Leveraging Real-Life Facts to Make Random Passwords More Memorable

Mahdi Nasrullah Al-Ameen[1], Kanis Fatema[1], Matthew Wright[1], and Shannon Scielzo[2]

[1] Department of Computer Science and Engineering
[2] Department of Psychology
The University of Texas at Arlington
Arlington, TX, USA

Abstract. User-chosen passwords fail to provide adequate security. System-assigned random passwords are more secure but suffer from memorability problems. We argue that the system should remove this burden from users by assisting with the memorization of randomly assigned passwords. To meet this need, we aim to apply the scientific understanding of long-term memory. In particular, we examine the efficacy of augmenting a system-assigned password scheme based on textual recognition by providing users with *verbal cues*—real-life facts corresponding to the assigned keywords. In addition, we explore the usability gain of including images related to the keywords along with the verbal cues. We conducted a multi-session in-lab user study with 52 participants, where each participant was assigned three different passwords, each representing one study condition. Our results show that the textual recognition-based scheme offering verbal cues had a significantly higher login success rate (94%) as compared to the control condition, i.e., textual recognition without verbal cues (61%). The comparison between textual and graphical recognition reveals that when users were provided with verbal cues, adding images did not significantly improve the login success rate, but it did lead to faster recognition of the assigned keywords. We believe that our findings make an important contribution to understanding the extent to which different types of cues impact the usability of system-assigned passwords.

Keywords: Usable security; system-assigned passwords; memorability; verbal cues

1 Introduction

Traditional user-chosen textual passwords suffer from security problems because of password reuse and predictable patterns [12,37]. Users are tasked with creating a password that should be both secure and memorable, but they typically lack information about what is secure in the face of modern cracking and attacks tools, as well as how to construct memorable strings, memorize them quickly, and accurately recall them later. Faced with this challenge, users often compromise on security and create a weak but memorable password. While policies have been deployed to get users to create stronger passwords [18, 37], such policies do not necessarily lead to more secure passwords but do adversely affect memorability [32, 37].

Studies in psychology have shown that recognition, such as identifying an assigned picture from a set, is an easier memory task than recall, such as traditional textual passwords [5, 41, 42]. Inspired by these findings, researchers have proposed and examined recognition-based authentication schemes as alternatives to pure recall-based schemes in hopes that by reducing the memory burden on users, more secure passwords can be generated. Wright et al. [44] implemented the concept of recognition for a text-based scheme, where users are shown several portfolios of keywords (e.g., "Cheetah", "Mango", "Camera", etc.), and one keyword per portfolio serves as the authentication secret that they have to recognize during login. Passfaces [1] is an example of a graphical recognition-based scheme, which is now commercially available and deployed by a number of large websites.[3]

To ensure security, the commercial Passfaces [1] product assigns a random image for each portfolio instead of allowing users to choose. With system-assigned passwords, the user does not have to guess whether a password is secure, and the system can ensure that all passwords offer the desired level of security. Additionally, while password reuse could pose a serious security threat [12], using system-assigned passwords ensures that users do not reuse a password (or modification thereof) already used on another account. Unfortunately, it is difficult for most people to memorize system-assigned passwords for both textual [44] and graphical recognition [16]. Thus, it still remains a critical challenge to design an authentication scheme that offers satisfactory memorability for system-assigned random passwords.

1.1 Contributions

To this end, we draw upon several prominent theories of cognitive psychology to enhance the memorability of system-assigned recognition-based passwords. In particular, we examine the impact of offering *verbal cues*, i.e., real-life facts related to the system-assigned keywords. For example, "Cheetah is faster than any other land animal" is a verbal cue for the keyword "Cheetah". The use of cues facilitates a detailed encoding that helps to transfer the authentication information (e.g., assigned keywords) from the working memory to long-term memory at registration [6], helping users recognize their keywords when logging in later. We provide a detailed discussion on these memorization processes in §3.

The study of Wright et al. [44] found insufficient memorability for textual recognition, where the keywords in a portfolio remained same but were shown at different positions each time that portfolio was loaded. The authors anticipated that showing the keywords in the same position each time would improve the

[3] http://www.realuser.com/ shows testimonials about Passfaces from customers.

memorability for recognition-based schemes and suggested the approach to be examined in future work. We adopt suggestion of Wright et al. [44] to design our study conditions by showing the keywords in a portfolio in the same position each time that portfolio is loaded. We also accommodate the *variant response* feature in our schemes to gain resilience against observation attacks like shoulder surfing (see §3.5 for details).

To examine the impact of verbal cues in improving the memorability for textual recognition, we design a scheme, *TextV*: **Text**ual Recognition with **V**erbal cues, and compare it with the *Control* condition that requires users remembering the assigned keywords without the help of verbal cue. In addition, we aim to understand whether adding images related to the keywords contributes to higher memorability than when users are provided with just verbal cues. To achieve the goal, we design another scheme, *GraphicV*: **Graphic**al Recognition with **V**erbal cues, and compare it with the TextV scheme. To the best of our knowledge, no study yet has compared textual and graphical recognition-based schemes in terms of usability.

In our within-group study with 52 participants, every participant was assigned three different passwords, each representing one study condition. The major findings from our study include:

- In contrast to the suggestion of Wright et al. [44], keeping the position of keywords fixed in a portfolio did not provide a satisfactory login success rate (61.5%).
- Verbal cues made a significant contribution in improving the login success rate for textual recognition (94.2%).
- Despite the *picture superiority effect* (see §3), we found no significant difference between textual and graphical recognition in terms of login success rate when both conditions included verbal cues.
- We did find, however, a significant improvement in login time for graphical recognition as compared to textual recognition, even though the number of attempts for successful logins did not differ significantly between these conditions.

We organize the rest of this paper as follows: In §2, we give an overview of notable authentication schemes with a discussion on their limitations and the respective scopes of possible improvements. In §3, we explain from the perspective of cognitive psychology how the design choices for our study conditions are set up. We then describe our study procedure in §4 and present the results in §5. In §6, we discuss the findings from our study and highlight the possible directions for future research, followed by a conclusion in §7.

2 Related Work

In this section, we give a brief overview of notable textual and graphical password schemes in which we highlight why existing schemes are insufficient.

2.1 Textual Password Schemes

Traditional passwords. Traditional user-chosen textual passwords are fraught with security problems because of password reuse and predictable patterns [12, 37]. Different password restriction policies (e.g., increasing the minimum password length, requiring a combination of different types of characters, and using password strength meters) have been deployed to get users to create stronger passwords [18, 37]. However, in separate studies, Proctor et al. [32] and Shay et al. [37] report that such policies do not necessarily lead to more secure passwords but do adversely affect memorability in some cases.

Mnemonic Passwords. Kuo et al. [27] studied passwords based on mnemonic phrases, in which the user chooses a memorable phrase and uses a character (often the first letter) to represent each word in the phrase. Results [27] show that user-selected mnemonic passwords are slightly more resistant to brute-force attacks than traditional passwords. However, mnemonic passwords are found to be more predictable when users choose common phrases to create their passwords. A properly chosen dictionary may further increase the success rate in guessing mnemonic passwords [27].

System-assigned passwords. System-assigned random textual password schemes are more secure but fail to provide sufficient memorability, even when natural-language words are used [36, 44]. Wright et al. [44] compared the usability of three different system-assigned textual password schemes: Word Recall, Word Recognition, and Letter Recall. None of these schemes had sufficient memorability rates.

PTP. Forget et al. [19, 21] proposed the Persuasive Text Passwords (PTP) scheme, in which the user first creates a password, and PTP improves its security by placing randomly-chosen characters at random positions into the password. PTP is resilient against attacks exploiting password reuse and predictable patterns. Unfortunately, the memorability for PTP is just 25% when two random characters are inserted at random positions [19].

Cognitive questions. Furnell et al. [22] revealed the potential of cognitive questions and reported a high level of user satisfaction in using that for primary authentication. However, Just and Aspinall [26] identified the usability and security problems of using cognitive questions for authentication, and several other studies [33, 35] reported the vulnerability of this approach to targeted guessing attacks.

2.2 Graphical Password Schemes

Graphical password schemes can be divided into three categories [7], based on the kind of memory leveraged by the systems: i) Drawmetric (recall-based), ii) Locimetric (cued-recall-based), and iii) Cognometric (recognition-based).

Drawmetric. The user is asked to reproduce a drawing in this category of graphical passwords. In *Draw-a-Secret (DAS)* [25], a user draws on top of a grid, and the password is represented as the sequence of grid squares. Nali and Thorpe [28] have shown that users choose predictable patterns in DAS that include drawing symmetric images with 1-3 pen strokes, using grid cell corners and lines (presumably as points of reference) and placing their drawing approximately in the center of the grid.

BDAS [15] intends to reduce the amount of symmetry in the user's drawing by adding background images, but this may introduce other predictable behaviors such as targeting similar areas of the images or image-specific patterns [7]. DAS and BDAS have recall rates of no higher than 80%.

Locimetric. The password schemes in this category present users with one or more images as a memory cue to assist them selecting their particular points on the image(s). In the *Passpoints* [8] scheme, users select a sequence of click-points on a single image as their password. *Cued Click-Points (CCP)* [10] is a modified version of Passpoints, where users sequentially choose one click-point on each of five images. Dirik et al. [14] developed a model that can predict 70-80% of users' click positions in Passpoints. To address this issue, Chiasson et al. proposed *Persuasive Cued Click-Points (PCCP)* [11, 20], in which a randomly-positioned viewport is shown on top of the image during password creation, and users select their click-point within this viewport. The memorability for PCCP was found to be 83-94%.

In a follow-up study, Chiasson et al. [9] found predictability in users' click points, showing that in Passpoints, the click points are roughly evenly spaced across the image, in straight lines starting from left to right, and either completely horizontal or sloping from top to bottom. The authors [9] indicate that predictability is still a security concern for PCCP.

Cognometric. In this recognition-based category of graphical passwords, the user is asked to recognize and identify their password images from a set of distractor images. *Passfaces* [1] is the most studied cognometric scheme as it is commercially deployed by a number of large websites. The commercial Passfaces [1] product assigns a random set of faces instead of allowing users to choose, since the research [13] has found that users select predictable faces, biased by race, gender, and attractiveness of faces. However, Everitt et al. [16] show that users have difficulty in remembering system-assigned Passfaces.

Davis et al. [13] proposed the *Story* scheme, in which users select a sequence of images as their password and, to aid memorability, are encouraged to mentally construct a story to connect those images. During login, users have to identify their images in accurate order from a panel of decoy images. Though the user choices in Story are found to be more varied than the face-recognition-based scheme, the results still display some exploitable patterns, and the user study showed a memorability rate of about 85% [13].

In a recent study [4], Al-Ameen et al. found satisfactory memorability by combining various cues for graphical recognition, which suggests that the use of

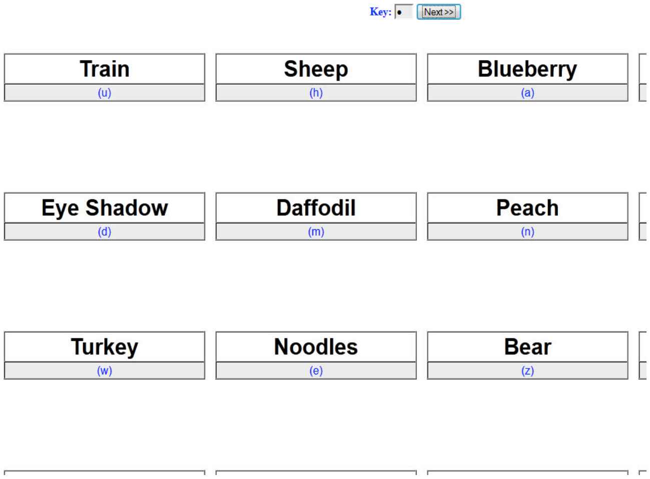


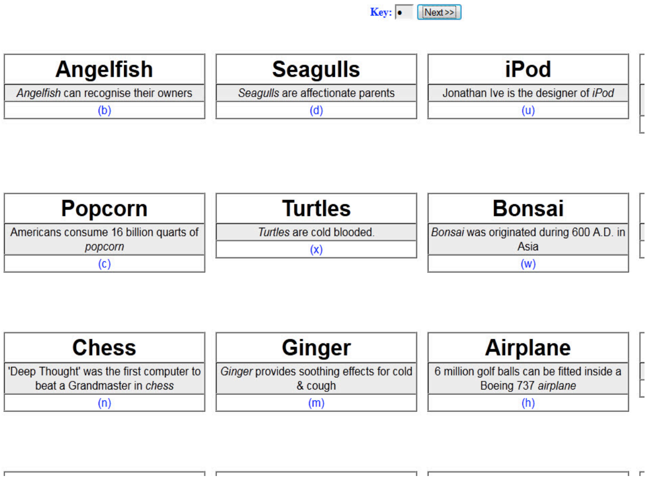
Fig. 1. A partial screen shot of the *Control* condition during login. Users enter the key, a lowercase letter shown in parentheses, in the password field (on top) to select the corresponding keyword. The keys are randomly assigned to keyword each time the portfolio is loaded, where no two keywords share the same key. During login, users are shown five such portfolios, where each presents a distinct set of 16 keywords including one of the five assigned keywords.

cues is very promising and motivates further study. In their experiment [4], the authors did not examine the impact of different cues, nor they studied textual recognition. Our deeper investigation on this issue helps to understand how humans' cognitive abilities could be leveraged through verbal cues for enhanced memorability in system-assigned textual recognition-based passwords. We also compare textual and graphical recognition to explore the usability gain of accommodating images, when users are provided with verbal cues.

3 System Design

Hlywa et al. [24] provide a guideline to design recognition-based authentication schemes with password-level security. We follow this guideline to design our study conditions, where the user is assigned five keywords at registration and has to recognize each of the assigned keywords from a distinct portfolio of 16 keywords during login. A successful authentication requires the user to recognize all five keywords correctly. For an unsuccessful login, the user is shown an error message at the end of the login attempt but not informed on which portfolio the mistake was made.

In our study, we implement three different recognition-based schemes. In Control condition, users remember and recognize the assigned keywords without the help of verbal cues (see Figure 1). In TextV scheme, the system offers verbal cues to help users with the memorization and recognition of the assigned



Angelfish
Angelfish can recognise their owners
(b)

Seagulls
Seagulls are affectionate parents
(d)

iPod
Jonathan Ive is the designer of *iPod*
(u)

Popcorn
Americans consume 16 billion quarts of *popcorn*
(c)

Turtles
Turtles are cold blooded.
(x)

Bonsai
Bonsai was originated during 600 A.D. in Asia
(w)

Chess
'Deep Thought' was the first computer to beat a Grandmaster in *chess*
(n)

Ginger
Ginger provides soothing effects for cold & cough
(m)

Airplane
6 million golf balls can be fitted inside a Boeing 737 *airplane*
(h)

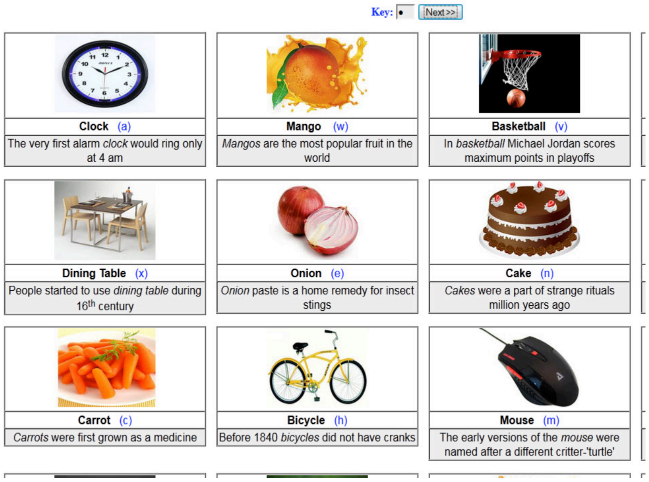
Fig. 2. A partial screen shot of *TextV* scheme during login. The facts corresponding to each keyword appear below that keyword.

keywords, where cues are shown both at registration and login (see Figure 2). In GraphicV scheme, the system provides users with images corresponding to the keywords along with the verbal cues (see Figure 3). In this section, we explain our design choices from the perspective of cognitive psychology and existing password literature.

3.1 Memory Retrieval

Users are required to perform a recognition task in our study. Researchers in psychology have found that recognition (identifying the correct item among a set of distractors) is easier than recall (reproducing the item from memory) [41] and have developed two main theories to explain this: *Generate-recognize theory* [5] and *Strength theory* [42].

Generate-recognize theory [5] speculates that recall is a two-phase process. In the generate phase, a list of candidate words is formed by searching long-term memory. Then, in the recognize phase, the list of words is evaluated to see if they can be recognized as the sought-out memory. According to this theory, recognition tasks do not utilize the generation phase and are thus faster and easier to perform. Strength theory [42] states that although recall and recognition involve the same memory task, recognition requires a lower threshold of strength that makes it easier. The point is commonly illustrated in examples from everyday life. For example, multiple choice questions are frequently easier than essay questions since the correct answer is available for recognition.



Clock (a) The very first alarm *clock* would ring only at 4 am	**Mango** (w) *Mangos* are the most popular fruit in the world	**Basketball** (v) In *basketball* Michael Jordan scores maximum points in playoffs
Dining Table (x) People started to use *dining table* during 16th century	**Onion** (e) *Onion* paste is a home remedy for insect stings	**Cake** (n) *Cakes* were a part of strange rituals million years ago
Carrot (c) *Carrots* were first grown as a medicine	**Bicycle** (h) Before 1840 *bicycles* did not have cranks	**Mouse** (m) The early versions of the *mouse* were named after a different critter-'turtle'

Fig. 3. A partial screen shot of GraphicV scheme during login. Each keyword is accommodated with the corresponding image.

3.2 Semantic Priming

Having a fixed set of objects in a certain place aids to augment *semantic priming*, which refers to recognizing an object through its relationship with other objects around it [1]. Semantic priming thus eases the recognition task [1]. For example, in Figure 3, the clock is not only in the upper-left-hand corner each time, but it is always next to the mango and above the dining table. This establishes a relationship between the objects and reinforces semantic priming. Thus, in each of our study conditions, the keywords in a portfolio remain same and presented at a fixed position whenever that portfolio is loaded.

3.3 Verbal Cues

We incorporate the scientific understanding of long-term memory to advance the usability properties of recognition-based authentication. According to the cognitive memory model proposed by Atkinson and Shiffrin [6], any new information is transferred to short-term memory (STM) through the sensory organs, where STM holds the information as *memory codes*, or mental representations of selected parts of the information. The information is transferred from STM to long-term memory (LTM), but only if it can be further processed and encoded (see the illustration in Figure 4). This encoding helps people to remember and retrieve the processed information efficiently over an extended period of time. To motivate such encoding, we examine the efficacy of providing verbal cues with the keywords.

If the system provides verbal cues, i.e., real-life facts related to the keywords, then users may focus their attention on associating the keywords with the corresponding cues, which should help to process and encode the information in

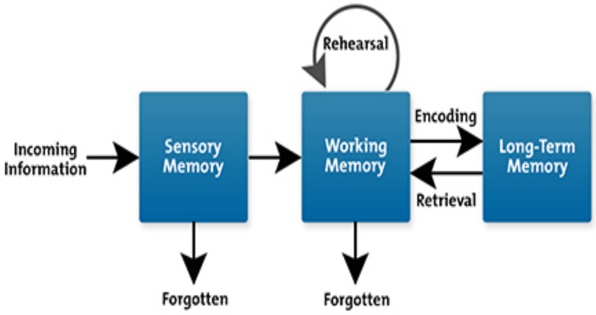


Fig. 4. Illustration of cognitive memory model

memory and store them in the long-term memory. For example, the keyword "Turtles" is associated with the verbal cue 'Turtles are cold blooded". The cues would also assist users to recognize the keywords in the future and thus enhance their memorability.

Psychology research [5,41] has shown that it is difficult to remember information spontaneously without memory cues, and this suggests that authentication schemes should provide users with cues to aid memory retrieval. *Encoding specificity theory* [40] postulates that the most effective cues are those that are present at the time of remembering. In TextV and GraphicV schemes, verbal cues are provided during registration, i.e., the learning period, and also at login.

3.4 Visual Memory

In GraphicV scheme, we leverage users' visual memory, in addition to offering verbal cues. Psychology research shows that the human brain is better at memorizing graphical information as compared to textual information [29, 31]. This is known as the *picture superiority effect*. Several explanations for the picture superiority effect have been proposed. The most widely accepted is *dual-coding theory* [31], which postulates that in human memory, images are encoded not only visually and remembered as images, but they are also translated into a verbal form (as in a description) and remembered semantically. Another explanation is the *sensory-semantic model* [29], which states that images are accompanied by more distinct sensory codes that allow them to be more easily accessed than text.

3.5 Variant Response

In the existing recognition-based schemes [1,24,44], mouse input is used to select a keyword or image, where the keywords/images in a portfolio remain the same but are positioned randomly each time that portfolio is loaded to compensate for shoulder surfing risk during login. However, the shoulder-surfing study of Tari et al. [38] reveals that recognition-based schemes with keyboard input provide higher resilience to shoulder surfing than schemes with mouse input, since the keyboard input associated with a particular keyword/image changes across the user's login sessions. This feature is called *variant response*, i.e., varying the user's responses across the login sessions [7].

For a recognition-based scheme providing variant response through varying keyboard inputs, the shoulder surfer needs to learn both the user's keystrokes and the corresponding keywords/images by looking at the keyboard and monitor. Tari et al.'s study [38] shows that observing both the monitor and keyboard at the same time is difficult.[4] Thus, the schemes in our study provide users with variant response feature, where each time a portfolio is loaded, a distinct lowercase letter `a-z` is assigned randomly as a *key* to one keyword on the page, and the user inputs the key letter corresponding to her assigned keyword into a single-character password field to move on to the next portfolio (see Figure 1, Figure 2 and Figure 3).

4 User Study

We now present the design of our user study, where we used a within-subjects design consisting of three experimental conditions. Using a within-subjects design controls for individual differences and permits the use of statistically stronger hypothesis tests. The study procedures were approved by our university's Institutional Review Board (IRB) for human subjects research.

4.1 Participants, Apparatus and Environment

For this experiment, we recruited 52 students (34 women, 18 men) through our university's Psychology Research Pool. Participants came from diverse backgrounds, including majors from Nursing, Psychology, Business, Environmental Science, Biochemistry, and Spanish Language. The age of the participants varied between 18 to 48 with a mean age of 22. Each participant was compensated with course credit for participation and was aware that her performance or feedback in this study would not affect the amount of compensation.

The lab studies were conducted with one participant at a time to allow the researchers to observe the users' interactions with the system. We created three realistic and distinct websites, including sites for banking, email, and social networking. The sites used the images and layouts from familiar commercial sites, and each of them was equipped with one of our three password schemes.

In our study, each of the five portfolios in a scheme consists of unique set of keywords and images that are not repeated in any other portfolio nor in any other scheme. In other words, we did not reuse any keywords or images. We collected the images and real-life facts (verbal cues) from free online resources.

[4] though we note that videotaping could overcome this.

4.2 Procedure

We conducted the experiment in two sessions, each lasting around 30 minutes. The second session took place one week after the first one to test users' memorization of the assigned passwords. A one-week delay is larger than the maximum average interval for a user between subsequent logins to any of her important accounts [23] and is also a common interval used in authentication studies (e.g., [3, 4, 15, 30, 44]).

Session 1. After signing a consent form, the participants were given an overview of our study. Then they performed registration for each of the three sites, each outfitted with a distinct scheme. The sites were shown to the participants at random order during registration. After registering with each scheme, participants performed a practice login with that scheme. They performed another practice login with each scheme after completing registration for all of the three sites. We did not collect data for these practice trials. They were asked to not record (e.g., write down or take a picture) their authentication secrets.

Session 2. The participants returned one week after registration and logged into each of the three sites using the assigned passwords. The sites were shown to the participants in random order, and they could make a maximum of five attempts for a successful login. After they had finished, we conducted an anonymous survey. Participants were then compensated and thanked for their time.

4.3 Ecological Validity

Most of our participants were young and all of them were university educated, which represents a large number of frequent Web users, but may not generalize to the entire population. They came from diverse majors. As the study was performed in a lab setting, we were only able to gather data from 52 participants. However, lab studies have been preferred to examine brain-powered memorability of passwords [17]. Since lab studies take place in a controlled setting, it helps to establish performance bounds and figure out whether field tests are worthwhile in future research. We believe that 52 provides a suitable sample size for a lab study as compared to the prior studies on password memorability [3, 4, 10, 11, 39, 43].

5 Results

We now discuss the results of our user study. To analyze our results, we use statistical tests and consider results comparing two conditions to be significantly different when we find $p < 0.05$. When comparing two conditions where the variable is at least ordinal, we use a Wilcoxon signed-rank test for the matched pairs of subjects and a Wilcoxon-Mann-Whitney test for unpaired results. Wilcoxon tests are similar to t-tests, but make no assumption about the distributions of

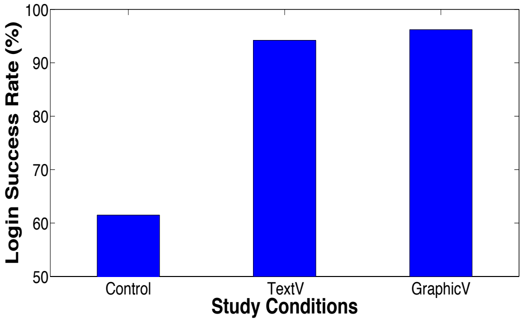


Fig. 5. Login success rates for the study conditions [Number of participants=52]

the compared samples, which is appropriate to the datasets in our conditions. Whether or not a participant successfully authenticated is a binary measure, and so we use either a McNemar's test (for matched pairs of subjects) or a chi-squared test (for unpaired results) to compare login success rates between two conditions. Here, we tested the following hypotheses:

Hypothesis 1

H_1: *The login success rate for TextV would be significantly higher than that for the Control condition.*

The TextV scheme offers verbal cues (i.e., real-life facts related to the keyword), where cues are shown both at registration and login. So, the users could memorize their keywords through associating them with the corresponding cues, which should help to process and encode the information to store them in long-term memory (see §3 for detailed discussion). Moreover, the cues would assist users to recognize the keywords in the future, which should enhance their memorability. Thus, we hypothesized that TextV scheme would have significantly higher login success rate than the Control condition.

Our results show that out of 52 participants in our study, 49 participants (94.2%) succeeded to log in using TextV, while 32 participants (61.5%) logged in successfully with the Control condition (see Figure 5). Whether or not a participant successfully authenticated is a binary measure, so we compare login success rates between conditions using McNemar's test. We found that the login success rate for TextV scheme was significantly higher than that for the Control condition, $\mathcal{X}^2(1, N = 52) = 12.2$, $p < 0.01$. Thus, H_1 is supported by these results.

Hypothesis 2

H_1: *The login success rate for GraphicV would be significantly higher than that for the TextV scheme.*

In GraphicV scheme, we accommodate images corresponding to the keywords, in addition to offering verbal cues. Psychology research reveals *picture*

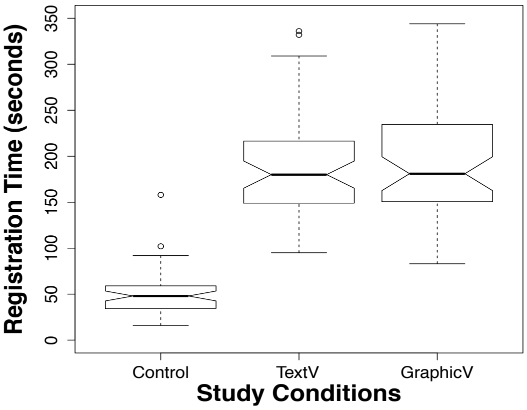


Fig. 6. Registration time for the study conditions

superiority effect showing that the human brain is better at memorizing graphical information as compared to textual information [29,31]. Thus, we hypothesized that the login success rate for GraphicV would be significantly higher than that for the TextV scheme.

We found that out of 52 participants in our study, 50 participants (96.2%) succeeded to log in using GraphicV scheme, and 49 participants (94.2%) logged in successfully with the TextV scheme. The results for McNemar's test show that there was no significant difference between TextV and GraphicV schemes in terms of login success rate, $\mathcal{X}^2(1, N = 52) = 0$, $p = 1$. Hence, H_2 is not supported by these results.

5.1 Registration Time

We illustrate the results for registration time in Figure 6. We found that the median registration times for Control, TextV, and GraphicV schemes were 48 seconds, 180 seconds, and 181 seconds, respectively. We use a Wilcoxon signed-rank test (appropriate for matched pairs of subjects) to evaluate two schemes in terms of registration time. The results show that the registration time for TextV ($V = 0$, $p < 0.01$) and GraphicV ($V = 1$, $p < 0.01$) were significantly less than that for the Control condition. We did not find a significant difference in registration time between TextV and GraphicV schemes ($V = 633.5$, $p = 0.62$).

5.2 Login Time and Number of Attempts

In this paper, *number of attempts* and *login time* respectively refer to the required attempts and time for successful logins only, unless otherwise specified. We do not get matched pairs of subjects while comparing two schemes in terms of login time or number of attempts for successful logins, since some participants who logged in successfully for one scheme failed in the other scheme. So, we use a

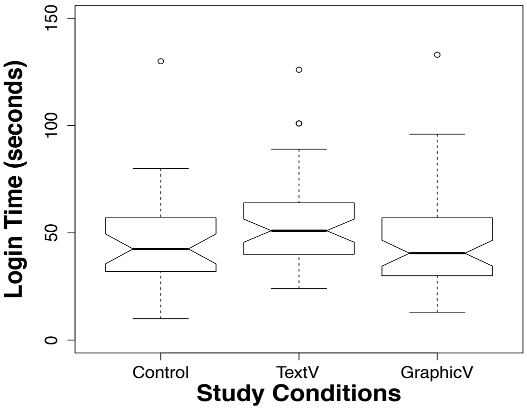


Fig. 7. Login time for the study conditions

Wilcoxon-Mann-Whitney test (appropriate for unpaired results) to evaluate two schemes in terms of login time and the number of attempts for successful logins.

Login Time. We illustrate our results for login time in Figure 7. We found that the median login time for Control, TextV, and GraphicV were 43 seconds, 51 seconds, and 41 seconds, respectively. The results for Wilcoxon-Mann-Whitney tests show that the login time for Control ($W = 569.5$, $p < 0.05$) and GraphicV ($W = 878.5$, $p < 0.05$) were significantly less than that for the TextV scheme. We did not find a significant difference in login time between Control and GraphicV ($W = 790$, $p = 0.93$).

Number of Attempts. The mean number of attempts for a successful login was less than two for each of the three study conditions, while the median was one in each case (see Table 1). The results for Wilcoxon-Mann-Whitney tests found no significant difference between any pair of study conditions in terms of the number of attempts for a successful login.

Table 1. Number of Attempts for Successful Logins [SD: Standard Deviation]

Study Conditions	Mean	Median	SD
Control	1.3	1	0.8
TextV	1.4	1	0.9
GraphicV	1.3	1	0.6

Table 2. Questionnaire responses for the usability of each of the three schemes. Scores are out of 10. * indicates that scale was reversed. *Med*: Median, *Mo*: Mode

Questions	Control		TextV		GraphicV	
	Med	*Mo*	*Med*	*Mo*	*Med*	*Mo*
I could easily sign up with this scheme	5	1	7.5	10	9	10
Logging in using this scheme was easy	5.5	1	7.5	10	9	10
Passwords in this scheme are easy to remember	5	1	7	10	8	10
I could easily use this scheme every day	5	4	7	10	8	10

5.3 User Feedback

We asked the participants to answer a set of 10-point Likert-scale questions (1: *strong disagreement*, 10: *strong agreement*) at the end of the second session, where a higher score indicates a more positive result for a scheme. We illustrate the results in Table 2. Since Likert scale data are ordinal, it is most appropriate to calculate mode and median for Likert-scale responses [34].

The feedback of the participants were overall positive (mode and median higher than neutral) for TextV and GraphicV schemes, however, the majority of participants reported concern about the usability of Control condition. The results for Wilcoxon signed-rank tests (appropriate for matched pairs of subjects) show that the user feedback was significantly better for TextV and GraphicV schemes in comparison to the Control condition; for *ease of registration*: TextV-Control ($V = 500$, $p < 0.05$), GraphicV-Control ($V = 118$, $p < 0.05$), *ease of login*: TextV-Control ($V = 567$, $p < 0.05$), GraphicV-Control ($V = 124$, $p < 0.05$), *memorability*: TextV-Control ($V = 577$, $p < 0.05$), GraphicV-Control ($V = 108.5$, $p < 0.05$), and *ease of everyday use*: TextV-Control ($V = 672$, $p < 0.05$), GraphicV-Control ($V = 27$, $p < 0.05$).

6 Discussion

System-assigned recognition-based passwords (e.g., Passfaces [1]) are now commercially available and deployed by a number of large websites. They fail, however, to gain satisfactory memorability [16], since it is difficult for most people to memorize system-assigned passwords. Our study explores a promising direction to improve memorability for these passwords by leveraging humans' cognitive abilities through verbal cues, and presents a comparison between textual and graphical recognition to understand the underlying usability gain of adding images, when users are provided with such memory cues.

We accommodate the scientific understanding of long-term memory to improve the memorability of system-assigned recognition-based passwords. As noted by Atkinson and Shiffrin [6], any new information is transferred from short-term memory to long-term memory, when it is duly processed and encoded. In our study, we explored the impact of verbal cues for an elaborate encoding of authentication information to ease recognition during login. As we compared TextV scheme with the Control condition, our results showed a significant improvement in login success rate when users were provided with verbal cues to aid textual recognition.

We design GraphicV scheme to examine the *picture superiority effect* when users are provided with verbal cues. As we compared TextV with GraphicV scheme, our results found no significant difference in login success rate. The login time for GraphicV was significantly less than that for TextV scheme, although we found no significant difference in number of attempts for successful logins. Thus, we infer that when verbal cues are provided, accommodating images with the keywords might not contribute to gain a significant improvement in login success rate, however, aids users with a faster recognition of the keywords, and so on, reduces the login time.

During registration with TextV and GraphicV schemes, the participants may have learned the assigned keywords by correlating them with the verbal cues. This then assisted them with the elaborate processing of the authentication information, but also contributed to the higher registration time compared to the Control condition. No significant difference was found between TextV and GraphicV schemes in terms of registration time.

Future Work. Now that lab-study results show promise for implementing verbal cues, it would be interesting to evaluate the approaches through a long-term field study with larger and more diverse populations, where we would explore the training effects on login performances over time. A recent field study [2] reveals that login time significantly decreases with the frequent use of a scheme due to training effects.

In future work, we would explore the efficacy of verbal cues for the people from different age groups. We would also make a deeper investigation to understand the impact of cues in improving the memorability of passwords for the people with different cognitive limitations.

7 Conclusion

In our study, we aimed to understand the impact of verbal cues on system-assigned recognition-based passwords, and designed three different study conditions to achieve this goal. In a study with 52 participants, we had a 94.2% login success rate for a textual recognition-based scheme offering verbal cues (TextV), which was significantly higher than that for the Control condition. To understand the usability gain of accommodating images for a scheme providing verbal cues, we compared TextV and GraphicV schemes, and found no significant difference in login success rate, although users required less time to recognize the keywords when they were accommodated with images. These findings shed light on a promising research direction to leverage humans' cognitive ability through verbal cues in gaining high memorability for system-assigned random passwords.

References

1. Passfaces corporation. The science behind Passfaces. White paper, `http://www.passfaces.com/enterprise/resources/white_papers.htm`
2. Al-Ameen, M.N., Wright, M.: A comprehensive study of the GeoPass user authentication scheme. Tech. rep., arXiv:1408.2852 [cs.HC] (2014)
3. Al-Ameen, M.N., Wright, M.: Multiple-password interference in the geopass user authentication scheme. In: USEC (2015)
4. Al-Ameen, M.N., Wright, M., Scielzo, S.: Towards making random passwords memorable: Leveraging users' cognitive ability through multiple cues. Tech. rep., arXiv:1503.02314 [cs.HC] (2015)
5. Anderson, J.R., Bower, G.H.: Recognition and recall processes in free recall. Psychological Review 79(2) (1972)
6. Atinkson, C.R., Shiffrin, M.R.: Human memory: A proposed system and its control processes. K.W. Spence and J.T. Spence (eds), Advances in the psychology of learning and motivation, New York academic press (1968)
7. Biddle, R., Chiasson, S., van Oorschot, P.: Graphical passwords: Learning from the first twelve years. ACM Computing Surveys 44(4) (2012)
8. Chiasson, S., Biddle, R., van Oorschot, P.C.: A second look at the usability of click-based graphical passwords. In: SOUPS (2007)
9. Chiasson, S., Forget, A., Biddle, R., van Oorschot, P.: User interface design affects security: Patterns in click-based graphical passwords. International Journal of Information Security 8(6) (2009)
10. Chiasson, S., van Oorschot, P.C., Biddle, R.: Graphical password authentication using cued click points. In: ESORICS (2007)
11. Chiasson, S., Stobert, E., Biddle, R., van Oorschot, P.: Persuasive cued click-points: design, implementation, and evaluation of a knowledge- based authentication mechanism. IEEE TDSC 9 (2012)
12. Das, A., Bonneau, J., Caesar, M., Borisov, N., Wangz, X.: The tangled web of password reuse. In: NDSS (2014)
13. Davis, D., Monrose, F., Reiter, M.: On user choice in graphical password schemes. In: USENIX Security (2004)
14. Dirik, A.E., Memon, N., Birget, J.C.: Modeling user choice in the passpoints graphical password scheme. In: SOUPS (2007)
15. Dunphy, P., Yan, J.: Do background images improve "Draw a Secret" graphical passwords? In: CCS (2007)
16. Everitt, K., Bragin, T., Fogarty, J., Kohno, T.: A comprehensive study of frequency, interference, and training of multiple graphical passwords. In: CHI (2009)
17. Fahl, S., Harbach, M., Acar, Y., Smith, M.: On the ecological validity of a password study. In: SOUPS (2013)
18. Florencio, D., Herley, C.: Where do security policies come from? In: SOUPS (2010)
19. Forget, A.: A World with Many Authentication Schemes. Ph.D. thesis, Carleton University (2012)
20. Forget, A., Chiasson, S., van Oorschot, P.C., Biddle, R.: Persuasion for stronger passwords: Motivation and pilot study. In: PT (2008)
21. Forget, A., Chiasson, S., van Oorschot, P., Biddle, R.: Improving text passwords through persuasion. In: SOUPS (2008)
22. Furnell, S., Papadopoulos, I., Dowland, P.: A long-term trial of alternative user authentication technologies. Information Management and Computer Security 12(2) (2004)

23. Hayashi, E., Hong, J.I.: A diary study of password usage in daily life. In: CHI (2011)
24. Hlywa, M., Biddle, R., Patrick, A.S.: Facing the facts about image type in recognition-based graphical passwords. In: ACSAC (2011)
25. Jermyn, I., Mayer, A., Monrose, F., Reiter, M., Rubin, A.: The design and analysis of graphical passwords. In: USENIX Security (1999)
26. Just, M., Aspinall, D.: Personal choice and challenge questions a security and usability assessment. In: SOUPS (2009)
27. Kuo, C., Romanosky, S., Cranor, L.F.: Human selection of mnemonic phrase-based passwords. In: SOUPS (2006)
28. Nali, D., Thorpe, J.: Analyzing user choice in graphical passwords. Tech. Rep. TR-04-01, School of Computer Science, Carleton University (2004)
29. Nelson, D.L., Reed, V.S., McEvoy, C.L.: Learning to order pictures and words: A model of sensory and semantic encoding. Journal of Experimental Psychology: Human Learning and Memory 3(5) (1977)
30. Nicholson, J., Coventry, L., Briggs, P.: Age-related performance issues for PIN and face-based authentication systems. In: CHI (2013)
31. Paivio, A.: Mind and Its Evolution: A Dual Coding Theoretical Approach. Lawrence Erlbaum: Mahwah, N.J. (2006)
32. Proctor, R.W., Lien, M.C., Vu, K.P.L., Schultz, E.E., Salvendy, G.: Improving computer security for authentication of users: Influence of proactive password restrictions. Behavior Research Methods, Instruments, and Computers 34(2) (2002)
33. Rabkin, A.: Personal knowledge questions for fallback authentication: Security questions in the era of Facebook. In: SOUPS (2008)
34. Robertson, J.: Stats: We're doing it wrong. `http://cacm.acm.org/blogs/blog-cacm/107125-stats-were-doing-it-wrong/fulltext` (April 2011)
35. Schechter, S., Brush, A.J.B., Egelman, S.: It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In: IEEE S&P (2009)
36. Shay, R., Kelley, P.G., Komanduri, S., Mazurek, M.L., Ur, B., Vidas, T., Bauer, L., Christin, N., Cranor, L.F.: Correct horse battery staple: Exploring the usability of system-assigned passphrases. In: SOUPS (2012)
37. Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F.: Encountering stronger password requirements: User attitudes and behaviors. In: SOUPS (2010)
38. Tari, F., Ozok, A., Holden, S.: A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In: SOUPS (2006)
39. Thorpe, J., MacRae, B., Salehi-Abari, A.: Usability and security evaluation of GeoPass: A geographic location-password scheme. In: SOUPS (2013)
40. Tulving, E., Thompson, D.M.: Encoding specificity and retrieval processes in episodic memory. Psychological Review 80(5) (1973)
41. Tulving, E., Watkins, M.: Continuity between recall and recognition. American Journal of Psych 86(4) (1973)
42. Wickelgren, W.A., Norman, D.A.: Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology 3 (1966)
43. Wiedenbeck, S., Waters, J., Birget, J., Brodskiy, A., Memon, N.: Authentication using graphical passwords: Effects of tolerance and image choice. In: SOUPS (2005)
44. Wright, N., Patrick, A.S., Biddle, R.: Do you see your password? Applying recognition to textual passwords. In: SOUPS (2012)