papers.htm, accessed December 2010.

[33] Renaud, K. 2009. Guidelines for Designing Graphical Authentication Interfaces. International Journal of Computer Security (IJICS). 3(1), 60 – 85.

[34] Renaud, K. & De Angeli, A. 2009. Visual Passwords: Cure all or snake oil? Communications of the ACM. 52(12), 135 – 140.

[35] Richardson-Klavehn, A., & Bjork, R.A. 1988 Measures of memory. Annual Review of Psychology. 39, 475 – 543.

[36] Sasse, M.A., Brostoff, S. & Weirich, D. 2001. Transforming the 'Weakest Link' – A Human/Computer Interaction Approach to Usable and Effective Security. BT Technology Journal.

[37] Schacter, D. L. 1987. Implicit memory: History and current status. Journal of Experimental Psychology: Learning. Memory, and Cognition.13. 501-518.

[38] Suo, X. & Zhu, Y. 2005. Graphical Passwords: a survey. Proceedings of the 21st Annual Computer Security Applications Conference. pp. 463-472.

[39] Thorpe, J., & Van Oorschot, P. C. 2007. Human seeded attacks and exploiting hot-spots in graphical passwords, in 16th USENIX Security Symposium, August 2007.
[40] Tulving, E., & Pearlstone, Z. 1966. Availability vs. accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior. 5, 381 – 391.
[41] Tulving, E., & Schacter, D. L. 1990. Priming and human memory systems. Science, 247, 301-396.
[42] Watkins, M. & Gardiner, J. M. 1979. An appreciation of the generate-recognize theory of recall. Journal of Verbal Learning and Verbal Behavior, 18, 687–704.
[43] Wiedenbeck S., Waters, J., Birget, JC., Brodskiy, A., & Memon, N. 2005 PassPoints: Design and longitudinal evaluation of a graphical password system. International Journal of Human-Computer Studies. 63(1-2), 102 – 127.
[44] Yan, J., Blackwell, A., Anderson, R., Grant, A. 2005. The Memorability and Security of Passwords. In L. F. Cranor & S. Garfinkel (Eds.), Security and Usability: designing secure systems that people can use (pp. 129-142). Sebastopol, CA: O'Reillly.