This is an Accepted Manuscript of an article published by Taylor & Francis in Behaviour & Information Technology on Dec 09, 2020, available online:
http://www.tandfonline.com/10.1080/0144929X.2020.1858161

On Improving the Memorability of System-assigned Recognition-based Passwords

Mahdi Nasrullah Al-Ameen*, Sonali T. Marne**, Kanis Fatema**, Matthew Wright †,

Shannon Scielzo ‡

Utah State University, The University of Texas at Arlington**, Rochester Institute of Technology* †,

The University of Texas Southwestern ‡

mahdi.al-ameen@usu.edu, {sonali.marne,kanis.fatema}@mavs.uta.edu, matthew.wright@rit.edu,

shannon.scielzo@UTSouthwestern.edu

(*Accepted: 25 Nov 2020, Published online: 09 Dec 2020*)

User-chosen passwords reflecting common strategies and patterns ease memorization but offer uncertain and often weak security, while system-assigned passwords provide higher security guarantee but suffer from poor memorability. We thus examine the technique to enhance password memorability that incorporates a scientific understanding of long-term memory. In particular, we examine the efficacy of providing users with *verbal cues*—real-life facts corresponding to system-assigned keywords. We also explore the usability gain of including images related to the keywords along with verbal cues. In our multi-session lab study with 52 participants, textual recognition-based scheme offering verbal cues had a significantly higher login success rate (94.23%) compared to the control condition, i.e., textual recognition without verbal cues (61.54%). We found that when users were provided with verbal cues, adding images contributed to faster recognition of the assigned keywords, and thus had an overall improvement in usability. So, we conducted a field study with 54 participants to further examine the usability of graphical recognition-based scheme offering verbal cues, which showed an average login success rate of 98% in a real-life setting and an overall improvement in login performance with more login sessions. These findings show a promising research direction to gain high memorability for system-assigned passwords.

Keywords: Usable security; System-assigned password; Memorability; Lab study; Field study.

1. Introduction

Traditional user-chosen textual passwords suffer from security problems because of password reuse and predictable patterns (Das et al. 2014; Ur et al. 2015; Jenkins et al. 2014), which is essential for protecting and mitigating threats to the information assets and technical resources available within computer-based systems (Crossler et al. 2013; Vu et al. 2007; Lowry, Dinev, and Willison 2017). Users bear the responsibility of ensuring the security of their account by creating a password that should be chosen with creativity and intelligence to achieve satisfactory security and memorability. Many users compromise on security with weak but memorable passwords. Policies requiring users to create longer passwords with different character types do not necessarily lead to more secure passwords, but they do adversely affect memorability in some cases (Shay et al. 2014; Vu et al. 2007; Campbell, Kleeman, and Ma 2007).

Studies in psychology have shown that recognition, such as identifying an assigned picture from a set, is an easier memory task than recall (Tulving and Watkins 1973; Anderson and Bower 1972; Wickelgren and Norman 1966). Inspired by these findings, researchers have proposed and examined recognition-based authentication schemes as alternatives to pure recall-based schemes (e.g., traditional textual password) in hopes that by reducing the memory burden on users, more secure passwords can be generated. Wright et al. (2012) 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 (Authentication 2004) is an example of a graphical recognition-based scheme that is now commercially available and deployed by many large websites.[1]

To ensure security, the commercial Passfaces (Authentication 2004) 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 (Das et al. 2014; Jenkins et al. 2014), 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 (Wright, Patrick, and Biddle 2012) and graphical recognition (Everitt et al. 2009). Thus, it still remains a critical challenge to design an authentication scheme that offers satisfactory

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

memorability for system-assigned random passwords.

1.1. *Contributions*

The study of Wright et al. (2012) anticipated that showing the keywords in the same position whenever a portfolio is loaded would improve the memorability for recognition-based password schemes, and suggested the approach to be examined in future work. We adopt the suggestion of Wright et al. (2012) to design our study conditions by showing the keywords in a portfolio in the same position each time a portfolio is loaded.

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 (Atkinson and Shiffrin 1968), helping users recognize their keywords when logging in later. We provide a detailed discussion on these memorization processes in §3.

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. (2012), keeping the position of keywords fixed in a portfolio did not provide a satisfactory login success rate (61.54%).
- Verbal cues made a significant contribution to improving the login success rate for textual recognition (94.23%).
- 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, although the login success rate (96.15%) for GraphicV was slightly higher than that of TextV scheme (94.23%).

- We did find, however, a significant improvement in login time for graphical recognition (i.e., GraphicV) as compared to textual recognition (i.e., TextV), even though the number of attempts for successful logins did not differ significantly between these conditions.

In our lab study, GraphicV scheme offered an overall improvement in usability as compared to TextV scheme, and thus, we conducted a field study to gain in-depth understanding on the usability of this scheme. A field study offers strong ecological validity and the best measure of login performance in a realistic setting (Biddle, Chiasson, and Van Oorschot 2012). We found that the memorability for GraphicV was satisfactory in a real-life setting with an average login success rate of 98%. Our field study found an overall improvement in login performance with more login sessions, including an 81% reduction in median login time to just 7 seconds by the 17^{th} login session.

2. Related Work

Passwords schemes are used for user authentication in various systems, including authentication in the Session Initiation Protocol (SIP) in mobile networks (Mishra 2016) and to improve security against various types of attacks, like password guessing (Crossler et al. 2013; Bonneau 2012), phishing (Sharifi et al. 2007), and shoulder-surfing (De Luca et al. 2013; Al-Ameen, Haque, and Wright 2016). In this section, however, we limit our discussion to schemes designed for authenticating users to their online accounts and aim to enhance guessing resilience and password-memorability.

In our literature review, we focused on knowledge-based authentication. We note that prior work (Mishra et al. 2015) has also proposed alternatives to such schemes, like using physical tokens (e.g., smart cards) for authentication. The extra hardware requirement adds costs, however, and is hard to extend to multiple accounts without creating a "necklace effect," where the user must carry an unwieldy number of tokens. Biometrics like fingerprints (Roy, Memon, and Ross 2017) can be useful for authenticating to devices, but they have the downside of not being easily updated if stolen or damaged. For these reasons, as well as cost and ease of deployment, knowledge-based authentication remains the dominant authentication technique for online accounts. For a more

extensive survey of the field of password replacement schemes, we suggest the work of Bonneau et al. (2012).

2.1. *Textual Password Schemes*

Traditional user-chosen textual passwords are fraught with security problems and are especially prone to password reuse and predictable patterns (Das et al. 2014; Campbell, Ma, and Kleeman 2011; Cazier and Medlin 2006; Jenkins et al. 2014; Ur et al. 2017). Ur et al. (2015) showed that users have many misconceptions that contribute to creating weak passwords. For example, many users believe that adding a special character at the end of a password makes it secure (Ur et al. 2015). Their study also showed that users could anticipate only targeted guessing attacks, believing that it is a secure approach to use a birthday or name as a password if that information is not available on social networking sites. More recently, Ur et al. (2016) showed that users have serious misconceptions about the impact of basing passwords on common phrases and including digits and keyboard patterns in passwords, which leads them to create weak and predictable authentication secrets. As reported by Tam et al. (2010), users engage in creating weak passwords because they do not see any immediate negative consequences to themselves.

Different password restriction policies have been deployed to get users to create stronger passwords (Campbell, Ma, and Kleeman 2011; Vu et al. 2007; Shay et al. 2014; Campbell, Kleeman, and Ma 2007; Mayer, Kirchner, and Volkamer 2017). These studies report, however, that such policies do not necessarily lead to more secure passwords. Worse still, they have been shown to adversely affect memorability. Due to the difficulty of remembering strong passwords, many users create weak passwords, even when they are aware of strategies for creating a strong password (Von Zezschwitz, De Luca, and Hussmann 2013). Zhang et al. (2009) leveraged list reduction and unique identifier methods to improve the memorability of passwords. Even with such techniques, however, password reuse and predictable patterns remain important unresolved issues. In a separate study, Shay et al. (2015) found that a multi-step password-creation process that provides guidance to users is not effective enough in creating strong passwords.

While user-chosen textual passwords fail to provide adequate security, Bonneau et al. (2012) suggested a set of usability, security, and deployability metrics that need to be addressed to provide a viable solution to the usability-security tension in online user authentication. In their metrics, system-assigned random password schemes are more secure than user-chosen passwords, but they

fail to provide sufficient memorability, even when natural language words are used (Shay et al. 2012; Wright, Patrick, and Biddle 2012; Keith, Shao, and Steinbart 2009). Forget et al. (2008a; 2012) proposed the Persuasive Text Passwords (PTP) scheme as a hybrid between user-selected and system-assigned passwords, but the memorability can be poor—as low as 25%.

2.2. *Graphical Password Schemes*

Graphical password schemes can be divided into three categories (Biddle, Chiasson, and Van Oorschot 2012) based on the kind of memory leveraged by the systems: i) Drawmetric (recall-based), ii) Locimetric (cued-recall-based), and iii) Cognometric (recognition-based). We cover these briefly here, and we suggest the survey paper by Biddle et al. (2012) for more detail.

2.2.1. Drawmetric

The user is asked to reproduce a drawing in this category of graphical passwords. In *Draw-a-Secret (DAS)*, a user draws on top of a grid, and the password is represented as the sequence of grid squares (Mayer, Monrose, and Rubin 1999). Nali and Thorpe (2004) 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* (Dunphy and Yan 2007) 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 (Biddle, Chiasson, and Van Oorschot 2012). DAS and BDAS have recall rates of no higher than 80%.

2.2.2. Locimetric

The password schemes in this category present users with one or more images as a memory cue to assist them in selecting their particular points on the image(s). In the *Passpoints* scheme (Chiasson, Biddle, and van Oorschot 2007; Wiedenbeck et al. 2005), users select a sequence of click-points on a single image as their password. *Cued Click-Points (CCP)* (Chiasson, Van Oorschot, and Biddle 2007) is a modified version of Passpoints, where users sequentially choose one click-point on each of five images. Dirik et al. (2007) developed a model that can predict 70-80% of users' click positions in Passpoints. To address this issue, researchers proposed *Persuasive Cued Click-Points (PCCP)*, 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 (Forget et al. 2008b; Chiasson et al. 2012). The memorability for PCCP was found to be 83-94%.

2.2.3. 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* (Authentication 2004) is the most studied cognometric scheme as it is commercially deployed by a number of large websites. The commercial Passfaces (Authentication 2004) product assigns a random set of faces instead of allowing users to choose, since the research (Davis, Monrose, and Reiter 2004) has found that users select predictable faces, biased by race, gender, and attractiveness of faces. However, Everitt et al. (2009) show that users have difficulty in remembering system-assigned Passfaces.

Davis et al. (2004) 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% (Davis, Monrose, and Reiter 2004). To reduce predictability, the Deja Vu scheme (Dhamija, Perrig et al. 2000) uses random art images instead of the images of human faces or common objects. Mihajlov et al. (2016), however, further identified the predictability of user-choice in recognition-based graphical passwords in terms of color, shape, and category.

In the Photographic Authentication system (Pering et al. 2003), users are required to provide their own set of digital photos during registration, such that at login, they can recognize their own photos from decoy photos. The decoy images are randomly selected from the images collected from other users. Pering et al. found that participants had a 90% login success rate with this scheme. However, this scheme is likely vulnerable to the *guessing-by-acquaintances attack* because of the use of user-selected personal photos (Tullis and Tedesco 2005).

In a recent study (Al-Ameen, Wright, and Scielzo 2015), the authors found satisfactory memorability by combining various cues for graphical recognition, which suggests that the use of cues is very promising and motivates further study. In their lab experiment (Al-Ameen, Wright, and Scielzo 2015), the authors did not examine the impact of different cues, nor did they study textual recognition. Our deeper investigation on this issue helps to understand how humans' cognitive abil-

Table 1. An Overview on the Security and Usability of Recognition-based Authentication Schemes [NR: Not Reported, Photo. Auth.: Photographic Authentication]

Password Scheme	Theoretical Entropy (bits)	System-Assigned	Lab Study	Field Study	Login Success Rate	Login Time (sec.)
Passfaces/Face	13	Yes	No	Yes	72%	88
Story	12	No	No	Yes	85%	NR
Photo. Auth.	20	No	Yes	No	95%	40
Deja Vu	16	No	Yes	No	90%	36
Textual Recognition	20	Yes	Yes	No	61.54%	42.50
TextV	20	Yes	Yes	No	94.23%	51
GraphicV	20	Yes	Yes	Yes	96.15%	40.50

ities 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 and conducted both lab and field studies to examine our scheme's usability.

We focused on recognition-based password schemes in our study. In Table 1, we compare our schemes with existing recognition-based password schemes that are designed for online user authentication. We note that some schemes have been evaluated through multiple studies, where the login performance of users might vary across different studies. For the sake of simplicity in the presentation in Table 1, we note the minimum login success rate and maximum login time of an authentication scheme reported in any study, which also indicates the minimum usability offered by that password scheme in terms of these metrics. For example, in our lab study, the login success rate and login time of GraphicV scheme was 96.15% and 40.50 seconds, respectively. In our field study, the login success rate of GraphicV remained consistently 100% from the 17^{th} session, where the login time became 7 seconds by the 17^{th} session. In this regard, we reported 96.15% as the login success rate and 40.50 seconds as the login time of GraphicV scheme (see Table 1).

3. System Design

Hlywa et al. (2011) 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. 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.

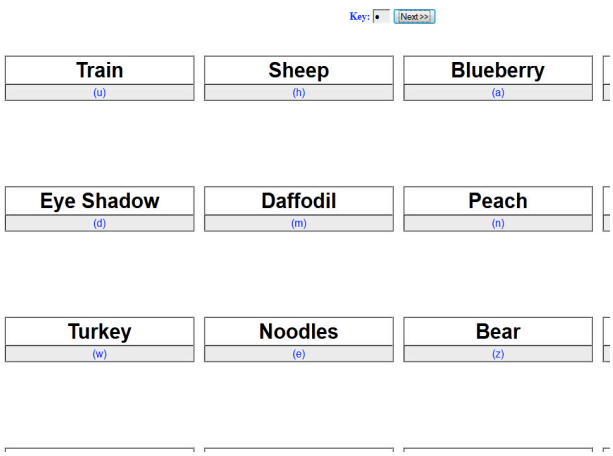

Figure 1. A partial screen shot of the *Control* condition during login. Users enter the key, a lowercase letter shown in parentheses, in the password field (on top) to select the corresponding keyword. No two keywords share the same key. During login, users are shown five such portfolios, where each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

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 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) (Tulving and Watkins 1973) and have developed two main theories to explain this: *Generate-recognize theory* (Anderson and Bower 1972) and *Strength theory* (Wickelgren and Norman 1966).

Generate-recognize theory (Anderson and Bower 1972) 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

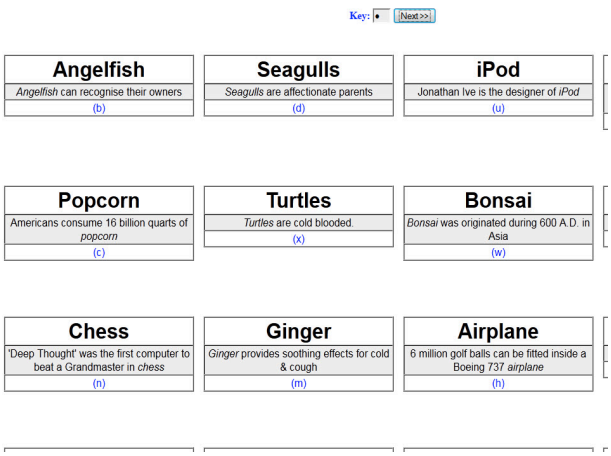

Figure 2. A partial screen shot of *TextV* scheme during login. The facts corresponding to each keyword appear below that keyword.

and are thus faster and easier to perform. Strength theory (Wickelgren and Norman 1966) 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.

3.2. *Semantic Priming*

Having a fixed set of objects in a particular place aids to augment *semantic priming*, which refers to recognizing an object through its relationship with other objects around it (Authentication 2004). Semantic priming thus eases the recognition task (Authentication 2004). 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 the same and presented at a fixed position whenever a 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 (1968), any new information is transferred to short-term memory (STM) through the sensory organs, where STM holds the information as *memory codes* or mental represen-

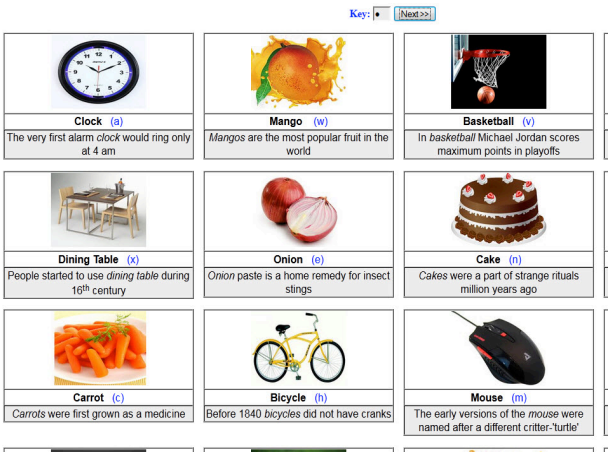

Figure 3. A partial screen shot of GraphicV scheme during login. Each keyword is accommodated with the corresponding image.

tations 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. This encoding helps people remember and retrieve the processed information efficiently over an extended period. 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 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 in recognizing the keywords in the future and thus enhancing their memorability.

Psychology research (Anderson and Bower 1972; Tulving and Watkins 1973) 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* (Tulving and Thomson 1973) 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. Based on these findings, we present our first hypothesis:

Hypothesis 1 (H_1): *The login success rate for TextV (and GraphicV), which offers cues at registration and login, will be significantly higher than that for the Control, which does not.*

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 (Paivio 2014; Nelson, Reed, and McEvoy 1977). 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* (Paivio 2014), which postulates that images are encoded in human memory 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 of picture superiority effect is the *sensory-semantic model* (Nelson, Reed, and McEvoy 1977), which states that images are accompanied by more distinct sensory codes that allow them to be more easily accessed than the textual information. Considering these findings and theories, we present our second hypothesis:

Hypothesis 2 (H_2): *The login success rate for GraphicV, which offers images, will be significantly higher than that for the TextV scheme, which does not.*

3.5. *Input Type*

In existing recognition-based password schemes (Authentication 2004; Hlywa, Biddle, and Patrick 2011; Wright, Patrick, and Biddle 2012), mouse input is used to select a keyword or an image. The study of Tari et al. (2006) showed that using keyboard input provides higher resilience to shoulder-surfing attacks than using mouse input. So, we use keyboard input for the schemes in our study, where a lowercase letter `a-z` is assigned 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, 2, and 3). The user-entered letter in the password field is shown as an asterisk to reduce the risk of shoulder surfing.

4. Study I: Lab Study

We used a within-subjects design in our lab study, which consists 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 a unique set of keywords and images that are not repeated in any other portfolio nor 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.

4.2. *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 (Hayashi and Hong 2011) and is also a common interval used in authentication studies (e.g., (Nicholson, Coventry, and Briggs 2013; Al-Ameen and Wright 2015; Wright, Patrick, and Biddle 2012; Dunphy and Yan 2007; Al-Ameen, Wright, and Scielzo 2015)).

4.2.1. 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 not to record (e.g., write down or take a picture)

their authentication secrets.

4.2.2. 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.

5. Study II: Field Study

The results from our lab study show that GraphicV performed best in terms of usability. We thus conducted a field study to further examine the usability of this scheme in a real-life setting. We conducted this study on a class with both undergraduate and graduate students. At the beginning of the study, the students were informed that we developed a website to let them access course study materials and their grades on exams and assignments.[2] With a projector, the experimenter showed the students how GraphicV scheme works. The students were then asked to complete their registration with this scheme, with each student in the class given a username. To protect against unauthorized access, students' usernames were pre-stored in the system so that only students in this class could create accounts, one per username. The mean registration time for GraphicV scheme was 265 seconds (median: 241 seconds, standard deviation: 110 seconds).

The GraphicV system was active for 74 days. Out of 64 students in this class, 54 students (10 women and 44 men with a mean age of 25) gave positive consent to use their login information for the study and signed consent forms before participating in an anonymous paper-based survey. They were compensated with extra credit in a class assignment for participating in this survey, and an alternative assignment was offered for those who did not want to participate. None of the students had participated before in a password-related user study.

In this field study, the users could log in at any time from anywhere using a desktop or laptop computer. During authentication, we started counting login time after the username had been entered. A successful attempt required the user to correctly enter both her username and GraphicV password. An unsuccessful attempt refers only to sessions where the username was correct, but the

[2]Grades were posted in a file containing all students' grades and anonymized by replacing names with a code given to each student.

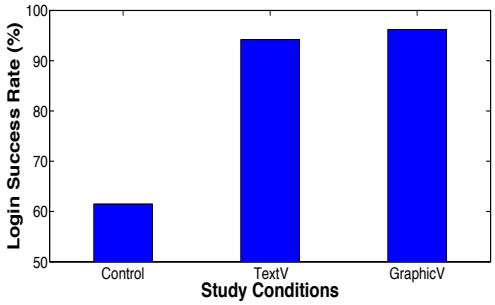

Figure 4. *Lab Study:* Login success rates for the study conditions [Number of participants=52]

GraphicV password was incorrect. We found that the participants always entered their username correctly.

If a participant could not log in because of forgetting the password, she had to send an email to the experimenter from her `.edu` email account, and in response, she would receive an email with a link that would lead her through the registration process to relearn the system-assigned password. Two participants were required to relearn their password within the first few days of the study. Thereafter, no participant was required to relearn her GraphicV password during the study.

6. Lab Study Results

We use statistical tests to analyze our results 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 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 will 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 by associating

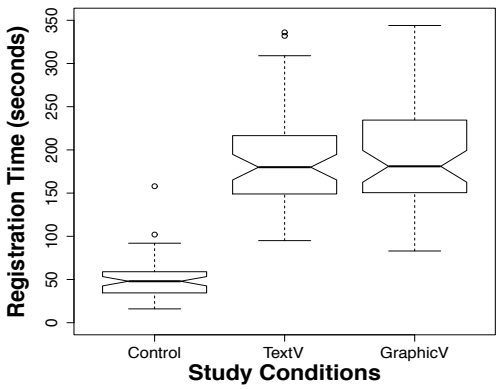

Figure 5. *Lab Study:* Registration time for the study conditions

them with the corresponding cues, which should help process and encode the information to store them in long-term memory (see §3 for a detailed discussion). Moreover, the cues would help users recognize the keywords in the future, which should enhance their memorability. Thus, we hypothesized that TextV scheme would have a significantly higher login success rate than the Control condition.

Our results show that out of 52 participants in our study, 49 participants (94.23%) succeeded in logging in using TextV, while 32 participants (61.54%) logged in successfully with the Control condition (see Figure 7). 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, $\mathcal{X}^2(1, N = 52) = 12.20$, $p < 0.01$ (effect size = 1.10). Thus, H_1 is supported by these results.

Hypothesis 2 (H_2): *The login success rate for GraphicV will 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 superiority effect* showing that the human brain is better at memorizing graphical information as compared to textual information (Paivio 2014; Nelson, Reed, and McEvoy 1977). 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 (96.15%) succeeded in logging in using GraphicV scheme, and 49 participants (94.23%) 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, $\mathcal{X}^2(1, N = 52) = 0$, $p = 1$ (effect size

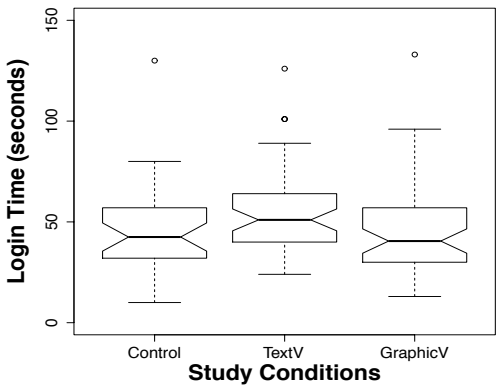

Figure 6. *Lab Study:* Login time for the study conditions

= 0). Hence, H_2 is not supported by these results.

6.1. *Registration Time*

We illustrate the results for registration time in Figure 5. 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 ($V = 0$, $p < 0.01$; effect size = 12.11) and GraphicV ($V = 1$, $p < 0.01$; effect size = 12.10) were significantly higher than that for the Control condition. We did not find a significant difference in registration time between TextV and GraphicV schemes ($V = 633.50$, $p = 0.62$; effect size = 7.19).

6.2. *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,

Table 2. *Lab Study:* Number of Attempts for Successful Logins [SD: Standard Deviation]

Study Conditions	Mean	Median	SD
Control	1.25	1	0.76
TextV	1.39	1	0.86
GraphicV	1.32	1	0.55

Table 3. *Lab Study:* 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

	Control		TextV		GraphicV	
Questions	*Med*	*Mo*	*Med*	*Mo*	*Med*	*Mo*
I could easily sign up with this scheme	5	1	7.50	10	9	10
Logging in using this scheme was easy	5.50	1	7.50	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

since some participants who logged in successfully for one scheme failed in the other scheme. So, we use a 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.

6.2.1. Login Time.

We illustrate our results for login time in Figure 6. We found that the median login time for Control, TextV, and GraphicV were 42.50 seconds, 51 seconds, and 40.50 seconds, respectively. The results for Wilcoxon-Mann-Whitney tests show that the login time for Control ($W = 569.50$, $p < 0.05$; effect size = 0.41) and GraphicV ($W = 878.52$, $p < 0.05$; effect size = 0.50) were significantly less than that for the TextV scheme. We did not find a significant difference in login time between Control and GraphicV ($W = 790$, $p = 0.93$; effect size = 0.02).

6.2.2. 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 2). 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.

6.3. *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 3. Since Likert scale data are ordinal, it is most appropriate to calculate mode and median for Likert-scale responses (Robertson 2011).

The feedback of the participants were overall positive (mode and median higher than neutral) for

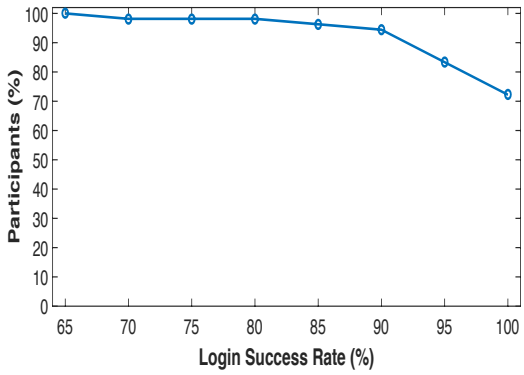

Figure 7. *Field Study:* Login success rate (54 participants).

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 ($V = 500$, $p < 0.05$; effect size = 8.12), GraphicV-Control ($V = 118$, $p < 0.05$; effect size = 11.10), *ease of login*: TextV-Control ($V = 567$, $p < 0.05$; effect size = 7.64), GraphicV-Control ($V = 124$, $p < 0.05$; effect size = 11.05), *memorability*: TextV-Control ($V = 577$, $p < 0.05$; effect size = 7.57), GraphicV-Control ($V = 108.50$, $p < 0.05$; effect size = 11.18), and *ease of everyday use*: TextV-Control ($V = 672$, $p < 0.05$; effect size = 6.93), GraphicV-Control ($V = 27$, $p < 0.05$; effect size = 11.88).

7. Field Study Results

In this section, we present our results and analysis on the login performance of users with GraphicV in a real-life setting.

7.1. *Overall Login Performance*

In our field study, we recorded 1349 login sessions for 54 participants, where a single login session (or *login*) by a participant may include multiple attempts to authenticate successfully. To find the full distribution of the number of attempts needed for a successful login, we did not limit the number of attempts a participant can make during a login session.

Participants performed 25 logins on average (median: 23, standard deviation: 13). We measured the average login performance of each participant in her login sessions (see Figure 7 and 8) and

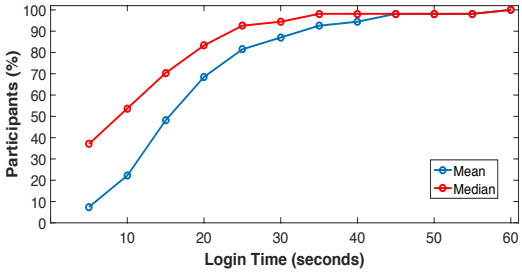

Figure 8. *Field Study:* Login time.

calculated the overall login performances for all of the 54 participants over 1349 login sessions. The overall login success rate was 98%. Users required 1.10 attempts (on average) per successful login, and the median login time was 9 seconds (mean: 15 seconds, standard deviation: 24 seconds).

To illustrate the login performance in more detail, Figures 7 and 8 show empirical cumulative distribution functions (ECDFs) of login performance statistics taken over the users in our study (in Figure 7, the x-axis is shown with increasing success rates and thus appears reversed).

Figure 7 shows the login success rates among participants. GraphicV proved sufficiently memorable for nearly all of our participants. 72% of participants had a 100% login success rate and 94% had at least a 90% success rate. We note that all of our participants logged in successfully within two attempts on average.

The performance for login time was more mixed. Figure 8 shows the average login time among participants. The mean login time was 15 seconds or less for 48% of participants and 20 seconds or less for 69% of participants. The median login time was 5 seconds or less for 37% of participants, 10 seconds or less for 54% of participants, and 20 seconds or less for 83% of participants.

7.2. *Training Effects*

To determine the extent of any training effects for GraphicV users in a real-world setting, we analyzed the change in login performance over login sessions. We illustrate the results at x^{th} login session ($x = 1, 5, 9, 13, 17, 21, 25, 29, 33$), where there are three login sessions between each value of x. Here, we consider up to the 33^{rd} login session, since using the higher values of x would make for a rather small sample size (e.g., 7 users for $x = 41$).

We note that the sample size shrinks for each successive value of $x > 5$ (see Table 4). As we are looking for a training effect, we may be concerned about the remaining population of users being more adept at using the system than those who have stopped logging in. Our results, however, show that the number of login sessions performed by a participant did not have a strong correlation with

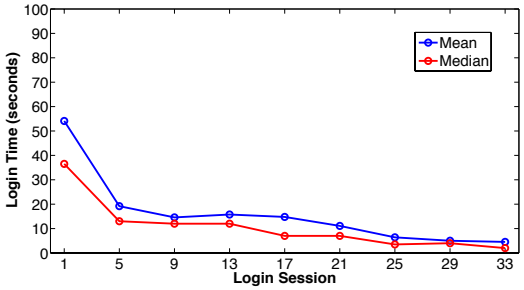

Figure 9. *Field Study:* The change in login time over the sessions.

Table 4. *Field Study:* Number of participants in the x^{th} login session.

x	1	5	9	13	17	21	25	29	33
Participants	54	54	52	44	38	34	22	15	13

her login success rate ($r = 0.37$).

In our field study, the login success rate and the number of attempts for successful logins were satisfactory right from the first login session, and thus, we see minimal training effects on these metrics of login performance. In particular, the login success rate was 96% in the first login session, which stayed above 96% in the subsequent login sessions and remained consistently 100% from the 17^{th} session. Also, for all login sessions, the mean number of attempts for successful login was less than 1.40 and the median number of attempts was 1.

The training effect was most prominent for login time as shown in Figure 9. A given (x, y) point in Figure 9 represents the average login performance (y) of the participants calculated over the x^{th} login session of each individual. Note that the x^{th} login session of any given participant likely occurred at a different time than that of other participants. The number of participants varied for different values of x (login session), since the participants performed different numbers of logins. Table 4 represents the number of participants in each of x^{th} login sessions.

As shown in Figure 9, the median login time was 37 seconds in the first login session, which decreased to 13 seconds in the 5^{th} login session and reduced to 7 seconds in the 17^{th} login session, an 81% reduction. The mean and median login times decreased over the login sessions, where we find an exception for mean login time at the 13^{th} login session.

By comparing login times for pairs of sessions, we can apply significance tests to find significant improvements due to training effects. Below, we report results for pairs of sessions (i, j), where a login time being significantly shorter in the j^{th} session than in the i^{th} session also means that it was shorter in later sessions $k > j$. Using Wilcoxon-Mann-Whitney tests,[3] we find the following

[3]Wilcoxon tests are similar to t-tests, but make no assumption about the distributions of the compared samples, which is

pairs to have significantly different login times: 1^{st} and 5^{th} ($W = 2313, p < 0.01$; effect size = 1.55), 5^{th} and 17^{th} ($W = 1265, p < 0.05$; effect size = 0.49), 9^{th} and 21^{st} ($W = 1095, p < 0.05$; effect size = 0.46), 13^{th} and 25^{th} ($W = 677, p < 0.05$; effect size = 0.75), 17^{th} and 33^{rd} ($W = 347, p < 0.05$; effect size = 0.64), 21^{st} and 33^{rd} ($W = 315, p < 0.05$; effect size = 0.69). Our analysis suggests that although most of the performance improvement occurs in the first few sessions, users continue to get moderately faster at logging in even after 21 sessions.

8. Study III: A Small-scale Field Study

In this section, we report on a pilot field study that we conducted for a 56-bit version of the GraphicV scheme. A scheme offering 56-bit passwords, also called *cryptographic passwords* (Biddle, Chiasson, and Van Oorschot 2012; Bonneau and Schechter 2014), provides much greater resistance against guessing than the 20-bit version of the scheme. This higher level of protection is important for high-stakes scenarios, such as a password for enterprise login or as a master key to protect other credentials, e.g., in a password manager (Bonneau and Schechter 2014). We note that 20 bits is considered sufficient against online guessing attacks, which is sufficient protection for most uses (Florêncio, Herley, and Van Oorschot 2014).

In our future work, we would conduct a large-scale field study on GraphicV scheme offering 56 bits of entropy, where users are required to recognize 14 images, each from a distinct portfolio of 16 images. To understand this scheme's potential in providing cryptographic security and identify the scope for improvement before conducting a large-scale study, we conducted a small-scale field study on a graduate class with 12 students. The study procedure was the same as in Study II, and all of the 12 students agreed to participate in this study. One student took part in both of our field studies.

The study continued for 43 days, in which we recorded 499 login sessions. The overall login success rate was 98%. Users required 1.10 attempts on average per successful login, while the median login time was 29 seconds. In this case, the median login time was less than 15 seconds for 33% of participants and 25 seconds or less for 42% of participants. We found an improvement in login time with more login sessions, where the median login decreased to 15 seconds by the 33^{rd} login session, a 75% reduction compared to the median login time in the first session.

The results of this study demonstrate that GraphicV offering cryptographic strength has the

appropriate to the datasets in our conditions.

potential to be further studied in future work, though it requires further attention for improvements in login time.

9. Discussion

In this section, we highlight the impact of verbal and graphical cues based on our findings in the lab study. We then discuss the deployment of our scheme in a real-life setting, training effects, and amenability to lockout rules, as supported by the results of our field study. We conclude this section by noting the limitations of our study and pointing to the scopes of future work.

9.1. *Impact of Verbal 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 (1968), 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 the login success rate when users were provided with verbal cues to aid textual recognition.

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 and 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.

9.2. *Impact of Graphical Cues*

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 the login success rate. The login time for GraphicV was significantly less than that for TextV scheme, although we found no significant difference in the 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 the login success rate but aids users with

faster recognition of the keywords, and so on, to reduce the login time.

9.3. *Deployment in a Real-life Setting*

As pointed out by Biddle et al. (2012), a field study offers strong ecological validity and the best measure of login performance in a realistic setting. Our field study shows a satisfactory memorability for GraphicV with an overall login success rate of 98%.

The deployment of a secure and memorable authentication scheme is important not only for the everyday computer and Internet usage of people (Al-Ameen and Kocabas 2020; Boss et al. 2015), but also to provide security for emerging technologies (Roman, Zhou, and Lopez 2013), maintain security and privacy in information management systems (Silic, Barlow, and Back 2017; Chatterjee, Sarker, and Valacich 2015), offer secure collaboration among professionals in sensitive profession (Watkins et al. 2016; McGregor et al. 2017; Watkins et al. 2017), and to address the general security concerns within business and organizational settings (Lowry et al. 2015; Haque et al. 2020; Siponen, Mahmood, and Pahnila 2014; Safa, Von Solms, and Furnell 2016; Dang-Pham, Pittayachawan, and Bruno 2017). The deployment of GraphicV in a real-life scenario does not require any change in the current authentication server compared to traditional textual passwords. In this regard, a textual password comprising of lowercase letters (used to select system-assigned keywords) would be stored at the server for each user. At the client-end, users do not need to memorize the characters used to select the keyword; rather, they could remember the system-assigned keyword with the help of given memory cues. During authentication, users recognize the keywords and select them by entering the corresponding lowercase letters that remain fixed across the login sessions.

9.4. *Training effect*

Since the prior field studies on system-assigned passwords did not present a detailed analysis of the training effect, it remains of particular interest to the research community to learn how login performances change over login sessions in a long-term field study. In our field study, the login success rate and the number of attempts for successful logins were satisfactory right from the first login session, while training effects played an important role in the improvement of login time with more login sessions.

Our field study found an overall improvement in login performance with more login sessions,

including an 81% reduction in median login time to just 7 seconds by the 17^{th} login session, while the login success rate remained consistently 100% from the 17^{th} session. So, it is clear that studying training effects in a field study provides a deeper understanding on the usability of a scheme.

Although it is difficult to remember a set of random letters (Al-Ameen, Haque, and Wright 2014), with the regular use of GraphicV, meaning repeatedly entering in the same letters, it is possible that participants may remember the letters as well, due to training effect. To know more about this, we asked participants to write down the letters in a paper at the end of study, where about half (44%) of the participants were able to correctly recall all five letters that they had to type for selecting their keywords. It is not clear, though, if the participants memorized the letters with regular entry or put additional effort into memorization. We now plan to conduct a study to investigate deeper into this issue.

9.5. *Lockout rules*

Lockout rules (Florêncio, Herley, and Coskun 2007) are implemented in many systems to protect against online guessing attacks. To implement a lockout rule that is both secure and convenient for legitimate users, it is important to figure out the number of attempts an actual user would usually require to log in successfully. Our field study gives insight into this issue, as we found that 100% of participants made, at most, two attempts on average to authenticate successfully. Thus, GraphicV is amenable to reasonable lockout rules.

9.6. *Limitations and Future Work*

In our studies, most of the participants were young, and all were university educated, which may not generalize to the entire population. However, they are still representative of a large number of frequent Web users. In our lab study, we had 52 participants from diverse majors, which we believe provides a suitable sample size for a lab study as compared to the prior studies on password memorability (Thorpe, MacRae, and Salehi-Abari 2013; Chiasson et al. 2012; Al-Ameen, Wright, and Scielzo 2015; Chiasson, Van Oorschot, and Biddle 2007; Al-Ameen, Haque, and Wright 2014).

In our field study, given that the participants performed a real-life task that mattered in their specific situation, the task had reasonable ecological validity. Now that our results for field study with young participants show promise, we would examine the usability of GraphicV scheme for senior users in our future work. It is not yet clear how our scheme would perform for people with

cognitive limitations (e.g., learning disabilities). So, in our future work, we would evaluate this scheme for people with learning disabilities to better understand its usability for the broadest possible set of users.

10. Conclusion

System-assigned recognition-based passwords (e.g., Passfaces (Authentication 2004)) are now commercially available and deployed by a number of large websites. They fail, however, to gain satisfactory memorability (Everitt et al. 2009), 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 we present 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 found that verbal cues played a significant role in improving the login success rate for textual recognition, and adding images contributed to a significant improvement in login time. The GraphicV scheme, which performed best in terms of usability in our lab study, was further evaluated through a field study. The memorability for GraphicV was satisfactory in a real-life setting, while the login time significantly improved with more login sessions because of training effects. To the best of our knowledge, ours is the first field study to explore training effects on the login performance of a system-assigned recognition-based password scheme.

Finally, in our pilot study on GraphicV of cryptographic-strength, we found that there is potential for high login success and moderate login times even for high-security applications. These findings point towards a promising future research direction in leveraging humans' cognitive strength through memory cues in gaining high memorability for system-assigned random passwords.

References

Al-Ameen, Mahdi Nasrullah, S M Taiabul Haque, and Matthew Wright. 2014. *Q-A: Towards the Solution of Usability-Security Tension in User Authentication.* Technical report. arXiv:1407.7277 [cs.HC].

Al-Ameen, Mahdi Nasrullah, SM Taiabul Haque, and Matthew Wright. 2016. "Leveraging autobiographical memory for two-factor online authentication." *Information & Computer Security* .

Al-Ameen, Mahdi Nasrullah, and Huzeyfe Kocabas. 2020. ""I cannot do anything": User's Behavior and

Protection Strategy upon Losing, or Identifying Unauthorized Access to Online Account." In *Symposium on Usable Privacy and Security (Poster Session),* .

Al-Ameen, Mahdi Nasrullah, and Matthew Wright. 2015. "Multiple-password Interference in the Geopass User Authentication Scheme." In *Proc. Workshop Usable Secur.(USEC),* 1–6.

Al-Ameen, Mahdi Nasrullah, Matthew Wright, and Shannon Scielzo. 2015. "Towards Making Random Passwords Memorable: Leveraging Users' Cognitive Ability Through Multiple Cues." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,* 2315–2324. ACM.

Anderson, John R, and Gordon H Bower. 1972. "Recognition and Retrieval Processes in Free Recall." *Psychological Review* 79 (2): 97.

Atkinson, Richard C, and Richard M Shiffrin. 1968. "Human Memory: A Proposed System and its Control Processes." *Psychology of Learning and Motivation* 2: 89–195.

Authentication, Real User Personal. 2004. "The Science Behind Passfaces." *White Paper, June* .

Biddle, Robert, Sonia Chiasson, and Paul C Van Oorschot. 2012. "Graphical Passwords: Learning from the First Twelve Years." *ACM Computing Surveys (CSUR)* 44 (4): 19.

Bonneau, Joseph. 2012. "The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords." In *IEEE Symposium on Security and Privacy,* 538–552. IEEE.

Bonneau, Joseph, Cormac Herley, Paul C Van Oorschot, and Frank Stajano. 2012. "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes." In *IEEE Symposium on Security and Privacy,* 553–567. IEEE.

Bonneau, Joseph, and Stuart E Schechter. 2014. "Towards Reliable Storage of 56-bit Secrets in Human Memory." In *USENIX Security Symposium,* 607–623.

Boss, S.R., D.F. Galletta, P.B. Lowry, G.D. Moody, and P. Polak. 2015. "What do Users Have to Fear? Using Fear Appeals to Engender Threats and Fear that Motivate Protective Security Behaviors." *MIS Quarterly* 39 (4).

Campbell, John, Dale Kleeman, and Wanli Ma. 2007. "The Good and not so Good of Enforcing Password Composition Rules." *Information Systems Security* 16 (1): 2–8.

Campbell, John, Wanli Ma, and Dale Kleeman. 2011. "Impact of Restrictive Composition Policy on User Password." *Behaviour & Information Technology* 30 (3): 379–388.

Cazier, Joseph A, and B Dawn Medlin. 2006. "Password Security: An Empirical Investigation into E-commerce Passwords and their Crack Times." *Information Systems Security* 15 (6): 45–55.

Chatterjee, Sutirtha, Suprateek Sarker, and Joseph S Valacich. 2015. "The Behavioral Roots of Information Systems Security: Exploring Key Factors Related to Unethical IT Use." *Journal of Management Information Systems* 31 (4): 49–87.

Chiasson, Sonia, Robert Biddle, and Paul C van Oorschot. 2007. "A Second Look at The Usability of

Click-based Graphical Passwords." In *Proceedings of the 3rd Symposium on Usable Privacy and Security,* 1–12. ACM.

Chiasson, Sonia, Elizabeth Stobert, Alain Forget, Robert Biddle, and Paul C Van Oorschot. 2012. "Persuasive Cued Click-points: Design, Implementation, and Evaluation of a Knowledge-based Authentication Mechanism." *IEEE Transactions on Dependable and Secure Computing* 9 (2): 222–235.

Chiasson, Sonia, Paul C Van Oorschot, and Robert Biddle. 2007. "Graphical Password Authentication Using Cued Click Points." In *European Symposium on Research in Computer Security,* 359–374. Springer.

Crossler, Robert E, Allen C Johnston, Paul Benjamin Lowry, Qing Hu, Merrill Warkentin, and Richard Baskerville. 2013. "Future Directions for Behavioral Information Security Research." *Computers & Security* 32: 90–101.

Dang-Pham, Duy, Siddhi Pittayachawan, and Vince Bruno. 2017. "Why Employees Share Information Security Advice? Exploring the Contributing Factors and Structural Patterns of Security Advice Sharing in the Workplace." *Computers in Human Behavior* 67: 196–206.

Das, Anupam, Joseph Bonneau, Matthew Caesar, Nikita Borisov, and XiaoFeng Wang. 2014. "The Tangled Web of Password Reuse." In *The Network and Distributed System Security Symposium,* Vol. 1423–26.

Davis, Darren, Fabian Monrose, and Michael K Reiter. 2004. "On User Choice in Graphical Password Schemes." In *USENIX Security Symposium,* 11–24.

De Luca, Alexander, Emanuel Von Zezschwitz, Laurent Pichler, and Heinrich Hussmann. 2013. "Using Fake Cursors to Secure On-screen Password Entry." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2399–2402. ACM.

Dhamija, Rachna, Adrian Perrig, et al. 2000. "Deja Vu-A User Study: Using Images for Authentication." In *USENIX Security Symposium,* 4–18.

Dirik, Ahmet Emir, Nasir Memon, and Jean-Camille Birget. 2007. "Modeling User Choice in the PassPoints Graphical Password Scheme." In *Proceedings of the 3rd Symposium on Usable Privacy and Security,* 20–28. ACM.

Dunphy, Paul, and Jeff Yan. 2007. "Do Background Images Improve "Draw a Secret" Graphical Passwords?." In *Proceedings of the 14th ACM Conference on Computer and Communications Security,* 36–47. ACM.

Everitt, Katherine M, Tanya Bragin, James Fogarty, and Tadayoshi Kohno. 2009. "A Comprehensive Study of Frequency, Interference, and Training of Multiple Graphical Passwords." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 889–898. ACM.

Florêncio, Dinei, Cormac Herley, and Baris Coskun. 2007. "Do Strong Web Passwords Accomplish Anything?." In *USENIX Summit on Hot Topics in Security,* Vol. 7.

Florêncio, Dinei, Cormac Herley, and Paul C Van Oorschot. 2014. "An Administrator's Guide to Internet Password Research." In *28th Large Installation System Administration Conference,* 35–52.

Forget, Alain. 2012. "A World with Many Authentication Schemes." Ph.D. thesis, Carleton University.

Forget, Alain, Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. 2008a. "Improving Text Passwords through Persuasion." In *Proceedings of the 4th Symposium on Usable Privacy and Security,* 1–12. ACM.

Forget, Alain, Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. 2008b. "Persuasion for Stronger Passwords: Motivation and Pilot Study." In *International Conference on Persuasive Technology,* 140–150. Springer.

Haque, SM Taiabul, MD Romael Haque, Swapnil Nandy, Priyank Chandra, Mahdi Nasrullah Al-Ameen, Shion Guha, and Syed Ishtiaque Ahmed. 2020. "Privacy Vulnerabilities in Public Digital Service Centers in Dhaka, Bangladesh." In *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development,* 1–12.

Hayashi, Eiji, and Jason Hong. 2011. "A Diary Study of Password Usage in Daily Life." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2627–2630. ACM.

Hlywa, Max, Robert Biddle, and Andrew S Patrick. 2011. "Facing the Facts about Image Type in Recognition-based Graphical Passwords." In *Proceedings of the 27th Annual Computer Security Applications Conference,* 149–158. ACM.

Jenkins, Jeffrey L, Mark Grimes, Jeffrey Gainer Proudfoot, and Paul Benjamin Lowry. 2014. "Improving Password Cybersecurity through Inexpensive and Minimally Invasive Means: Detecting and Deterring Password Reuse through Keystroke-dynamics Monitoring and Just-in-time Fear Appeals." *Information Technology for Development* 20 (2): 196–213.

Keith, Mark, Benjamin Shao, and Paul Steinbart. 2009. "A Behavioral Analysis of Passphrase Design and Effectiveness." *Journal of the Association for Information Systems* 10 (2): 2.

Lowry, Paul Benjamin, Tamara Dinev, and Robert Willison. 2017. "Why Security and Privacy Research Lies at the Centre of the Information Systems (IS) Artefact: Proposing a Bold Research Agenda." *European Journal of Information Systems* 26 (6): 546–563.

Lowry, Paul Benjamin, Clay Posey, Rebecca (Becky) J Bennett, and Tom L Roberts. 2015. "Leveraging Fairness and Reactance Theories to Deter Reactive Computer Abuse Following Enhanced Organisational Information Security Policies: An Empirical Study of the Influence of Counterfactual Reasoning and Organisational Trust." *Information Systems Journal* 25 (3): 193–273.

Mayer, Ian Jermyn Alain, Fabian Monrose, and Michael K Reiter Aviel D Rubin. 1999. "The design and analysis of graphical passwords." In *Proceedings of the 8th USENIX Security Symposium,* 1–14.

Mayer, Peter, Jan Kirchner, and Melanie Volkamer. 2017. "A Second Look at Password Composition Policies in the Wild: Comparing Samples from 2010 and 2016." In *Symposium on Usable Privacy and Security,* 13–28.

McGregor, Susan E, Elizabeth Anne Watkins, Mahdi Nasrullah Al-Ameen, Kelly Caine, and Franziska

Roesner. 2017. "When the Weakest Link is Strong: Secure Collaboration in the Case of the Panama Papers." In *26th USENIX Security Symposium,* 505–522.

Mihajlov, Martin, Borka Jerman-Blažič, and Anita Ciunova Shuleska. 2016. "Why that Picture? Discovering Password Properties in Recognition-based Graphical Authentication." *International Journal of Human–Computer Interaction* 32 (12): 975–988.

Mishra, Dheerendra. 2016. "Design of a Password-based Authenticated Key Exchange Protocol for SIP." *Multimedia Tools and Applications* 75 (23): 16017–16038.

Mishra, Dheerendra, Ashok Kumar Das, Ankita Chaturvedi, and Sourav Mukhopadhyay. 2015. "A Secure Password-based Authentication and Key Agreement Scheme Using Smart Cards." *Journal of Information Security and Applications* 23: 28–43.

Nali, Deholo, and Julie Thorpe. 2004. *Analyzing User Choice in Graphical Passwords.* Technical report.

Nelson, Douglas L, Valerie S Reed, and Cathy L McEvoy. 1977. "Learning to Order Pictures and Words: A Model of Sensory and Semantic Encoding." *Journal of Experimental Psychology: Human Learning and Memory* 3 (5): 485.

Nicholson, James, Lynne Coventry, and Pam Briggs. 2013. "Age-related Performance Issues for Pin and Face-based Authentication Systems." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 323–332. ACM.

Paivio, Allan. 2014. *Mind and its Evolution: A Dual Coding Theoretical Approach.* Psychology Press.

Pering, Trevor, Murali Sundar, John Light, and Roy Want. 2003. "Photographic Authentication through Untrusted Terminals." *IEEE Pervasive Computing* 2 (1): 30–36.

Robertson, Judy. 2011. "Stats: We're Doing it Wrong." `http://cacm.acm.org/blogs/blog-cacm/107125-stats-were-doing-it-wrong/fulltext`.

Roman, Rodrigo, Jianying Zhou, and Javier Lopez. 2013. "On the Features and Challenges of Security and Privacy in Distributed Internet of Things." *Computer Networks* 57 (10): 2266–2279.

Roy, Aditi, Nasir Memon, and Arun Ross. 2017. "Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems." *IEEE Transactions on Information Forensics and Security* 12 (9): 2013–2025.

Safa, Nader Sohrabi, Rossouw Von Solms, and Steven Furnell. 2016. "Information Security Policy Compliance Model in Organizations." *Computers & Security* 56: 70–82.

Sharifi, Mohsen, Alireza Saberi, Mojtaba Vahidi, and Mohammad Zorufi. 2007. "A Zero Knowledge Password Proof Mutual Authentication Technique against Real-time Phishing Attacks." In *International Conference on Information Systems Security,* 254–258. Springer.

Shay, Richard, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Alain Forget, Saranga Komanduri, Michelle L Mazurek, William Melicher, Sean M Segreti, and Blase Ur. 2015. "A Spoonful of Sugar?:

The Impact of Guidance and Feedback on Password-creation Behavior." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,* 2903–2912. ACM.

Shay, Richard, Patrick Gage Kelley, Saranga Komanduri, Michelle L Mazurek, Blase Ur, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2012. "Correct Horse Battery Staple: Exploring the Usability of System-assigned Passphrases." In *Proceedings of the Eighth Symposium on Usable Privacy and Security,* 7. ACM.

Shay, Richard, Saranga Komanduri, Adam L Durity, Phillip Seyoung Huh, Michelle L Mazurek, Sean M Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2014. "Can Long Passwords be Secure and Usable?." In *Proceedings of the 32nd ACM Conference on Human Factors in Computing Systems,* 2927–2936. ACM.

Silic, Mario, Jordan B Barlow, and Andrea Back. 2017. "A New Perspective on Neutralization and Deterrence: Predicting Shadow IT Usage." *Information & management* 54 (8): 1023–1037.

Siponen, Mikko, M Adam Mahmood, and Seppo Pahnila. 2014. "Employees' Adherence to Information Security Policies: An Exploratory Field Study." *Information & management* 51 (2): 217–224.

Tam, Leona, Myron Glassman, and Mark Vandenwauver. 2010. "The Psychology of Password Management: a Tradeoff between Security and Convenience." *Behaviour & Information Technology* 29 (3): 233–244.

Tari, Furkan, Ant Ozok, and Stephen H Holden. 2006. "A Comparison of Perceived and Real Shoulder-surfing Risks between Alphanumeric and Graphical Passwords." In *Proceedings of the Second Symposium on Usable Privacy and Security,* 56–66. ACM.

Thorpe, Julie, Brent MacRae, and Amirali Salehi-Abari. 2013. "Usability and Security Evaluation of Geopass: A Geographic Location-password Scheme." In *Proceedings of the Ninth Symposium on Usable Privacy and Security,* 14. ACM.

Tullis, Thomas S, and Donna P Tedesco. 2005. "Using Personal Photos as Pictorial Passwords." In *Extended Abstracts on Human Factors in Computing Systems (CHI),* 1841–1844. ACM.

Tulving, Endel, and Donald M Thomson. 1973. "Encoding Specificity and Retrieval Processes in Episodic Memory." *Psychological Review* 80 (5): 352.

Tulving, Endel, and Michael J Watkins. 1973. "Continuity between Recall and Recognition." *The American Journal of Psychology* 739–748.

Ur, Blase, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, et al. 2017. "Design and Evaluation of a Data-driven Password Meter." In *Proceedings of the CHI Conference on Human Factors in Computing Systems,* 3775–3786. ACM.

Ur, Blase, Jonathan Bees, Sean M Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. "Do Users' Perceptions of Password Security Match Reality?." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* 3748–3760. ACM.

Ur, Blase, Fumiko Noma, Jonathan Bees, Sean M Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2015. ""I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab." In *Proceedings of the Eleventh Symposium on Usable Privacy and Security,* 123–140.

Von Zezschwitz, Emanuel, Alexander De Luca, and Heinrich Hussmann. 2013. "Survival of the Shortest: A Retrospective Analysis of Influencing Factors on Password Composition." In *IFIP Conference on Human-Computer Interaction,* 460–467. Springer.

Vu, Kim-Phuong L, Robert W Proctor, Abhilasha Bhargav-Spantzel, Bik-Lam Belin Tai, Joshua Cook, and E Eugene Schultz. 2007. "Improving Password Security and Memorability to Protect Personal and Organizational Information." *International Journal of Human-Computer Studies* 65 (8): 744–757.

Watkins, Elizabeth Anne, Mahdi Nasrullah Al-Ameen, Franziska Roesner, Kelly Caine, and Susan McGregor. 2017. "Creative and Set in Their Ways: Challenges of Security Sensemaking in Newsrooms." In *7th USENIX Workshop on Free and Open Communications on the Internet (FOCI),* .

Watkins, Elizabeth Anne, Franziska Roesner, Susan McGregor, Byron Lowens, Kelly Caine, and Mahdi Nasrullah Al-Ameen. 2016. "Sensemaking and Storytelling: Network Security Strategies for Collaborative Groups." In *International Conference on Collaboration Technologies and Systems,* 622–623. IEEE.

Wickelgren, Wayne A, and Donald A Norman. 1966. "Strength Models and Serial Position in Short-term Recognition Memory." *Journal of Mathematical Psychology* 3 (2): 316–347.

Wiedenbeck, Susan, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. 2005. "PassPoints: Design and Longitudinal Evaluation of a Graphical Password System." *International journal of human-computer studies* 63 (1-2): 102–127.

Wright, Nicholas, Andrew S Patrick, and Robert Biddle. 2012. "Do You See Your Password?: Applying Recognition to Textual Passwords." In *Proceedings of the Eighth Symposium on Usable Privacy and Security,* 8. ACM.

Zhang, Jie, Xin Luo, Somasheker Akkaladevi, and Jennifer Ziegelmayer. 2009. "Improving Multiple-password Recall: An Empirical Study." *European Journal of Information Systems* 18 (2): 165–176.

Appendix A. Survey Questionnaire

Lab Study:

Q. What is you gender?

Q. What is your age?

Q. What is your major?

Participants were shown the following 10-point Likert-scale questions (1: Strongly Disagree, 10: Strongly Agree) after their login attempts with each of the three schemes in Session 2.

a. I could easily sign up with this scheme.

b. Logging in using this scheme was easy.

c. Passwords in this scheme are easy to remember.

d. I could easily use this scheme every day.

Field Study:

Q. What is you gender?

Q. What is your age?