

The Impact of Cues and User Interaction on the Memorability of System-assigned  
Random Passwords

by

MAHDI NASRULLAH AL-AMEEN

Presented to the Faculty of the Graduate School of  
The University of Texas at Arlington in Partial Fulfillment  
of the Requirements  
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

Copyright © by Mahdi Nasrullah Al-Ameen  
All Rights Reserved

## ABSTRACT

The Impact of Cues and User Interaction on the Memorability of System-assigned  
Random Passwords

Mahdi Nasrullah Al-Ameen, Ph.D.

The University of Texas at Arlington,

Supervising Professor: Matthew Wright

Given the choice, users produce passwords reflecting common strategies and patterns that ease recall but offer uncertain and often weak security. Addressing this usability-security tension in user authentication remains the key research issue in password studies for decades. In this thesis, we aim to understand how humans' cognitive abilities could be leveraged to design more secure and memorable authentication schemes. To achieve this goal, we draw upon multiple theories from cognitive psychology and implement them in the context of improving memorability for system-assigned random passwords.

We argue that while the system assigns random passwords, it should also help users with memorization and recall. We investigated the feasibility of this approach with CuedR, a novel *cued-recognition* authentication scheme that provides users with multiple cues (visual, verbal, and spatial) and lets them choose the cues that best fit their learning process for later recognition of system-assigned keywords. The lab study on CuedR showed promise for providing multiple cues with a 100% memorability rate over the span of one week.

The study on CuedR did not examine the individual impact of each cue. Thus, we performed a second study to explore deeper into this issue, where we examined the efficacy of *spatial cues* (fixed position of images), *verbal cues* (phrases/facts related to the images), and employing *user interaction* (learning images through writing a short description at registration) to improve the memorability of system-assigned passwords based on face images and object images. In our multi-session lab study with 56 participants, we had a 98% login success rate for a scheme offering spatial and verbal cues (ObjectSV), while a scheme based on user interaction had a 95% login success rate for face images (FaceSUI) and a 93% login success rate for object images (ObjectSUI). Our analysis shows that verbal cues and user interaction made an important contribution to gain significantly higher login success rate as compared to the control conditions representing existing graphical password schemes.

Since the combination of spatial and verbal cues performed best in the second study in providing satisfactory memorability for system-assigned recognition-based graphical passwords, in the third study, we examined the impact of combining spatial and verbal cues for system-assigned recognition-based textual passwords. We 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 spatial and verbal cues (TextV), which was significantly higher than that for the control condition providing only spatial cues. 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. To note, the GraphicV scheme in this study is similar to the ObjectSV scheme in the second study.

The results from these lab studies showed that a cued-recognition-based scheme (e.g., GraphicV/ObjectSV) offering users with spatial and verbal cues for object images performed best in terms of memorability. So, we conducted a field study for a further understanding on the usability of this scheme in a real-life scenario, where the memorability for GraphicV scheme was quite satisfactory with an average login success rate of 98%. Our analysis also shows that login time significantly improved with more login sessions because of training effect.

We believe that our research makes an important contribution to understand how humans' cognitive abilities could be leveraged to design more secure and memorable authentication schemes.

## TABLE OF CONTENTS

ABSTRACT . . . . .	iii
LIST OF ILLUSTRATIONS . . . . .	xi
LIST OF TABLES . . . . .	xiii
Chapter	Page Chapter
1. Introduction . . . . .	1
1.1 Statement of the Problem . . . . .	1
1.2 Contributions . . . . .	2
1.2.1 CuedR . . . . .	3
1.2.2 The Impact of Cues and User Interaction on Graphical Recognition . . . . .	4
1.2.3 The Impact of Cues on Textual Recognition . . . . .	5
1.2.4 Field Study . . . . .	6
1.3 Organization of Thesis . . . . .	7
1.4 Related Publications . . . . .	8
1.5 Other Publications . . . . .	8
2. Background . . . . .	11
2.1 Textual Password Schemes . . . . .	11
2.1.1 Traditional passwords . . . . .	11
2.1.2 Mnemonic Passwords . . . . .	12
2.1.3 System-assigned passwords . . . . .	12
2.1.4 PTP . . . . .	12
2.1.5 Cognitive questions . . . . .	13

2.2	Graphical Password Schemes . . . . .	13
2.2.1	Drawmetric . . . . .	13
2.2.2	Locimetric . . . . .	14
2.2.3	Cognometric . . . . .	15
2.3	Location-Passwords . . . . .	15
2.3.1	Other schemes . . . . .	16
2.4	Conclusion . . . . .	17
3.	CuedR: A Cued-Recognition Based User Authentication Scheme . . . . .	19
3.1	CuedR: System Design . . . . .	19
3.2	The Science behind CuedR . . . . .	21
3.2.1	Long-Term Memory . . . . .	21
3.2.2	Memory Retrieval . . . . .	22
3.2.3	Memory Cues . . . . .	23
3.3	CuedR: Through The Lens of Password Literature . . . . .	24
3.3.1	[1] Theoretical password space meeting the security policy of the intended domain . . . . .	25
3.3.2	[2] Avoiding exploitable reductions in security due to user choice of passwords . . . . .	25
3.3.3	[3] At least mild resistance to shoulder surfing and key logging, through variant response . . . . .	25
3.3.4	[4] Cues aiding memorability . . . . .	27
3.3.5	[5] Usability as close as possible to, or better than, textual passwords . . . . .	27
3.3.6	[6] Implicit feedback to legitimate users, when passwords are multi-part . . . . .	28
3.3.7	[7] Leveraging pre-existing user-specific knowledge where possible	28

3.4	User Study . . . . .	29
3.4.1	Participants, Apparatus and Environment . . . . .	29
3.4.2	Procedure . . . . .	29
3.4.3	Ecological Validity . . . . .	30
3.5	Results . . . . .	31
3.5.1	Memorability . . . . .	31
3.5.2	Registration and Login Time . . . . .	32
3.5.3	Impact of Portfolios on Usability . . . . .	32
3.5.4	User Perception on the Efficacy of Different Cues . . . . .	33
3.5.5	User Feedback on Usability and Applicability . . . . .	34
3.5.6	Pilot study: Memorability for Multiple CuedR Passwords . . . . .	36
3.6	Discussion . . . . .	37
3.6.1	CuedR: The Impact . . . . .	37
3.6.2	CuedR: The Acceptance . . . . .	38
3.6.3	CuedR: The Applications . . . . .	39
3.7	Conclusion . . . . .	40
4.	The Impact of Cues and User Interaction on the Memorability of System- Assigned Recognition-Based Graphical Passwords . . . . .	41
4.1	System Design . . . . .	41
4.1.1	Visual Memory . . . . .	42
4.1.2	Face Recognition . . . . .	43
4.1.3	Long-Term Memory . . . . .	43
4.1.4	Variant Response . . . . .	48
4.2	User Study . . . . .	49
4.2.1	Participants, Apparatus and Environment . . . . .	49
4.2.2	Procedure . . . . .	50



4.2.3	Ecological Validity . . . . .	51
4.3	Results . . . . .	51
4.3.1	Registration Time . . . . .	55
4.3.2	Login Time and Number of Attempts . . . . .	57
4.4	Discussion . . . . .	59
4.4.1	Cued-Recognition . . . . .	59
4.4.2	User Interaction . . . . .	61
4.4.3	Cued-Recognition vs. User Interaction . . . . .	62
4.4.4	Face recognition vs. Object Recognition . . . . .	63
4.5	Conclusion . . . . .	64
5.	The Impact of Cues on Textual Recognition . . . . .	65
5.1	System Design . . . . .	65
5.2	User Study . . . . .	67
5.2.1	Participants, Apparatus and Environment . . . . .	68
5.2.2	Procedure . . . . .	69
5.2.3	Ecological Validity . . . . .	70
5.3	Results . . . . .	70
5.3.1	Registration Time . . . . .	73
5.3.2	Login Time and Number of Attempts . . . . .	74
5.3.3	User Feedback . . . . .	75
5.4	Discussion . . . . .	76
5.5	Conclusion . . . . .	77
6.	Field Study: Exploring The Usability of GraphicV Scheme in a Real-life Scenario . . . . .	78
6.1	System Design . . . . .	78
6.2	Study Design . . . . .	79

6.3	Results and Analysis . . . . .	81
6.3.1	Overall Login Performance . . . . .	82
6.3.2	Training Effects . . . . .	84
6.3.3	User feedback . . . . .	87
6.4	Discussion . . . . .	88
6.5	Conclusion . . . . .	89
7.	Conclusion and Research Directions . . . . .	90
	REFERENCES . . . . .	93

## LIST OF ILLUSTRATIONS

Figure	Page
3.1 A partial screen shot during login. The facts corresponding to each keyword appear on the left side of the screen. The key is shown in parenthesis next to each keyword and also in the rightmost column of the table . . . . .	20
3.2 Illustration of cognitive memory model . . . . .	21
3.3 Box plot showing times for registration and logging in for each session	31
3.4 Responses to the question: “How often did the following cues assist you in recognizing keywords in CuedR?” . . . . .	33
4.1 The Study Model to Understand the Impact of Cues and User Interaction on Recognition-based Graphical Authentication . . . . .	42
4.2 A partial screen shot of ObjectSV scheme during login. The facts corresponding to each image appear below that image. Users enter the key, a lowercase letter shown in parentheses, in the password field (on top) to select the corresponding image. The keys are randomly assigned to images each time the portfolio is loaded, where no two images share the same key. During login, users are shown five such portfolios, where each presents a distinct set of 16 images including one of the five assigned images. . . . .	45
4.3 A partial screen shot of FaceSUI scheme during login. During login, users are shown five such portfolios, each presents a distinct set of 16 images including one of the five assigned images. . . . .	47

4.4	Login success rates for the study conditions [Number of participants=56]	52
4.5	Registration time for the study conditions . . . . .	55
4.6	Login time for the study conditions . . . . .	56
5.1	A partial screen shot of the <i>Control</i> 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. . . . .	66
5.2	A partial screen shot of <i>TextV</i> scheme during login. The facts corresponding to each keyword appear below that keyword. . . . .	67
5.3	A partial screen shot of <i>GraphicV</i> scheme during login. Each keyword is accommodated with the corresponding image. . . . .	68
5.4	Login success rates for the study conditions [Number of participants=52]	71
5.5	Registration time for the study conditions . . . . .	71
5.6	Login time for the study conditions . . . . .	73
6.1	<i>Field Study</i> : Number of logins by the participants, where a $(x, y)$ point represents the percentage of participants ( $y\%$ ) who conducted at least $x$ logins, either successful or unsuccessful. . . . .	79
6.2	<i>Field Study</i> : Login success rate (54 participants). . . . .	82
6.3	<i>Field Study</i> : Mean number of attempts. . . . .	82
6.4	<i>Field Study</i> : Login time. . . . .	83
6.5	<i>Field Study</i> : The change in login success rate over the sessions. . . . .	84
6.6	<i>Field Study</i> : The change in mean number of attempts over the sessions.	84
6.7	<i>Field Study</i> : The change in login time over the sessions. . . . .	85

## LIST OF TABLES

Table	Page	
3.1	Questionnaire responses for the usability of CuedR. Scores are out of 10. * indicates that scale was reversed. SD: Standard Deviation . . . . .	34
3.2	The applicability of CuedR for different online accounts. Scores are out of 10. . . . .	35
4.1	Number of Attempts for Successful Logins [SD: Standard Deviation] .	58
5.1	Number of Attempts for Successful Logins [SD: Standard Deviation] .	73
5.2	Questionnaire responses for the usability of each of the three schemes. Scores are out of 10. * indicates that scale was reversed. <i>Med</i> : Median, <i>Mo</i> : Mode . . . . .	75
6.1	<i>Field Study</i> : Registration time (seconds). SD: Standard Deviation . .	81
6.2	<i>Field Study</i> : Overall login performance [Login Sessions: 1349, Login Success Rate: 98%]. SD: Standard Deviation. . . . .	83
6.3	<i>Field Study</i> : Number of participants in the $x^{th}$ login session. . . . .	85
6.4	<i>Field Study</i> : Significance tests (Wilcoxon-Mann-Whitney tests) for login time between pairs of login sessions. We consider results comparing two conditions to be significantly different when we find $p < 0.05$ . . .	86
6.5	Questionnaire responses for the usability of GraphicV scheme. Scores are out of 10. * indicates that scale was reversed. <i>Med</i> : Median, <i>Mo</i> : Mode . . . . .	87

## CHAPTER 1

### Introduction

#### 1.1 Statement of the Problem

*“In the space of one hour, my entire digital life was destroyed”*, reported Matt Honan, technology writer and editor for Wired magazine, after several of his important online accounts were hijacked by attackers [1]. A recent study [2] found that 30% of users experienced account hijacking for at least one important online account, resulting in loss of privacy and reputation. It is widely believed that strong passwords could prevent account hijacking to a good extent [2].

Traditional user-chosen textual passwords suffer from security problem because of password reuse and predictable patterns [3, 4], in which users bear the responsibility of ensuring security for their online account through a password that should be chosen with creativity and intelligence so that it achieves satisfactory security and memorability. For many users, this is a lot of work, and in many cases they compromise on security and create a weak but memorable password.

A recent study [2] reveals that with the advancement of digital technology and widespread use of the Internet, users now better realize the importance of strong passwords than anytime before, and many of them intend to create secure passwords but still fail to achieve a good balance between security and memorability. We term this issue as *usability-security tension in user authentication*. Different password restriction policies have been deployed to get users to create stronger passwords [4, 5]. Such policies, however, do not necessarily lead to more secure passwords but do adversely affect memorability in some cases [6, 4].

We argue that the existing password systems fail to fully address users’ cognitive limitations or leverage humans’ cognitive strengths. Thus, despite a large body of research, it remains a critical challenge to build an authentication system that offers both high memorability and guessing resilience.

## 1.2 Contributions

We address this challenge in our research, and aim to understand how humans’ cognitive abilities could be leveraged to design more secure and memorable authentication schemes. To achieve the goal, we draw upon multiple theories from cognitive psychology, and examine the efficacy of memory cues and user interaction to improve the memorability of system-assigned random passwords. In this section, we highlight our major findings in the context of addressing usability-security tension in user authentication.

A number of important cognitive propensities have been considered by the researchers to explore better alternatives to traditional textual passwords. For example, recognition is an easier memory task than recall [7, 8, 9], and due to the *picture superiority effect*, the human brain is better at memorizing graphical information as compared to textual information [10, 11]. These are the core ideas behind the design of recognition-based graphical passwords, such as Passfaces [12], which is now commercially available and deployed by a number of large websites.<sup>1</sup>

In recognition-based graphical passwords, such as Passfaces [12], users are shown several portfolios of pictures of faces (e.g., four portfolios of nine faces each), and one face per portfolio serves as the authentication secret that they have to recognize during login. Previously, users in Passfaces could select images from the portfolio for their authentication secret. Davis et al. [13], however, found that users select predictable

---

<sup>1</sup><http://www.realuser.com/> shows testimonials about Passfaces from customers.

images. As a result, the commercial Passfaces [12] product now 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. Unfortunately, it is difficult for most people to memorize system-assigned passwords [14, 15, 16]. The commercial deployment of recognition-based graphical passwords (e.g., Passfaces [12]) and its demonstrated potential mean that improvements to such schemes would be very valuable contributions.

### 1.2.1 CuedR

We argue that while the system assigns random passwords, it should also help with memorization and recall. We investigate the feasibility of this approach with *CuedR*, a novel *cued-recognition* authentication scheme that provides users with multiple cues (visual, verbal, and spatial) and lets them choose the cues that best fit their learning process for later recognition of system-assigned keywords. In our study, we found a 100% memorability for CuedR over the span of one week, and 84% of participants preferred to use it in real life as a replacement to traditional textual passwords.

The lab study on CuedR showed promise for providing multiple cues to aid recognition, however, the study did not examine the individual impact of each cue. Thus, we plan to perform a separate study to explore deeper into this issue, where we examine the efficacy of *spatial cues* (fixed position of images), *verbal cues* (phrases/facts related to the images), and employing *user interaction* (learning images through writing a short description at registration) to improve the memorability of system-assigned passwords based on face images and object images.



### 1.2.2 The Impact of Cues and User Interaction on Graphical Recognition

The use of cues facilitates a detailed encoding that helps to transfer the authentication information (e.g., assigned images) from the working memory to long-term memory at registration [17], helping users recognize their images when logging in later. We also explore the efficacy of requiring user interaction at registration, in which we have users apply their observation and imagination to type a short description about assigned images. In the course of such observations and thinking on the assigned images, users get more familiar with them and consequently succeed to recognize those images from the set of decoys during login. This process engages users' action-event memory [18], in addition to their visual memory [10, 11], and aids in the elaborate encoding of the authentication secret in long-term memory [17] (see §4.1 for details).

The commercially deployed Passfaces scheme uses face images instead of object images, and it is unclear which should be used. We examine this issue in our study. Considering both human faces and objects as images, along with cues and interaction, we design seven different study conditions (see Figure 4.1). In our within-group study with 56 participants, every participant was assigned seven different graphical passwords, each representing one study condition. One week after registration, participants had a 98% login success rate for a scheme, *ObjectSV* offering users with spatial and verbal cues for object images, while the scheme based on user interaction had a 95% login success rate for face images and a 93% login success rate for object images. All of these were significantly higher than the control conditions representing existing graphical password schemes. The major findings from our study include:

- Verbal cues make a significant contribution in improving the memorability for object-recognition-based graphical passwords.

- Spatial cues do not contribute significantly to improve memorability for either face or object recognition.
- User interaction is an effective approach to enhance memorability for both face and object recognition.

### 1.2.3 The Impact of Cues on Textual Recognition

While verbal cues made an important contribution in improving the memorability for system-assigned graphical passwords, we conduct another study to examine the impact of verbal cues in enhancing the usability for textual recognition. To achieve the goal, we design a scheme, *TextV*: **T**extual 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. So, we design another scheme, *GraphicV*: **G**raphical Recognition with **V**erbal cues, and compare it with the *TextV* scheme <sup>2</sup>. To the best of our knowledge, no study yet has compared textual and graphical recognition-based schemes in terms of usability.

The study of Wright et al. [15] 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 (i.e., offering spatial cues) would improve the memorability for recognition-based schemes and suggested the approach to be examined in future work. We adopt suggestion of Wright et al. [15] to design our study conditions by offering spatial cues. In our within-group study with 52 participants,

---

<sup>2</sup>GraphicV scheme is same as the ObjectSV scheme

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. [15], keeping the position of keywords fixed in a portfolio (i.e., offering spatial cues) did not provide a satisfactory login success rate for textual recognition (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 §5.1), 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.

#### 1.2.4 Field Study

Based on the findings from these lab studies, it is clear that GraphicV (i.e., ObjectSV) scheme offering users with spatial and verbal cues for object images performed best in terms of memorability. So, we conduct a field study to further examine the usability of this scheme in a real-life scenario. Although a field study is challenging to perform because of the resources and time required, they offer strong ecological validity and the best measure of login performance in a realistic setting [19]. We conducted a 74-day-long field study, including 1349 login sessions from 54 participants. Our results show that GraphicV scheme offered satisfactory memorability in our real-world context, with an overall login success rate of 98%.

In this field study, we analyze the login performance distribution among users for a detailed understanding of the usability issues. A training effect occurs when

users get better at entering their password over time. We give an insight into this issue by examining the change in login performance over login sessions, where we identified an overall improvement in login performance with more login sessions, with some modest fluctuations.

### 1.3 Organization of Thesis

We organize this thesis proposal as follows:

- **Chapter 2** gives an overview of the notable textual, graphical, and geographic location-password schemes, and shows why the existing schemes are insufficient to address the usability-security tension in user authentication.
- **Chapter 3** presents a novel cued-recognition based authentication scheme, CuedR, where we explain the design features of this scheme from the perspective of cognitive psychology and existing password studies, followed by reporting the results of a lab-based user study.
- **Chapter 4** describes our study model and the corresponding results revealing the impact of memory cues and user interaction on the memorability of system-assigned recognition-based graphical passwords.
- **Chapter 5** presents the results showing the impact of verbal cues on textual recognition-based system-assigned random passwords and the usability benefits of adding images for a textual recognition-based scheme that offers users with verbal cues.
- **Chapter 6** describes the study procedure and results of a real-life field study examining the usability of system-assigned recognition-based graphical password that provides users with spatial and verbal cues for object images.

#### 1.4 Related Publications

- **Mahdi Nasrullah Al-Ameen**, Kanis Fatema, Matthew Wright, and Shannon Scielzo. Leveraging Real-Life Facts to Make Random Passwords More Memorable. In European Symposium on Research in Computer Security (ESORICS). September 2015 [20].
- **Mahdi Nasrullah Al-Ameen**, Kanis Fatema, Matthew Wright, Shannon Scielzo. The Impact of Cues and User Interaction on the Memorability of System Assigned Recognition-Based Graphical Passwords. In Symposium on Usable Privacy and Security (SOUPS). July 2015 [21].
- **Mahdi Nasrullah Al-Ameen**, Matthew Wright, Shannon Scielzo. Towards Making Random Passwords Memorable: Leveraging Users Cognitive Ability Through Multiple Cues. In 33rd ACM Conference on Human Factors in Computing Systems (CHI). April 2015 [22].

#### 1.5 Other Publications

- **Mahdi Nasrullah Al-Ameen**, Matthew Wright. Multiple-Password Interference in the GeoPass User Authentication Scheme. In NDSS Workshop on Usable Security (USEC). February 2015 [23].
- Ruj Akavipat, **Mahdi N. Al-Ameen**, Apu Kapadia, Zahid Rahman, Roman Schlegel, and Matthew Wright. ReDS: A Framework for Reputation-Enhanced DHTs. In IEEE Transactions on Parallel and Distributed Systems (TPDS), Special Issue on Trust, Security, and Privacy. Vol. 25, No. 2, pp. 321-331. 2014 [24].

- **Mahdi N. Al-Ameen**, Matthew Wright. Design and Evaluation of Persea, a Sybil-resistant DHT. In ACM Symposium on Information, Computer and Communications Security (AsiaCCS). June 2014 [25].
- **Mahdi Nasrullah Al-Ameen**, Matthew Wright. Persea : A Sybil-resistant Social DHT. In Poster Session: ACM Conference on Data and Application Security and Privacy (CODASPY). February 2013 [26].
- **Mahdi Nasrullah Al-Ameen**, Charles Gatz, and Matthew Wright. SDA-2H: Understanding the Value of Background Cover against Statistical Disclosure. In Journal of Networks Special Issue on Selected Papers from ICCIT 2011, Vol. 7 No. 12, pp. 1943-1951. December 2012 [27].
- **Mahdi Nasrullah Al-Ameen**, Mehrab Shahriyar, A. Billah. Time and Space Efficient Algorithm for Consumer's Priority Product Management. In International Conference on Computer and Information Technology (ICCIT). December 2012 [28].
- **Mahdi Nasrullah Al-Ameen**, Charles Gatz, Matthew Wright. SDA-2H : Understanding the Value of Background Cover Against Statistical Disclosure. In International Conference on Computer and Information Technology (ICCIT). December 2011 [29].
- **Mahdi Nasrullah Al-Ameen**, Md. Monjurul Hasan, Asheq Hamid, Making FindBugs More Powerful. In International Conference on Software Engineering and Service Sciences (ICSESS). July 2011 [30].
- **Mahdi Nasrullah Al-Ameen**, Matthew Wright. iPersea: The Improved Persea with Sybil Detection Mechanism. Technical Report, arXiv: 1412.6883. December 2014 [31].

- **Mahdi Nasrullah Al-Ameen**, Matthew Wright. A Comprehensive Study of the GeoPass User Authentication Scheme. Technical Report, arXiv:1408.2852. August 2014 [32].
- **Mahdi Nasrullah Al-Ameen**, S M Taiabul Haque, Matthew Wright. Q-A: Towards the Solution of Usability-Security Tension in User Authentication. Technical Report, arXiv:1407.7277. July 2014 [33].
- **Mahdi Nasrullah Al-Ameen**. An Intelligent Fire Alert System Using Wireless Mobile Communicaiton. Technical Report, arxiv: 1308.0327. August 2013 [34].

## CHAPTER 2

### Background

In this section, we give an overview of notable textual and graphical password schemes in which we highlight why existing schemes are insufficient. We also discuss about location-passwords, a recent inclusion in password studies. We end by summarizing the main points and arguing what the next steps should be in the study of password designs.

#### 2.1 Textual Password Schemes

##### 2.1.1 Traditional passwords

Traditional user-chosen textual passwords are fraught with security problems and are especially prone to password reuse and predictable patterns [3, 4, 35]. Ur et al. [36] showed that misconceptions of users contribute to creating weak passwords. For example, many users believe that adding a special character at the end of a password instantly makes it secure [36]. The study [36] also showed that users could anticipate only the targeted guessing attack, believing that it is a secure approach to use birthday or name as a password if those data are not available on social networking site (e.g., Facebook). In a recent study [37], Ur et al. showed that users have serious misconceptions about the impact of basing passwords on common phrases and including digits and keyboard patterns in passwords, which lead them to creating weak and predictable authentication secrets.

Different password restriction policies have been deployed to get users to create stronger passwords [35, 4, 38]. The studies. [4, 38] report, however, that such policies



do not necessarily lead to more secure passwords, but they do adversely affect memorability in several cases. In another study [39], the authors found that multi-step password-creation process through providing guidance to users is not effective enough in creating strong passwords.

### 2.1.2 Mnemonic Passwords

Kuo et al. [40] 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 [40] 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 [40].

### 2.1.3 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 [14, 15]. Wright et al. [15] 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.

### 2.1.4 PTP

Forget et al. [41, 42] 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 [42].

### 2.1.5 Cognitive questions

Furnell et al. [43] revealed the potential of cognitive questions and reported a high level of user satisfaction in using that for primary authentication. However, Just and Aspinall [44] identified the usability and security problems of using cognitive questions for authentication, and several other studies [45, 46] reported the vulnerability of this approach to targeted guessing attacks.

## 2.2 Graphical Password Schemes

Graphical password schemes can be divided into three categories [19], based on the kind of memory leveraged by the systems: i) Drawmetric (recall-based), ii) Locimetric (cued-recall-based), and iii) Cognometric (recognition-based).

### 2.2.1 Drawmetric

The user is asked to reproduce a drawing in this category of graphical passwords. In *Draw-a-Secret (DAS)* [47], a user draws on top of a grid, and the password is represented as the sequence of grid squares. Nali and Thorpe [48] 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* [49] 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 [19]. DAS and BDAS have recall rates of no higher than 80% when users have to remember a single

drawmetric password. To the best of our knowledge, no multiple-password study was conducted on drawmetric graphical passwords.

### 2.2.2 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* [50] scheme, users select a sequence of click-points on a single image as their password. *Cued Click-Points (CCP)* [51] is a modified version of *Passpoints*, where users sequentially choose one click-point on each of five images. Dirik et al. [52] 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)* [53, 54], 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. [55] 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 [55] indicate that predictability is still a security concern for PCCP.

In a multiple-password study on locimetric graphical passwords (e.g., *Passpoints*) [56], the authors reported that 57% (15/26) of the participants were able to recall their graphical passwords successfully after two weeks of registration, and the average login time varied between 18 to 47 seconds.

### 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* [12] is the most studied cognometric scheme as it is commercially deployed by a number of large websites. The commercial Passfaces [12] 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. Everitt et al. [16] show that users have difficulty in remembering system-assigned Passfaces. As reported in the study [16], the participants accessing four different facial passwords each week had a failure rate of 15.23% after a month, when each password was used once a week. Here, the average login time was 29.7 seconds.

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].

## 2.3 Location-Passwords

Geographic location-passwords is a recent category in password research. In these schemes, users select one or more locations in an online map (e.g. Google Maps) as their password. To the best of our knowledge, three schemes [57, 58, 59] in this password-category have been proposed to date, where GeoPass [57], proposed by Thorpe et al. in 2013, shows most potential in terms of usability and security.

In GeoPass, the user’s password is a single location on an online map (Google Maps). This secret location, known both as the *location-password* and just *geopass*<sup>1</sup>, is selected by the user at registration by right-clicking on the map. The search bar helps to make navigation faster by enabling the user to type the name of a place. Also, typing leads to a drop-down menu suggesting locations in which the searched item may appear. Zooming and panning are also enabled via the Google Maps API. Using the convention that a higher-numbered zoom level represents being zoomed in closer, the initial zoom level is 2, and GeoPass allows the user to click on a location at a minimum zoom level of 16. A successful login requires the users to click within a 21x21 pixel box around the location-password they had set. We refer readers to Thorpe et al.’s paper [57] for in-depth discussion on the features of GeoPass.

Thorpe et al. [57] conducted a nine-day-long user study on GeoPass with three sessions: two in a lab-setting and one online. The login success rate was 97%, and the median login time was found to be no greater than 30 seconds [57]. The security analysis [57] showed that the theoretical password space for GeoPass is  $2^{36.9}$ , such that only 11% of online guessing attacks might be successful after allowing for  $2^{16}$  guesses. In this respect, reasonable lockout rules [60] should make GeoPass sufficiently resilient against such attacks. The user feedback on GeoPass was encouraging, and a number of users showed interest to use the scheme in real life [57].

### 2.3.1 Other schemes

There are two other schemes that use map locations as an authentication secret: one proposed by Spitzer [59] and another one is called PassMap [58]. PassMap requires the user to choose two locations and the scheme by Spitzer [59] requires five or seven locations at different zoom levels to be selected as the location-password. Thorpe

---

<sup>1</sup>in lowercase to avoid confusion with the system name

et al. [57] have shown that GeoPass is more usable than other digital-map-based schemes [58, 59] because of its requirement to click on a single location and normalized error tolerance to a given zoom level. The login success rate in GeoPass (97%) was found to be higher than that in PassMap (92.59%).

Although GeoPass showed a high potential in a lab study, it is still unknown how usable the scheme would be in a real-life setting, and what would be the memorability for multiple location-passwords. As noted by Biddle et al. [19], multiple-password interference is a major usability concern, and only a handful of graphical password schemes have been evaluated with an interference study. To the best of our knowledge, no field study or interference study has been conducted yet on geographic location-passwords.

## 2.4 Conclusion

As we see from the above discussion, textual password schemes that provide better memorability are vulnerable to guessing attacks (e.g., traditional and mnemonic passwords). On the other hand, textual passwords with higher resilience against guessing attacks suffer from memorability problems (e.g., system assigned passwords, PTP).

Graphical passwords are found to provide better memorability than textual passwords [56]. However, when users are allowed to create their graphical passwords, they choose predictable images or patterns [48, 52, 55, 13]. If the images are assigned by the system to improve security, memorability is decreased [16]. Password managers fail to provide a suitable solution, as they suffer from usability (in implementation) and security (e.g., single point of failure) problems. Additionally, two recent papers extensively examine security problems in a range of password managers [61, 62].

Location-passwords show potential to address this usability-security tension, however, the memorability for multiple location-passwords is not examined yet.

We argue that existing password systems fail to fully address users' cognitive limitations or leverage humans' cognitive strengths. Thus, despite a large body of research, it remains a critical challenge to build an authentication system that offers both high memorability and guessing resilience. We address this challenge in our research, and aim to understand how humans' cognitive abilities could be leveraged to design more secure and memorable authentication schemes.

## CHAPTER 3

### CuedR: A Cued-Recognition Based User Authentication Scheme

In this chapter, we propose a cued-recognition based authentication scheme: CuedR, which provides users with multiple cues to help recognizing the system-assigned keywords. In §3.1, we describe the system design of CuedR, and explain our design choices from two different perspective: cognitive psychology (see §3.2) and the findings from existing password literature (see §3.3). In §3.4, we describe our study procedure to evaluate CuedR in a lab setting, and present the results of our user study in §3.5, followed by a discussion on our findings in §3.6.

#### 3.1 CuedR: System Design

In CuedR, six keywords are randomly assigned to the user each from a distinct portfolio (e.g., animals, fruits, or vehicles), where each portfolio presents 26 keywords. To aid memorability, our scheme offers graphical, verbal, and spatial cues corresponding to each keyword. In particular, each keyword (e.g., “Zebra”) has an image (a picture of a zebra), a number and a phrase related to the keyword (“2. Each zebra has a unique pattern of stripes.”), and the position of both the image and phrase are fixed in both absolute terms and relative to the other images and phrases (Zebra is at the top between Cheetah and Elephant). See Figure 3.1 for a screenshot illustrating these features.

Each time a portfolio is loaded, each of the 26 lowercase letters **a-z** is assigned randomly as a *key* to one keyword on the page. The user inputs the key letter corresponding to her keyword into a single-character password field to move on to



No.	Facts	Key
1	Cheetah is faster than any other land animal	w
2	Every zebra has unique pattern of stripes	i
3	Elephants can get sunburned	t
4	Rabbits have a near 360-degree vision & can see behind them	z
5	Longhorn Cattles have become the symbol of Texas	g
6	A sheep named Dolly was the first cloned mammal	r
7	Penguins lost flying-ability million years ago	b
8	Koalas have similar fingerprints to humans	n
9	On average, cats spend 2/3 of everyday sleeping	y
10	Giraffe can weigh as much as a pick up truck	j
11	Turtles are cold blooded	x
12	Panda is the symbol of peace in China	u
13	Roar of lion can be heard from 5 miles away	c
14	A male deer is called 'buck'	e
15	Hippos eat mostly grass	o










		
1. Cheetah (w)	2. Zebra (i)	3. Elephant (t)
		
7. Penguin (b)	8. Koala (n)	9. Cat (y)
		
13. Lion (c)	14. Deer (e)	15. Hippopotamus (o)

Figure 3.1. A partial screen shot during login. The facts corresponding to each keyword appear on the left side of the screen. The key is shown in parenthesis next to each keyword and also in the rightmost column of the table.

the next portfolio. The key letter changes every time to provide the *variant response* property [19]. Schemes with this property have been shown to provide higher resilience to shoulder surfing and simple keystroke loggers than schemes like traditional textual passwords in which the same letters are entered at every login [63, 19].

**User authentication.** At user registration in CuedR, the system randomly selects six portfolios (without replacement) and the user is assigned one keyword from each of these portfolios. The user enters the key corresponding to the assigned keyword to get forwarded to the next portfolio. In case of a wrong entry she will immediately be informed about the error and will need to enter the correct one. During login, the user recognizes her system-assigned keyword from the portfolio and enters the key corresponding to that keyword into a small password field. A successful authentication requires the user to correctly enter keys for all six of her assigned keywords.

When the user makes a mistake during login, CuedR shows the user a portfolio that is different from her next assigned portfolio. A legitimate user can recognize this *implicit feedback* as an indicator that something was wrong and that she should

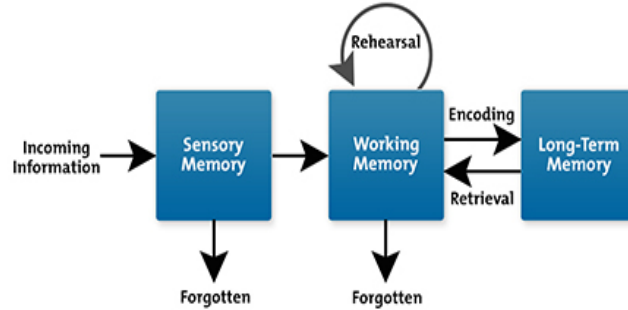


Figure 3.2. Illustration of cognitive memory model.

go back and correct the mistake, while an attacker will not know which portfolios are correct. Implicit feedback is thus a desirable feature to enhance usability when passwords have multiple parts [19].

**Password storage.** For secure storage of the user’s authentication secret, the six keywords can be concatenated together with a salt and hashed using a slow hash function like bcrypt [64] or PBKDF2 [65]. Implicit feedback can be implemented by making the selection of the next portfolio a function of the current portfolio and the keyword selected, which is independent of the correctness of the responses. Thus, correctness only needs to be checked after all keys have been entered.

### 3.2 The Science behind CuedR

In this section, we explain from the perspective of cognitive psychology how our design choices are set up to provide high memorability.

#### 3.2.1 Long-Term Memory

We incorporate the scientific understanding of long-term memory to advance the scheme’s usability properties. According to the cognitive memory model proposed by Atkinson and Shiffrin [17], 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. In this respect, an elaborative encoding would take place if the information can be associated with something meaningful, such as *cues*. This encoding helps people to remember and retrieve the processed information efficiently over an extended period of time (see the illustration in Figure 3.2).

In CuedR, users focus their attention to learn keywords through associating them with the corresponding cues, which should help to process and encode the keywords in memory and store them in the LTM. The cues would assist the user to recognize the keywords in the future, which should enhance their memorability.

### 3.2.2 Memory Retrieval

We designed CuedR to require users to perform a recognition task. 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) [7] and have developed two main theories to explain this: *Generate-recognize theory* [8] and *Strength theory* [9].

Generate-recognize theory [8] 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 [9] 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.

### 3.2.3 Memory Cues

Psychology research [8, 7] 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* [66] postulates that the most effective cues are those that are present at the time of remembering. In CuedR, cues are provided during registration, i.e., the learning period, and also at login.

#### 3.2.3.1 Why use multiple cues?

In CuedR, the user has five different references (cues) that she can leverage to learn the keyword: (i) the image, (ii) the number from 1 to 26, (iii) the phrase or fact associated with the keyword, (iv) the absolute positions of the keyword, the image, and the phrase/fact, and (v) the positions of the keyword, image and phrase/fact relative to the other keywords, images and phrases. This combination brings together graphical (images), spatial (positions), and verbal (facts, numbers) information. Thus, a user may focus on just those cues that she finds most appropriate to her learning process, while the other cues may provide additional support for memorability.

**Graphical cues.** Psychology research [10, 11] reveals that the human brain is better at memorizing graphical information as compared to textual information. This is known as the *picture superiority effect*, which motivates us to include graphical cues (images) in our scheme.

**Verbal and spatial cues.** While images are generally effective cues, not all users may have a strong visual memory. Additionally, many graphical password schemes require good vision and motor skills, which elderly users [67] may lack. Thus, we provide verbal and spatial cues in addition to graphical cues to let users leverage their cognitive ability in memorizing the keywords. Yan et. al. [68] examined the influence of phrases in increasing the memorability of passwords, which inspires us to accommodate a common phrase or fact for each keyword as a verbal cue.

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 [12]. Semantic priming thus eases the recognition task [12]. In CuedR, the keywords and cues in a portfolio remain same and presented at a fixed position whenever that portfolio is loaded, which establishes a relationship between them and reinforces semantic priming.

### 3.3 CuedR: Through The Lens of Password Literature

Through a comprehensive survey on 25 different graphical password schemes, Biddle et al. [19] identified seven features that should be offered by an ideal graphical password system. The authors [19] state, “We expect tomorrow’s ideal graphical password systems may have many of the following desirable characteristics, reflecting lessons learned from proposals to date.”

In this section, we analyze how CuedR addresses these seven features and explain our design choices based on the findings from the literature on passwords.

### 3.3.1 [1] Theoretical password space meeting the security policy of the intended domain

Well-known recognition-based schemes, such as Passfaces [12] and Story [13], originally provided no more than 13 bits of theoretical entropy. Later, Hlywa et al. [69] conducted a study on recognition based graphical passwords with 20 bits of entropy, since 20 bits of entropy with reasonable lockout rules is considered sufficient to prevent online brute-force attacks [5]. In CuedR, we use more than 20 bits of entropy, in particular 28 bits, to maintain comparability with prior studies on system-assigned passwords [15]. During login in CuedR, a user has to recognize her keyword from a portfolio of 26 distinct keywords. This is required six times, once for each portfolio. The password space is thus  $\log_2(26)^6 \approx 28$  bits. CuedR can be used with a range of entropy values by varying either the number of keywords or portfolios.

### 3.3.2 [2] Avoiding exploitable reductions in security due to user choice of passwords

Statistical password distributions are often not equiprobable due to scheme-dependent predictability of user choices [70]. In CuedR, passwords are randomly assigned by the system, which provides two security benefits. First, the effective password space in CuedR is same as the theoretical space. Second, the system gains user-choice resilience and thus provides robustness against online guessing attacks that exploit password reuse, personal information and predictable strategies [3, 4].

### 3.3.3 [3] At least mild resistance to shoulder surfing and key logging, through variant response

Variant response refers to varying how the password is entered across different login sessions, which is an important feature to offer robustness against shoulder surfing and keystroke loggers [19].

### 3.3.3.1 Shoulder surfing

It is difficult in practice to observe both keyboard and monitor at the same time. Thus, graphical password schemes that include the variant response feature with keyboard entry provide higher resilience to shoulder surfing compared to traditional textual passwords and graphical passwords with mouse input [63]. In a study by Tari et al. [63], participants playing the role of shoulder surfers were able to gain 73% of non-dictionary passwords, 26% of dictionary passwords, and 62% of graphical passwords with mouse input, but just 11% of graphical passwords with variant response.

CuedR offers the variant response feature, where the user enters a key corresponding to her keyword using the keyboard, and watching only keyboard entries is not sufficient for a shoulder surfer, as the key associated with each keyword changes with every login attempt. The entered key is shown as an asterisk or dot (as with a regular password) to minimize the risk of shoulder surfing.

Variant response does not protect against an attacker who can use a video camera to record both the monitor and keystrokes at the same time, and attackers may gain the user's credentials when they are assigned during registration. Thus, we only claim that CuedR provides mild resistance to shoulder surfing through variant response, which conforms to the desired level of security in this regard [19]. We recommend that users register in a secure environment (e.g., avoiding public terminals) to ensure better security against shoulder surfing.

### 3.3.3.2 Keystroke and mouse loggers

Keystroke loggers record keyboard input and mouse loggers capture mouse actions to make the user's credentials available for retrieval by remote attackers [19].

Biddle et al. [19] state that a system provides resilience against keystroke/mouse loggers when the keyboard/mouse entries for authentication vary across subsequent login sessions. Thus, the variant response feature in CuedR offers better resilience against basic keystroke loggers compared to a password system where the same letters are entered during every login session. CuedR is clearly resilient to mouse loggers, as it does not use mouse input.

### 3.3.4 [4] Cues aiding memorability

While different variants of system-assigned passwords failed to provide satisfactory memorability [15, 41, 42, 14], CuedR achieves a good memorability through associating each keyword with a set of cues and letting users choose the appropriate one(s) to their learning process<sup>1</sup>. We describe the basis for the effectiveness of cues in the previous section, and we report on user perceptions of different cues in the results section.

### 3.3.5 [5] Usability as close as possible to, or better than, textual passwords

Shay et al. [38] performed a comprehensive study on the memorability of user-chosen textual passwords following different composition policies, where *basic12* was the simplest form of passwords in which the user had to create a password of at least 12 characters without any composition requirements or dictionary check. Participants reported the least difficulty to create and remember a *basic12* password. After two days, 86% of participants who wrote down their password could log in (76% on the first attempt), while 75% of participants who did not write down their password could log in (61% on the first attempt). For CuedR, after one week, we found that 100%

---

<sup>1</sup>We note that direct comparison between different studies should be taken with caution



of participants could log in (89% on the first attempt). So, we see that memorability is better than textual passwords with moderate security requirements.

Although the login time for CuedR is high compared with traditional textual passwords, users mostly disagreed with the notion that the scheme is too time consuming (see Table 3.1). Overall, users reported satisfaction with the usability of CuedR and 84% preferred to use it in real life as a replacement to traditional textual passwords.

### 3.3.6 [6] Implicit feedback to legitimate users, when passwords are multi-part

Implicit feedback instantly notifies a user when she makes a mistake, instead of showing her an error message at the end of all entries. Due to its implicit nature, this feedback should only be recognizable and useful to the legitimate user. An attacker who does not know about a user’s portfolios must make all six guesses in CuedR to learn whether he has succeeded or not. Implicit feedback has already been shown to have satisfactory user acceptance in a cued-recall based scheme [51]. To accommodate this feature in CuedR, we build distinct portfolios of images (i.e., “animals”, “fruits”, “flowers”, etc.) so that a user can clearly distinguish among the portfolios at a glance and quickly realize her mistake.

### 3.3.7 [7] Leveraging pre-existing user-specific knowledge where possible

Leveraging pre-existing user-specific knowledge, for example answering cognitive questions or recognizing personal images from decoys could make the scheme vulnerable to targeted guessing attacks (e.g., guessing by acquaintances). So, we did not include this feature in CuedR to ensure security. Since CuedR offers good memorability, it is not clear if user-specific knowledge is required, though it could help to reduce the cognitive burden on users. Exploring ways to securely leverage

user-specific knowledge for authentication could be an interesting venue for future work.

### 3.4 User Study

We now present the design of our user study to evaluate the usability and memorability of CuedR. The study procedures were approved by our university's Institutional Review Board (IRB) for human subjects research.

#### 3.4.1 Participants, Apparatus and Environment

For this experiment, we recruited 37 students (25 women, 12 men) through our university's Psychology Research Pool. Participants came from diverse backgrounds, including majors from Nursing, Psychology, Business, Political Science, Biology, Physical Science, and Social Work. The age of the participants varied between 17 to 30 with a mean age of 21. They make regular use of the Internet and websites that require authentication. 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 user's interaction with the system. For this study, we built 18 different portfolios (e.g., animals, fruits, flowers, and vehicles), and collected the images (graphical cues) and phrases/facts (verbal cues) from free online resources.

#### 3.4.2 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 memorization of the password. Note that the one-week delay is larger than the maximum average interval

for a user between her subsequent logins to any of her important accounts [71]. One week is also a common interval used in authentication studies (e.g., [72, 15, 49]).

*Session 1.* After signing a consent form, the participants performed a practice trial with CuedR to compensate for novelty effect. We did not collect data for this practice trial. At registration, six portfolios were randomly chosen by the system and a user was assigned at most one keyword from each portfolio. Then participants were asked to spend 60 seconds in completing a mental rotation test (MRT) puzzle shown on the computer screen, which helps to clear their working memory [73]. Participants were then given questionnaire that gathered demographic information, and were asked to log into the same site with CuedR (*login 1*). They were asked to not write down their authentication secrets.

*Session 2.* The participants returned after one week of registration, and logged into the site using CuedR (*login 2*). After they had finished, we conducted an anonymous paper-based survey. Participants were then compensated and thanked for their time.

### 3.4.3 Ecological Validity

Our participants were young and university educated, which represents a large number of frequent Web users, but may not generalize to the entire population. They came from diverse majors including Nursing, Psychology, Physical Science, Business, etc. As the study was performed in a lab setting, we were only able to gather data from 37 participants. We believe that 37 provides a suitable sample size for a lab study as compared to the prior studies on password memorability [57, 54, 51, 74].

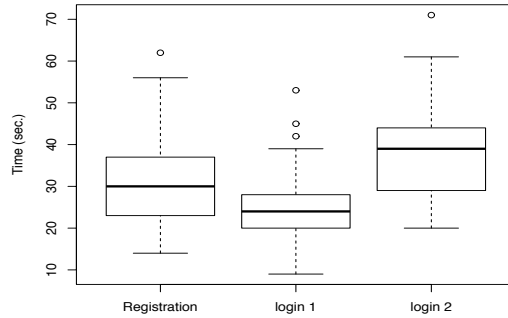


Figure 3.3. Box plot showing times for registration and logging in for each session.

### 3.5 Results

In this section, we discuss the results of our user study. We label the login performance of participants in session 1 and session 2 as *login 1* and *login 2*, respectively. We evaluated the usability of CuedR via all metrics suggested in the literature [75]: memorability, login time, number of login attempts, and user feedback. In addition, we analyzed the impact of portfolios on login performance and user perceptions on the effectiveness of different cues. We also discuss the results of pilot study on the memorability of multiple CuedR passwords.

#### 3.5.1 Memorability

We observed a 100% login success rate for CuedR in both *login 1* and *login 2*. In *login 1*, all the participants successfully recognized the keywords on the first attempt. In *login 2*, 89% of participants succeeded on the first attempt to recognize all six keywords. The other four participants (11%) recognized five of out of six keywords on the first attempt. Three participants corrected their mistake on the second attempt, and the other participant succeeded on the third attempt.

### 3.5.2 Registration and Login Time

The registration and login time for CuedR are illustrated in Figure 3.3. The mean time for registration was 31.2 seconds (median: 30 seconds, SD: 10.5 seconds). The mean time for successful login were 25.7 seconds (median: 24.0 seconds, SD: 8.3 seconds) in *login 1*, and 38.0 seconds (median: 39.0 seconds, SD: 11.4 seconds) in *login 2*. A paired-samples t-test reveals that login time in *login 1* was significantly less than that in *login 2*,  $t(36) = 7.81$ ,  $p < 0.01$ . This was expected, as participants performed *login 1* shortly after learning the keywords. To note, the reported registration and login time include the time to download images.

The login time in CuedR is in line with that in prior recognition based schemes offering 28 bits of entropy [69, 15]. We note that our results for login time are likely conservative, since they measure initial use. A recent field study [32] reveals that login time decreases with the frequent use of a scheme due to training effects. These findings are in agreement with our user feedback, where the participants reported that with practice, they could quickly recognize the keywords (see Table 3.1).

### 3.5.3 Impact of Portfolios on Usability

In our study, all the participants succeeded to recognize their keywords irrespective of the type of portfolios in both *login 1* and *login 2*. In *login 1*, no participant made any mistake in any portfolio, and thus there was no difference among portfolios for the number of attempts to succeed. In *login 2*, four participants (11%) required multiple attempts to succeed (see the results for *Memorability*), where one-way ANOVA test results show that there was no significant difference among portfolios in terms of the number of attempts required to successfully recognize the keywords,  $F(17, 220) = 1.16$ ,  $p = 0.31$ . In addition, we conducted a post-hoc pairwise compari-

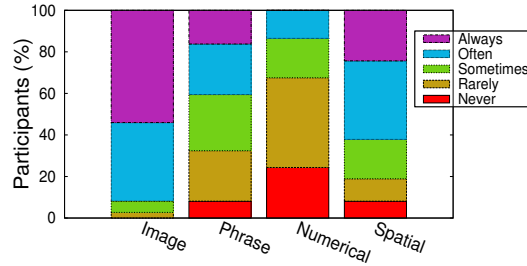


Figure 3.4. Responses to the question: “How often did the following cues assist you in recognizing keywords in CuedR?”.

son using Tukey’s HSD (Honestly Significant Difference), which reveals no significant difference between any pair of portfolios for the number of attempts to succeed.

Our one-way ANOVA test results demonstrate that there was no significant difference among different portfolios in terms of the time to learn the keyword during registration,  $F(17, 220) = 0.76$ ,  $p = 0.71$ , or recognize the keyword either in *login 1*,  $F(17, 220) = 1.16$ ,  $p = 0.31$ , or in *login 2*,  $F(17, 220) = 0.59$ ,  $p = 0.87$ . In addition, we conducted a post-hoc pairwise comparison using Tukey’s HSD, which did not find any significant difference between any pair of portfolios in either registration time or in login time. These findings indicate that the usability in recognizing keywords did not vary significantly across different portfolios used in our study.

### 3.5.4 User Perception on the Efficacy of Different Cues

To understand user perception on the importance of different cues in aiding recognition, we asked them at the end of second session, “How often did the following cues assist you in recognizing keywords in CuedR?” In response, for each cue they selected one of five options: *Never*, *Rarely*, *Sometimes*, *Often*, or *Always*. Our results show that participants report using multiple cues to varying degrees to help recognize their keywords (see Figure 3.4). In particular, 92% of participants reported that the images were *always* or *often* helpful to recognize keywords, while 62%, 40%, and 14%

Table 3.1. Questionnaire responses for the usability of CuedR. Scores are out of 10. \* indicates that scale was reversed. SD: Standard Deviation

Questions	Mode	Median	Mean	SD
I could easily sign up with CuedR	10	9.0	9.0	1.3
The login using CuedR was easy	10	10.0	9.5	0.7
Keywords are easy to remember in CuedR	10	10.0	9.4	0.8
*I found CuedR too time-consuming (i.e., I did not find CuedR too time consuming)	10	7.0	6.4	2.6
With practice, I could quickly enter my password in CuedR	10	10.0	9.8	0.6
I could easily use CuedR every day	10	9.0	8.8	1.3
I could easily use CuedR every week	10	9.0	9.0	1.3

of participants, respectively reported that spatial, phrase, and numerical cues were *always* or *often* helpful in recognizing keywords. The participants' diverse choices for cues to aid recognition and their high login success rate support our anticipation that letting users choose the appropriate cue(s) to their learning process aids the memorability for system assigned random passwords.

### 3.5.5 User Feedback on Usability and Applicability

We asked the participants to answer two sets of 10-point Likert-scale questions (1: *strong disagreement*, 10: *strong agreement*) at the end of the second session. We reversed some of the questions to avoid bias; the scores marked with (\*) were reversed before calculating the modes, medians, and means. So, a higher score always indicates a more positive result for CuedR. To design the questionnaire, we carefully followed the guidelines provided in the existing password literature [57, 51, 50], including using nearly identical questions to those from other studies.

### 3.5.5.1 Usability

Participants showed a high degree of satisfaction with the usability (e.g., memorability, ease of login, ease of using either weekly or daily) of CuedR. Their feedback was also positive (mode, median, and mean higher than neutral) regarding login time, and they indicated that with practice they could log in quickly using CuedR (see Table 3.1). In our study, we could not test the usability of implicit feedback for CuedR, since most users did not make enough login mistakes to gain experience with it.

### 3.5.5.2 Applicability

At the end of second session, we asked 31 of the participants,<sup>2</sup> “Do you want to use CuedR in real life as a replacement to traditional textual passwords?” 84% responded ‘Yes’, 10% responded ‘Maybe’, and two participants responded ‘No’, where both of them mentioned that they would prefer traditional textual passwords in real life as they did not find any problems with them. User feedback about the applicability of CuedR in different online accounts is illustrated in Table 3.2.

Table 3.2. The applicability of CuedR for different online accounts. Scores are out of 10.

Online accounts	Mode	Median	Mean	SD
Bank	10	8.0	7.4	2.6
E-mail	10	9.0	8.1	2.1
Social Networking	10	9.0	7.7	2.4
University Portal	10	8.0	8.2	1.9
E-commerce	10	9.0	7.8	2.5

---

<sup>2</sup>We failed to ask the first six participants.



### 3.5.6 Pilot study: Memorability for Multiple CuedR Passwords

It is common in password research to report a single-password study in the first article of a new authentication scheme, which helps to establish performance bounds and figure out whether multiple-passwords tests are worthwhile in future research. A recent survey [19] reported that out of 25 graphical password schemes proposed to date, only three have been evaluated through a multiple-password study, and none of these study results was reported in the first article. Since the use of multiple passwords is an important issue for deployment, however, we conducted a pilot study for multiple passwords, in addition to reporting the detailed results of a single-password study.

The study procedure was same as that in our single-password study, except that each participant was assigned three CuedR passwords (18 keywords, in total) instead of one. To administer this experiment, we created three different websites outfitted with CuedR and presented the sites to participants as tabs in an open browser window. Participants were free to select the order of websites at registration and login, but the tabs were arranged the same way every time. For this study, we recruited 11 students (9 men, 2 women) who came from various majors of our university. We believe that 11 represents a suitable sample size for a pilot study [53].

In this study, all of the participants were able to log in successfully within three attempts in both *login 1* (same day of registration) and in *login 2* (one week after registration). In *login 1*, nine participants (82%) succeeded on the first attempt for all three CuedR passwords. One participant (9%) succeeded to log in using two CuedR passwords on the first attempt, where she recognized 17 keywords on the first attempt and corrected the lone mistake on the second attempt. Another participant succeeded on the first attempt for one CuedR password, where she successfully recognized 15 keywords on the first attempt and corrected the mistakes on the second attempt.

In *login 2*, six participants (55%) succeeded on the first attempt to recognize all 18 keywords. Four participants (36%) successfully recognized 17 keywords on the first attempt, i.e., they succeeded to log in using two CuedR passwords on the first attempt. For another CuedR password, two (18%) of these four participants succeeded on the second attempt and other two participants succeeded on the third attempt. One participant (9%) successfully recognized 16 keywords on the first attempt. In particular, she succeeded to log in using one CuedR password on the first attempt and succeeded on the second attempt for other two CuedR passwords.

### 3.6 Discussion

In this section, we discuss three important aspects of CuedR: i) impact, ii) acceptance, and iii) application. Here, the term *study* refers to our single-password study, unless otherwise specified. We conclude with a discussion on the scope for future research on CuedR.

#### 3.6.1 CuedR: The Impact

Existing password systems fail to fully address users' cognitive limitations or leverage humans' cognitive strengths. Thus, despite a large body of research, it still remains a critical challenge to build an authentication scheme that provides both guessing resilience and high memorability. CuedR represents a breakthrough, offering high memorability for system-assigned random passwords, and shows a promising research direction to leverage humans' cognitive abilities for user authentication.

System-assigned passwords provide higher security against guessing attacks than user-chosen passwords, but it is difficult for most people to memorize them [14, 15, 16]. Users have varying cognitive strengths and abilities, and it is hard to know in advance what will help a given user to remember her password. In CuedR, we

present a variety of visual, verbal, and spatial information related to randomly selected keywords in an organized way and then let users choose the appropriate cue(s) to their learning process. Further, the cues can work together. When we asked users to identify the cues they used, 83.8% of users reported of using multiple cues often or always. As one participant commented, “The image, phrase and number naturally correspond in my mind, and make it easy to remember.”

CuedR also shows that the cued-recognition class of password schemes, a new design point in the field, can be effective for user authentication. In particular, CuedR addresses each of the features needed in an effective graphical password scheme, as identified by Biddle et al. [19] from their comprehensive survey on the graphical password literature.

Here we mention a participant’s feedback that particularly drew our attention: *“The multiple cues make it a helpful password scheme for autistic persons, who find it cognitively difficult to create secure passwords.”* We appreciate such a thoughtful opinion and note it to be an important issue to be explored in future work.

### 3.6.2 CuedR: The Acceptance

In traditional user-chosen passwords, users bear the responsibility of ensuring security for their online account through a secure password that should be chosen with creativity and intelligence so that it achieves satisfactory memorability. For many users, this is a lot of work, and thus in many cases they compromise with security and create a weak but memorable password. A recent study [2] reveals that with the advancement of digital technology and widespread use of internet in recent years, users now better realize the importance of strong passwords than anytime before, and many of them intend to create secure passwords but just fail to achieve a good balance between security and memorability. So, rather than blaming users

for predictable passwords, researchers should improve how authentication systems address human cognitive abilities.

Participants in our study seem to be convinced with the security provided by a system-assigned password. In a post-experiment open-ended question where they were asked about their opinion of CuedR, most of them reported high satisfaction with its security features.

Since the participants were convinced with the security of six system-assigned keywords, and could efficiently recognize each keyword in a reasonable time (6.3 seconds, on average) after a week of registration, they found the overall login time acceptable and reported satisfaction with the usability of CuedR (see Table 3.1). 84% of participants preferred to use the scheme in real life as a replacement to traditional textual passwords.

### 3.6.3 CuedR: The Applications

Although most of the participants reported strong agreement about using CuedR for all of the given account types (see Table 3.2), we must be cautious to recommend its application since textual passwords have lower login times than CuedR. Traditional textual passwords are fraught with security problems that make them less than desirable, especially for high-security accounts [3]. Ideally, there should be a clear separation between the passwords used for low-security websites and high-security websites [76]. Thus, CuedR can be used as a standard authentication mechanism for online accounts with high security requirements and where logins occur relatively infrequently, such as financial (e.g., online banking, brokerage services) and e-commerce accounts [71]. A study by Hayashi and Hong [71] finds that users log into financial and e-commerce sites once a week on average, which is in agreement with the interval of one week before *login 2* in our study. By using CuedR for high-security accounts,

it helps to build the mental separation with lower security accounts and avoid attacks based on password reuse and predictable patterns.

As compared to other cognometric graphical password schemes that present users with images only, the deployment of CuedR may require more effort, where separate portfolios of keywords are built accommodating both graphical and verbal cues. We note that each commercial deployment can use a small set of portfolios for all of its users. For example, with 10 portfolios, a phisher could correctly guess the first two portfolios for a user only 1% of the time.

### 3.7 Conclusion

In this chapter, we present a novel authentication scheme, CuedR, which helps us to explore the efficacy of combining graphical, verbal, and spatial cues to improve the memorability of system-assigned random passwords. We also discuss the promise of CuedR in addressing the features of an effective graphical password scheme [19]. Although the login time is relatively high, our primary findings indicate high memorability for CuedR, suggesting that cued-recognition would be an important direction in password research to address the usability-security tension in authentication.

## CHAPTER 4

### The Impact of Cues and User Interaction on the Memorability of System-Assigned Recognition-Based Graphical Passwords

In this chapter, we draw upon several prominent theories of cognitive psychology to enhance the memorability of system-assigned recognition-based graphical passwords. In particular, we examine the impact of using memory cues and the efficacy of requiring user interaction at registration, in which we have users apply their observation and imagination to type a short description about assigned images. Considering both human faces and objects as images, along with cues and interaction, we design seven different study conditions (see Figure 4.1). In §4.1, 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 §4.2 and present the results in §4.3, followed by a discussion on our findings in §4.4.

#### 4.1 System Design

Passfaces [12] provides *PIN-level* security (13 bits of entropy), while an authentication scheme should offer at least 20 bits of entropy to attain *password-level* security [19]. Hlywa et al. [69] provide a guideline to design recognition-based graphical authentication schemes with password-level security, where the user is assigned five images at registration and has to recognize each of the assigned images from a distinct portfolio of 16 images during login. We follow this guideline to design our study conditions, where a successful authentication requires the user to recognize all five images correctly. For an unsuccessful login, the user is shown an error message

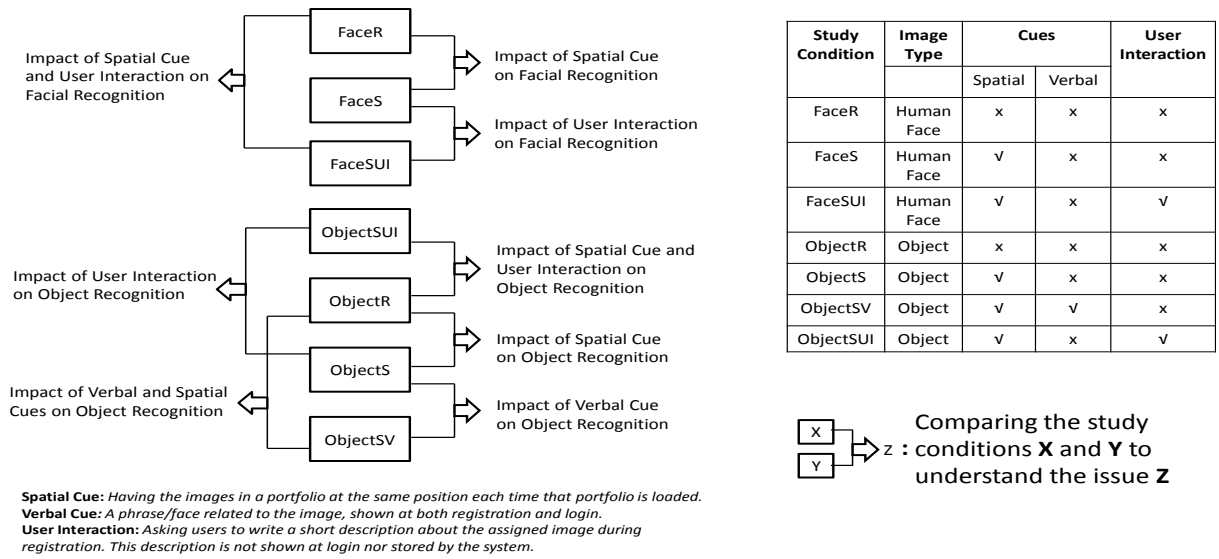


Figure 4.1. The Study Model to Understand the Impact of Cues and User Interaction on Recognition-based Graphical Authentication.

at the end of the login attempt but not informed on which portfolio the mistake was made.

Users are required to perform a recognition task in our study, since the psychology research [7, 8, 9] has found that recognition is easier than recall (see §3.2.2 for details). In this section, we explain from the perspective of cognitive psychology how our design choices are set up to understand the impact of cues and user interaction (at registration) in improving the memorability for system-assigned recognition-based (i.e., cognometric) graphical password schemes. We illustrate our study model in Figure 4.1.

#### 4.1.1 Visual Memory

In our study, we leverage users' visual memory. Psychology research shows that the human brain is better at memorizing graphical information as compared to textual information [10, 11]. 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* [10], 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* [11], which states that images are accompanied by more distinct sensory codes that allow them to be more easily accessed than text.

#### 4.1.2 Face Recognition

In our study, we consider face and object images separately, to understand the impact of cues and user interaction on each image type for recognition-based graphical authentication. Passfaces [12] uses face images, and prior research [77, 78] has shown evidences that there may be regions of the human brain dedicated to processing facial information and recognizing faces. Minnebusch et al. [78] demonstrate that three important regions of human brain, *fusiform face area (FFA)*, *superior temporal sulcus (STS)*, and *occipital face area (OFA)*, are recruited bilaterally (with some right hemisphere bias) while processing facial information. The results of functional MRI show that FFA in the brain gets activated more strongly while viewing faces as compared to other visual objects. STS is sensitive to dynamic aspects of face stimuli, such as gaze or expression. OFA is another important area of human brain, and it deals with the physical features of faces. The findings of Minnebusch et al. [78] are in agreement with the evidences shown by Haxby et al. [77] that suggest that face recognition is functionally different than recognizing other visual objects.

#### 4.1.3 Long-Term Memory

According to the cognitive memory model proposed by Atkinson and Shiffrin [17], an elaborative encoding (for new information) from short-term memory to long-term



memory takes place when the information can be associated with something meaningful, such as cues [17]. This encoding helps people to remember and retrieve the processed information efficiently over an extended period of time. To motivate such encoding, we examine two different approaches in our study:

- *Cued-recognition*: Providing memory cues (e.g., spatial, verbal) with the images, which would be shown both at registration and login.
- *User interaction*: Asking users to write a short description about the assigned images during registration. These descriptions would not be shown at login nor stored by the system.

To explore the impact of cues and user interaction on graphical recognition, we design a control condition for face recognition, in which the images in a portfolio remain the same but randomly positioned each time that portfolio is loaded, as in Passfaces [12]. In this chapter, we term this control condition *FaceR* (**F**ace images with **R**andom positioning). We design a similar control condition for object recognition that we call *ObjectR*.

#### 4.1.3.1 Cued-Recognition

Psychology research [8, 7] 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* [66] postulates that the most effective cues are those that are present at the time of remembering. In this study, we aim to understand the impact of spatial and verbal cues in improving the memorability of cognometric graphical passwords.

**Spatial Cues.** Having a fixed set of objects in a certain place helps to augment *semantic priming*, which refers to recognizing an object through its relationship with

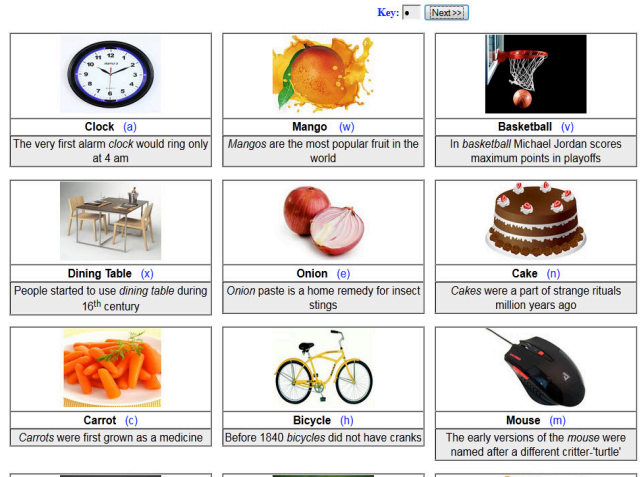


Figure 4.2. A partial screen shot of ObjectSV scheme during login. The facts corresponding to each image appear below that image. Users enter the key, a lowercase letter shown in parentheses, in the password field (on top) to select the corresponding image. The keys are randomly assigned to images each time the portfolio is loaded, where no two images share the same key. During login, users are shown five such portfolios, where each presents a distinct set of 16 images including one of the five assigned images..

other objects around it [12]. Semantic priming thus eases the recognition task [12]. In a graphical password scheme offering spatial cues, the images in a portfolio remain the same and are presented at a fixed position whenever that portfolio is loaded. For example, in Figure 4.2, 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, the schemes (except control conditions: FaceR and ObjectR) in our study offer spatial cues, while we design *FaceS* (**F**ace Images with **S**patial cues) and *ObjectS* (**O**bject Images with **S**patial cues) schemes to understand the precise impact of spatial cues on graphical recognition.

In the FaceS and ObjectS schemes, the images in a portfolio remain at the same position each time that portfolio is loaded. We compare FaceS with FaceR, and

ObjectS with ObjectR to show the impact of spatial cues on face recognition and object recognition, respectively.

**Verbal Cues.** If the system provides verbal cues, i.e., phrases/facts related to the images, then users may focus their attention on associating the images with the corresponding cues, which should help to process and encode the information in memory and store them in the long-term memory. The cues would also assist users to recognize the images in the future and thus enhance their memorability.

In our study, *ObjectSV* (**O**bject images with **S**patial and **V**erbal cues) scheme provides users with verbal cues. For example, the image of a ‘Dining Table’ is provided with the name of this object (‘Dining Table’), and a corresponding phrase/fact (“People started to use dining table during 16th century.”). Yan et. al. [68] examined the influence of phrases in increasing the memorability of passwords, which inspires us to accommodate a common phrase or fact for each image as a verbal cue. See Figure 4.2 showing a partial screen shot of the login screen for ObjectSV scheme.

ObjectSV is similar to CuedR except that it provides 20 bits of entropy, which is sufficient to offer password-level security [19]. In case of an incorrect entry during login with ObjectSV, the error message is shown at end instead of providing an implicit feedback, to gain higher resilience against observation attack in comparison to CuedR. Since ObjectSV does not offer implicit feedback, a portfolio does not require to accommodate the images of the same type, which in turn reduces the implementation challenges for this scheme.

We typically do not provide a physical description of an image (e.g., “A dining table has four legs.”) as a phrase or fact, since it is already visible in the image. Rather, ObjectSV offers an additional fact corresponding to the object (in image) as a verbal cue for helping users to better remember the image through correlating it

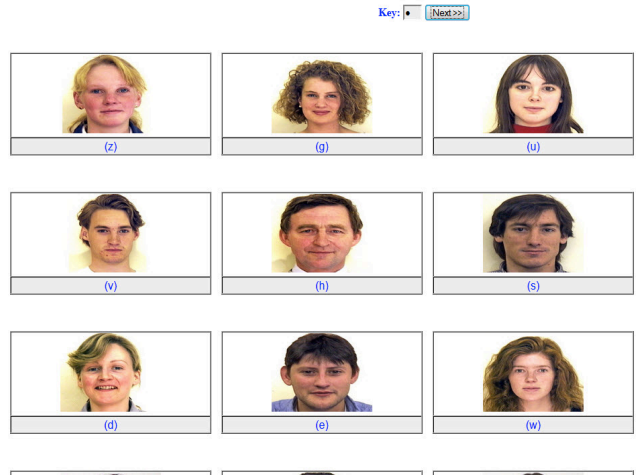


Figure 4.3. A partial screen shot of FaceSUI scheme during login. During login, users are shown five such portfolios, each presents a distinct set of 16 images including one of the five assigned images..

with the given cues. Thus, we did not accommodate verbal cues for face recognition, since it is not possible to provide users with facts about the anonymous face images.

We compare ObjectSV with ObjectR to examine the memorability gain through combining spatial and verbal cues, while the comparison between ObjectSV and ObjectS reveals a more precise impact of verbal cues.

#### 4.1.3.2 User Interaction

In our study, we implement user interaction through the schemes FaceSUI (**Face** images with **Spatial** cues and **User Interaction**) and ObjectSUI (**Object** images with **Spatial** cues and **User Interaction**)), in which the system asks users to describe each assigned image during registration. The user interface includes a text field for users to type a short description about the assigned image. The descriptions, provided by the users during registration, are not stored by the system nor shown in any form at login. The sole purpose of this approach is to make random images more familiar

to the users through motivating their deeper observations. See Figure 4.3 showing a partial screen shot of the login screen for FaceSUI scheme.

Unlike existing graphical password schemes, such as Passfaces [12], where users just use their visual memory to memorize the given images, FaceSUI and ObjectSUI schemes leverage both visual memory and action-event memory [18] for an elaborate encoding of authentication information (e.g., assigned images), which help users to remember and retrieve the processed information efficiently over an extended period of time.

We compare FaceSUI with FaceR and ObjectSUI with ObjectR to examine the memorability gain through combining spatial cues with user interaction, while the comparisons of FaceSUI with FaceS and ObjectSUI with ObjectS reveal the more precise impact of user interaction on face recognition and object recognition, respectively.

#### 4.1.4 Variant Response

In existing cognometric graphical password schemes [12, 69], mouse input is used to select an image, where the images in a portfolio remain the same but are positioned randomly each time that a portfolio is loaded to compensate for shoulder surfing risk during login. Since the existing recognition-based schemes [69] are presented through our control conditions, FaceR and ObjectR schemes also use mouse input to select images.

The shoulder-surfing study of Tari et al. [63] reveals that cognometric schemes with keyboard input provide higher resilience to shoulder surfing than schemes with mouse input, since the keyboard input associated with a particular 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 [19].

For a cognometric scheme providing variant response through varying keyboard inputs, the shoulder surfer needs to learn both the user’s keystrokes and the corresponding images by looking at the keyboard and monitor. Tari et al.’s study [63] shows that observing both the monitor and keyboard at the same time is difficult.<sup>1</sup> Thus, the schemes (except control conditions: FaceR and ObjectR) 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 image on the page, and the user inputs the key letter corresponding to her assigned image into a single-character password field to move on to the next portfolio (see Figure 4.2 and Figure 4.3). The user-entered letter in the password field is shown as an asterisk to reduce the risk of shoulder surfing.

## 4.2 User Study

We now present the design of our user study to explore the impact of cues and user interaction on the memorability of recognition-based graphical authentication. In this study, we used a within-subjects design consisting of seven experimental conditions (see Figure 4.1). 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.

### 4.2.1 Participants, Apparatus and Environment

For this experiment, we recruited 56 students (40 women, 16 men) through our university’s Psychology Research Pool. Participants came from diverse backgrounds, including majors from Nursing, Psychology, Business, Environmental Science, Bio-

---

<sup>1</sup>though we note that videotaping could overcome this.

chemistry, and Spanish Language. The age of the participants varied between 18 to 51 with a mean age of 21. 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 seven realistic and distinct websites, including sites for banking, social networking, email, movie streaming, and shopping. The sites used the images and layouts from familiar commercial sites, and each of them was equipped with one of our seven graphical password schemes.

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

#### 4.2.2 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 graphical passwords. A one-week delay is larger than the maximum average interval for a user between subsequent logins to any of her important accounts [71] and is also a common interval used in authentication studies (e.g., [72, 15, 49, 23, 79]).

*Session 1.* After signing a consent form, the participants were given an overview of our study. Then they performed registration for each of the seven 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 seven 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 seven sites using the assigned graphical 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 the participants had finished, they were compensated and thanked for their time.

#### 4.2.3 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 56 participants. However, lab studies have been preferred to examine brain-powered memorability of passwords [80]. 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 56 provides a suitable sample size for a lab study as compared to the prior studies on password memorability [57, 54, 51, 74, 23, 79].

### 4.3 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



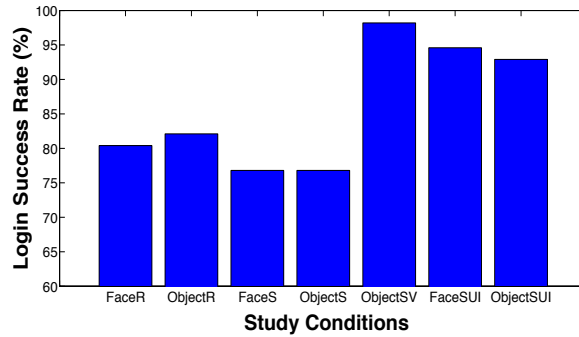


Figure 4.4. Login success rates for the study conditions [Number of participants=56].

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

*H1<sub>a</sub>: The login success rate for FaceS would be significantly higher than that for FaceR.*

*H1<sub>b</sub>: The login success rate for ObjectS would be significantly higher than that for ObjectR.*

In a graphical password scheme offering spatial cues, the images in a portfolio remain same and presented at a fixed position whenever that portfolio is loaded, which establishes a relationship between them and reinforces semantic priming (see §4.1 for details). Thus, we hypothesized that FaceS and ObjectS, offering spatial cues, would have significantly higher login success rates than FaceR and ObjectR, respectively, in which the position of images in a portfolio are randomly changed each time that portfolio is loaded.

Our results show that out of 56 participants in our study, 45 participants (80%) succeeded to log in using FaceR, while 43 participants (77%) logged in successfully with FaceS. For ObjectR and ObjectS schemes, 46 participants (82%) and 43 participants (77%) succeeded to log in, respectively (see Figure 4.4). Thus,  $H1_a$  and  $H1_b$  are not supported by these results.

Whether or not a participant successfully authenticated is a binary measure, so we compare login success rates between conditions using McNemar’s test. We did not find a significant difference in login success rate between FaceS and FaceR,  $\chi^2(1, N = 56) = 0.08, p = 0.77$ , nor between ObjectS and ObjectR,  $\chi^2(1, N = 56) = 0.36, p = 0.55$ .

## **Hypothesis 2**

*H2<sub>a</sub>: The login success rate for ObjectSV would be significantly higher than that for ObjectS.*

*H2<sub>b</sub>: The login success rate for ObjectSV would be significantly higher than that for ObjectR.*

The ObjectSV scheme offers spatial and verbal cues (i.e., phrase or facts related to the images), where cues are shown both at registration and login. So, the users could memorize their graphical passwords through associating them with the corresponding cues, which should help to process and encode the information to store them in long-term memory (see §4.1 for detailed discussion). Moreover, the cues would assist users to recognize the images in the future, which should enhance their memorability. Thus, we hypothesized that ObjectSV scheme would have significantly higher login success rate than ObjectS and ObjectR schemes.

We observed a 98% login success rate for ObjectSV scheme, while 55 out of 56 participants could log in successfully one week after registration. As we compare the

login success rate for ObjectSV scheme with that for ObjectS (77%) and ObjectR (82%), the results for McNemar’s test show that ObjectSV had a significantly higher login success rate than ObjectS,  $\chi^2(1, N = 56) = 10.08, p < 0.05$  and ObjectR,  $\chi^2(1, N = 56) = 7.11, p < 0.05$ . Hence,  $H2_a$  and  $H2_b$  are supported by these results.

### **Hypothesis 3**

*H3<sub>a</sub>: The login success rate for FaceSUI would be significantly higher than that for FaceS.*

*H3<sub>b</sub>: The login success rate for FaceSUI would be significantly higher than that for FaceR.*

*H3<sub>c</sub>: The login success rate for ObjectSUI would be significantly higher than that for ObjectS.*

*H3<sub>d</sub>: The login success rate for ObjectSUI would be significantly higher than that for ObjectR.*

FaceR and ObjectR schemes represent existing graphical password schemes that use just the visual memory of users [12, 69]. FaceSUI and ObjectSUI schemes leverage both visual memory and action-event memory [18], which contributes to an elaborate encoding of the assigned images and thus assists users with memorizing the processed information. In addition, FaceSUI and ObjectSUI schemes offer spatial cues. So, we hypothesized that the login success rate for FaceSUI would be significantly higher than that for FaceS and FaceR schemes, while ObjectSUI would have a significantly higher login success rate than ObjectS and ObjectR schemes.

Our results show that 53 participants (95%) in FaceSUI and 52 participants (93%) in ObjectSUI scheme logged in successfully one week after registration. As we compare the login success rate for FaceSUI with that for FaceS (77%) and FaceR

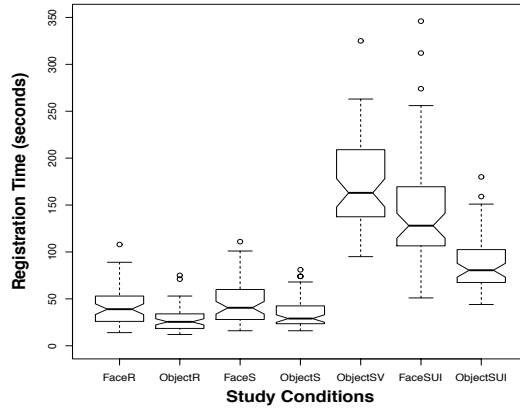


Figure 4.5. Registration time for the study conditions.

(80%), the results for McNemar’s tests show that FaceSUI had a significantly higher login success rate than FaceS,  $\chi^2(1, N = 56) = 8.1, p < 0.05$  and FaceR,  $\chi^2(1, N = 56) = 4.9, p < 0.05$ .

We also found that the login success rate for ObjectSUI was significantly higher than that for ObjectS,  $\chi^2(1, N = 56) = 7.11, p < 0.05$  and ObjectR,  $\chi^2(1, N = 56) = 4.17, p < 0.05$ . Thus,  $H3_a, H3_b, H3_c,$  and  $H3_d$  are supported by these results.

#### 4.3.1 Registration Time

We illustrate the results for registration time in Figure 4.5. We found that the median registration times for FaceR and FaceS were 39 seconds and 41 seconds, respectively, while FaceSUI scheme had a median registration time of 128 seconds. 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 FaceR ( $V = 1596, p < 0.05$ ) and FaceS ( $V = 1596, p < 0.05$ ) were significantly less than that for FaceSUI scheme. We did not find a significant difference in registration time between FaceR and FaceS ( $V = 789.5, p = 0.69$ ).

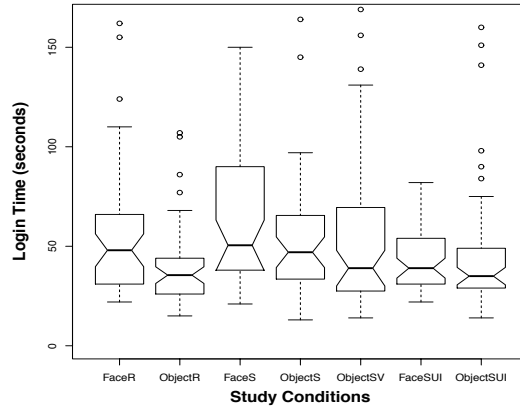


Figure 4.6. Login time for the study conditions.

Our results show that the median registration time for ObjectSUI scheme was 81 seconds, while ObjectR and ObjectS schemes had median registration time of 26 seconds and 29 seconds, respectively. The results for Wilcoxon signed-rank tests show that the registration time for ObjectR ( $V = 1595$ ,  $p < 0.05$ ) and ObjectS ( $V = 1582.5$ ,  $p < 0.05$ ) were significantly less than that for ObjectSUI. We also found that the registration time for ObjectR was significantly less than that for ObjectS ( $V = 1074.5$ ,  $p < 0.05$ ). In our study, ObjectSV scheme had a median registration time of 163 seconds, while the registration time for ObjectR ( $V = 1596$ ,  $p < 0.05$ ), ObjectS ( $V = 1596$ ,  $p < 0.05$ ), and ObjectSUI ( $V = 1566.5$ ,  $p < 0.05$ ) were significantly less than that for ObjectSV scheme.

We intend to see if the image type had any impact on the required time for learning system-assigned images (i.e., registration time). Our results show that the registration time for ObjectR was significantly less than that for FaceR ( $V = 147.5$ ,  $p < 0.05$ ). We also found that the registration time for ObjectS was significantly less than that for FaceS scheme ( $V = 249$ ,  $p < 0.05$ ), while the registration time for ObjectSUI was significantly less than that for FaceSUI ( $V = 39$ ,  $p < 0.05$ ).

### 4.3.2 Login Time and Number of Attempts

In this chapter, *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 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.

#### 4.3.2.1 Login Time

We illustrate our results for login time in Figure 4.6. We found that the median login time for FaceR and FaceS were 48 and 51 seconds, respectively, while FaceSUI had a median login time of 39 seconds. The results for Wilcoxon-Mann-Whitney tests show that the login time for FaceSUI scheme was significantly less than that for FaceS ( $W = 746.5, p < 0.05$ ). We did not find a significant difference in login time between FaceSUI and FaceR ( $W = 1017, p = 0.21$ ), nor between FaceR and FaceS ( $W = 1096, p = 0.20$ ).

Our results show that the median login time for ObjectSUI scheme was 35 seconds, while ObjectR and ObjectS schemes had median login times of 36 and 47 seconds, respectively. The results for Wilcoxon-Mann-Whitney tests show that the login time for ObjectSUI ( $W = 814, p < 0.05$ ) and ObjectR ( $W = 1310.5, p < 0.05$ ) were significantly less than that for ObjectS. We did not find a significant difference in login time between ObjectSUI and ObjectR ( $W = 1285, p = 0.53$ ).

We found a median login time of 39 seconds for ObjectSV. No significant difference was found in terms of login times when we compared ObjectSV with ObjectR

Table 4.1. Number of Attempts for Successful Logins [SD: Standard Deviation]

Study Conditions	Mean	Median	SD
FaceR	1.3	1	0.7
ObjectR	1.2	1	0.5
FaceS	1.3	1	0.6
ObjectS	1.3	1	0.7
ObjectSV	1.4	1	0.9
FaceSUI	1.1	1	0.3
ObjectSUI	1.1	1	0.4

( $W = 1489$ ,  $p = 0.13$ ), ObjectS ( $W = 1020.5$ ,  $p = 0.25$ ), and ObjectSUI ( $W = 1560.5$ ,  $p = 0.42$ ).

We compared ObjectR and FaceR to see if image type had any impact on login time given random image positioning. The results for Wilcoxon-Mann-Whitney tests show that the login time for ObjectR was significantly less than that for FaceR ( $W = 1312.5$ ,  $p < 0.05$ ). However, we did not find a significant difference in login time between ObjectS and FaceS ( $W = 1033$ ,  $p = 0.26$ ), nor between ObjectSUI and FaceSUI ( $W = 1498.5$ ,  $p = 0.44$ ).

#### 4.3.2.2 Number of Attempts

The mean number of attempts for a successful login was less than two for each of the seven schemes, while the median was one in each case (see Table 4.1). 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.

## 4.4 Discussion

Cognometric graphical passwords (e.g., Passfaces [12]) are now commercially available and deployed by a number of large websites, in which the images are assigned by the system to provide reasonable security guarantees. 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 new direction to improve memorability for these passwords by leveraging humans' cognitive abilities through cues and interaction.

### 4.4.1 Cued-Recognition

We accommodate the scientific understanding of long-term memory to improve the memorability of system-assigned cognometric passwords. As noted by Atkinson and Shiffrin [17], 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 spatial and verbal cues for an elaborate encoding of authentication information to ease recognition during login.

Our study on CuedR show the potential of combining multiple cues to aid recognition, where the participants were asked to rate the efficacy of each cue. The participants rated spatial cue to be more effective than verbal cues to aid recognition (see §3.5.4). We then made a deeper investigation of this issue in this study through a direct comparison between schemes offering different combinations of cues, and we found that spatial cues did not significantly contribute to enhance memorability, while verbal cues made a significant contribution in this regard. Thus, the findings from our study make an important contribution to understand the effectiveness of memory cues (e.g., spatial and verbal cues), and indicate that relying solely on user feedback



might not be a reliable approach to understand the impact of cues on password memorability.

#### 4.4.1.1 Spatial Cues

To understand the efficacy of spatial cues, we compared FaceS and ObjectS schemes with fixed positions of all images in a portfolio with FaceR and ObjectR schemes with random repositioning of the images. Our results show that spatial cues did not contribute to improve the login success rate for either facial recognition or object recognition.

In theory, spatial cues reinforce semantic priming and thus ease the recognition task [12]. It is possible that spatial cues are less effective when remembering multiple images, in which case it might create confusion when a user attempts to recognize the images using spatial cues. In our future work, we would perform a field study to explore if a higher login frequency could lead to training effects that could help users to benefit from spatial cues.

#### 4.4.1.2 Verbal Cues

We compared ObjectSV, which has spatial and verbal cues, with both the object-based control condition (ObjectR) and ObjectS, which has spatial cues but not verbal ones. We found a 98% login success rate for ObjectSV, which was significantly higher than that for ObjectS and ObjectR. Given that we also found no benefit in spatial cues alone, we conclude that providing verbal cues with the images played a significant role in improving memorability.

During registration with ObjectSV, the participants may have learned the assigned images 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 ObjectR and ObjectS. No significant difference was found in terms of login time or number of attempts for successful logins between ObjectSV and either ObjectR or ObjectS.

We observed an interesting anecdotal case from one participant. In the first session, he told us that he used to struggle in memorizing new information because of a severe injury on his head. At this point, we were interested to see his login performances in the second session, and we found that he could log in successfully only with the schemes offering verbal cues (e.g., ObjectSV) and leveraging user interaction (e.g., FaceSUI). At the end of second session, he said, “It is much too difficult to manage with just images. The fact [verbal cue] attached with an image help to encode the images.” Indeed, the benefit of verbal cues may not be important to all users, but may instead help users who struggle with graphical information alone, like the approximately 20% of participants who failed to login with ObjectS and ObjectR. If so, it may be good to individually tailor a scheme with different types of information for different users.

#### 4.4.2 User Interaction

We tested user interaction with the FaceSUI and ObjectSUI schemes, in which we have users describe their assigned images at registration so as to deepen users’ processing of authentication information. For clarity, we again note that the descriptions would be immediately destroyed and need not even be transferred to the server.

We compare FaceSUI with FaceR and ObjectSUI with ObjectR to examine the memorability gain through combining spatial cues with user interaction, while comparisons of FaceSUI with FaceS and ObjectSUI with ObjectS reveal the more precise impact of user interaction on face recognition and object recognition, respectively. Our results show that the login success rate for FaceSUI (95%) was significantly higher

than that for FaceS and FaceR and the login success rate for ObjectSUI (93%) was significantly higher than for ObjectS and ObjectR. It appears that user interaction played the major role to improve the success rates, since spatial cues by themselves did not help.

During registration with user interaction based schemes (e.g., FaceSUI and ObjectSUI), the participants wrote descriptions about the assigned images, which required significantly higher registration time for FaceSUI (in comparison to FaceS and FaceR) and ObjectSUI (in comparison to ObjectS and ObjectR).

#### 4.4.3 Cued-Recognition vs. User Interaction

In our study, we found a significant improvement in login success rate through cued-recognition (ObjectSV) and user interaction (FaceSUI, ObjectSUI). The login success rate in ObjectSV was found higher than ObjectSUI and FaceSUI schemes, but no significant difference was found in this regard. We did not find a significant difference in login time or number of attempts for successful logins, when comparing ObjectSV with ObjectSUI and FaceSUI schemes. However, the registration time for ObjectSUI and FaceSUI were significantly less than that for ObjectSV scheme, indicating that the participant required less time to learn the assigned images through writing a description as compared to memorizing images through correlating them with the given verbal cues.

The deployment of ObjectSV scheme may require more effort as compared to other cognometric graphical password schemes that present users with images only, since ObjectSV requires writing verbal cues in addition to the images.

The success of the user-interaction-based schemes depends on the involvement of users in describing the assigned images. Our observations during the study reveal that all of the participants in ObjectSUI and FaceSUI schemes put in effort to describe

the assigned images. Users in a lab study, however, are naturally open to take on requested tasks, while users in real life may get lazy. We plan to explore this issue deeper through a field study in a real-life setting and identify more ways for actively compelling users to engage with the interaction activity.

#### 4.4.4 Face recognition vs. Object Recognition

Hlywa et al. [69] conducted a study to examine the effect of image type on the usability of recognition-based graphical passwords, in which they focused on exploring that impact for randomly-positioned images (similar to FaceR and ObjectR in our study).<sup>2</sup> In this chapter, we provide a deeper understanding on this issue, while our investigation about the efficacy of cues and user interaction for face and object images lets us compare the face and object recognition in terms of registration time, login success rate, and login time for three different conditions: i) The images in a portfolio are randomly positioned each time that portfolio is loaded (FaceR, ObjectR), ii) The images in a portfolio remain at the same position each time that portfolio is loaded (FaceS, ObjectS), iii) The images in a portfolio are placed at the same position each time that portfolio is loaded, and users learn the graphical passwords through interaction, e.g., writing a description about the assigned images at registration (FaceSUI, ObjectSUI).

The registration time for ObjectSUI scheme was found to be significantly less than that for FaceSUI scheme, which indicates that it was less time consuming for the participants to describe object images in comparison to face images. We also found that the registration time for ObjectR and ObjectS were significantly less than that for

---

<sup>2</sup>For randomly-positioned images, our findings about the impact of image type in terms of login time and login success rate are similar to those of Hlywa et al. [69]. Their study did not evaluate the effect of image type on registration time.

FaceR and FaceS schemes, respectively. Thus, users seem to need less time for object images. We speculate that since object images can be selected to be rather distinct from each other within a portfolio both visually and semantically (see Fig. 3.1), it takes less time to memorize a particular assigned object than for faces, which have distinct details but a basic similarity.

In login performance, we found no significant difference in login success rates as we compared ObjectR with FaceR, ObjectS with FaceS, and ObjectSUI with FaceSUI. ObjectR had a significantly lower login time than FaceR, but no significant difference was found in login time as we compared ObjectS with FaceS, and ObjectSUI with FaceSUI. Random positioning may make visually distinctiveness more important to quick login times.

#### 4.5 Conclusion

In our study, we aimed to better understand the impact of cues and user interaction on system-assigned recognition-based graphical passwords, and designed seven different study conditions to achieve this goal. In a study with 56 participants, we had a 98% login success rate for a scheme offering spatial and verbal cues (ObjectSV), while a scheme based on user interaction had a 95% login success rate for face images (FaceSUI) and a 93% login success rate for object images (ObjectSUI). Our analysis show that verbal cues and user interaction made an important contribution to gain significantly higher login success rate as compared to the control conditions representing existing graphical password schemes. Contrary to the suggestions of user feedback from a prior study [79], we found that spatial cues were not effective. These findings shed light on a promising research direction to leverage humans' cognitive ability through cues and interaction in gaining high memorability for system-assigned random passwords.

## CHAPTER 5

### The Impact of Cues on Textual Recognition

In Chapter 4, we examined the impact of cues on graphical passwords. Our results show that verbal cues played an important role to significantly improve the memorability for recognition-based graphical authentication scheme. In this section, we examine the impact of verbal cues in enhancing the usability for textual recognition. In addition, we analyze whether adding images related to the keywords contributes to higher memorability than when users are provided with just verbal cues.

We organize the rest of the chapter as follows: In §5.1, we explain how the design choices for our study conditions are set up. We then describe our study procedure in §5.2 and present the results in §5.3. In §5.4, we discuss the findings from our study, followed by a conclusion in §5.5.

#### 5.1 System Design

Hlywa et al. [69] 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.

The study of Wright et al. [15] found insufficient memorability for textual recognition, where the keywords in a portfolio remained same but were shown at different

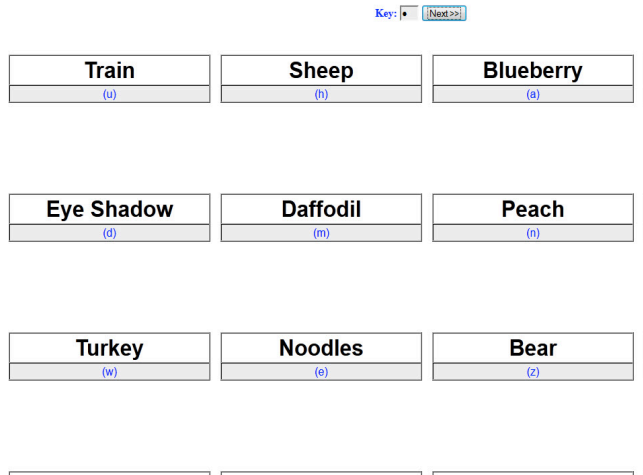


Figure 5.1. A partial screen shot of the *Control* condition during login. Users enter the key, a lowercase letter shown in parentheses, in the password field (on top) to select the corresponding keyword. The keys are randomly assigned to keyword each time the portfolio is loaded, where no two keywords 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..

positions each time that portfolio was loaded. The authors anticipated that showing the keywords in the same position each time would improve the memorability for recognition-based schemes and suggested the approach to be examined in future work. We adopt suggestion of Wright et al. [15] to design our study conditions by showing the keywords in a portfolio in the same position each time that a portfolio is loaded. We also accommodate the *variant response* feature in our schemes to gain resilience against observation attacks like shoulder surfing (see §4.1.4 for details).

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 5.1). 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 5.2). We compare TextV

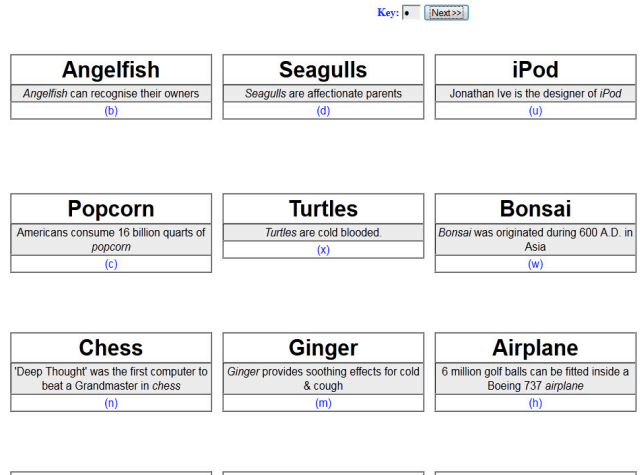


Figure 5.2. A partial screen shot of *TextV* scheme during login. The facts corresponding to each keyword appear below that keyword..

with the Control condition to examine the impact of verbal cues in improving the memorability for textual recognition.

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*: **Graphical** Recognition with **Verbal** cues, and compare it with the TextV scheme. In GraphicV scheme, the system provides users with images corresponding to the keywords along with the verbal cues (see Figure 5.3) <sup>1</sup>. To the best of our knowledge, no study yet has compared textual and graphical recognition-based schemes in terms of usability.

## 5.2 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 hy-

<sup>1</sup>GraphicV scheme is same as the ObjectSV scheme as noted in Chapter 4





Figure 5.3. A partial screen shot of GraphicV scheme during login. Each keyword is accommodated with the corresponding image..

pothesis tests. The study procedures were approved by our university's Institutional Review Board (IRB) for human subjects research.

### 5.2.1 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.

### 5.2.2 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 [71] and is also a common interval used in authentication studies (e.g., [72, 15, 49, 23, 79, 33]).

#### 5.2.2.1 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 to not record (e.g., write down or take a picture) their authentication secrets.

#### 5.2.2.2 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.

#### 5.2.3 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 [80]. 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 [57, 54, 51, 74, 23, 79, 33].

### 5.3 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 the compared samples,

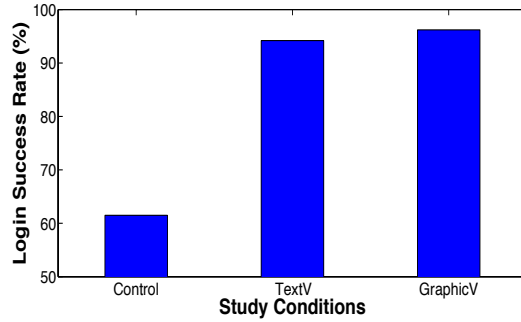


Figure 5.4. Login success rates for the study conditions [Number of participants=52].

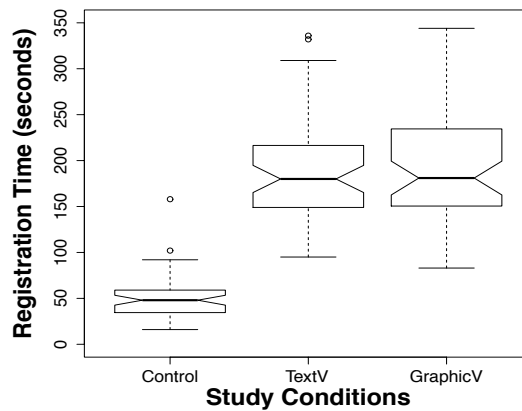


Figure 5.5. Registration time for the study conditions.

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 §5.1 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 5.4). 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,  $\chi^2(1, N = 52) = 12.2, p < 0.01$ . 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 superiority effect* showing that the human brain is better at memorizing graphical information as compared to textual information [10, 11]. 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

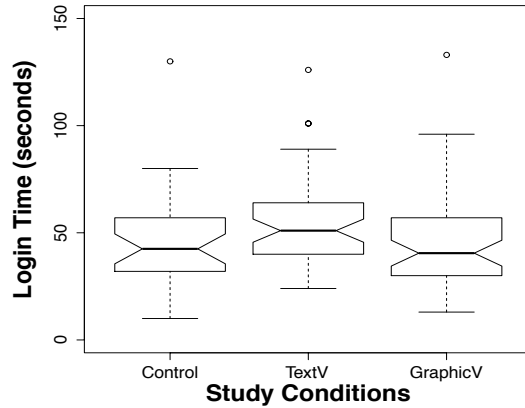


Figure 5.6. Login time for the study conditions.

was no significant difference between TextV and GraphicV schemes in terms of login success rate,  $\chi^2(1, N = 52) = 0, p = 1$ . Hence,  $H_2$  is not supported by these results.

### 5.3.1 Registration Time

We illustrate the results for registration time in Figure 5.5. 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 ( $V = 0, p < 0.01$ ) and GraphicV ( $V = 1, p < 0.01$ ) were significantly less than that for

Table 5.1. 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

the Control condition. We did not find a significant difference in registration time between TextV and GraphicV schemes ( $V = 633.5$ ,  $p = 0.62$ ).

### 5.3.2 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 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.

#### 5.3.2.1 Login Time.

We illustrate our results for login time in Figure 5.6. 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 ( $W = 790$ ,  $p = 0.93$ ).

#### 5.3.2.2 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 5.1). 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 5.2. 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	
	<i>Med</i>	<i>Mo</i>	<i>Med</i>	<i>Mo</i>	<i>Med</i>	<i>Mo</i>
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

### 5.3.3 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 5.2. Since Likert scale data are ordinal, it is most appropriate to calculate mode and median for Likert-scale responses [81].

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 ( $V = 27$ ,  $p < 0.05$ ).



## 5.4 Discussion

In this study, we explore a promising direction to improve the memorability for system-assigned random 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 [17], 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 condi-

tion. No significant difference was found between TextV and GraphicV schemes in terms of registration time.

## 5.5 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.

## CHAPTER 6

### Field Study: Exploring The Usability of GraphicV Scheme in a Real-life Scenario

The lab-study results from Chapter 4 and Chapter 5 have shown that a system-assigned authentication scheme offering users with verbal and spatial cues for recognizing object images performed best in terms of memorability. This scheme is termed *ObjectSV* in Chapter 4 and *GraphicV* in Chapter 5 to keep consistency with other study conditions in the corresponding user study. In this chapter, we use GraphicV to refer to this scheme.

A field study offers strong ecological validity and the best measure of login performance in a realistic setting [19]. We conducted a field study for a detailed understanding on the usability of GraphicV scheme in a real-life scenario. In this field study, we have made some changes in the system design of GraphicV scheme that we explain in §6.1. We describe the procedure of this field study in §6.2 and the results are presented in §6.3. We discuss the findings from this field study in §6.4, followed by a conclusion in §6.5.

#### 6.1 System Design

The system design of GraphicV scheme remains same as explained in the previous chapter, except the following changes that are made to improve the usability and deployability of this scheme in a real-life scenario.

In the modified version of GraphicV scheme, the next portfolio automatically shows when a letter is entered to select a keyword in the current portfolio at login. In the earlier version of this scheme as described in the previous chapter, users were

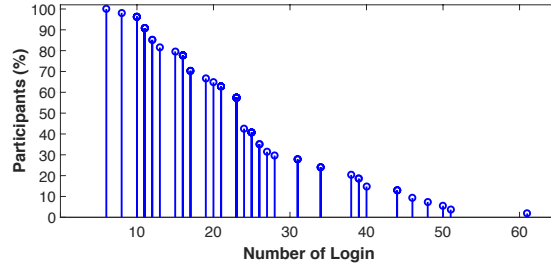


Figure 6.1. *Field Study*: Number of logins by the participants, where a  $(x, y)$  point represents the percentage of participants ( $y\%$ ) who conducted at least  $x$  logins, either successful or unsuccessful..

required to press a “Next” button each time to move to the next portfolio during login. This change in user interface design should contribute to reduce the login time and so on to improve the usability of GraphicV scheme.

To simplify the deployment of GraphicV scheme in a real-life scenario without making any change in the current authentication server of the traditional textual password, we keep the small-case letter for selecting a keyword unchanged across the login sessions. In this case, a textual password comprising of five small-case letters would be stored at the server-end for each user. At the client-end, users do not require to memorize the characters used to select the keywords, rather they could remember the system-assigned keywords representing the name of objects with the help of graphical, verbal, and spatial cues. At login, users recognize the keywords and select them by the entering the corresponding small-case letters that remain fixed across the login sessions. To note, this updated version of GraphicV scheme does not offer variant response feature, so the resilience to shoulder-surfing attack remains same as the traditional textual password.

## 6.2 Study Design

We conducted the field study on a class with both undergraduate and graduate students. At the beginning of study, the students were informed that we developed

a website to let them access course study materials and their grades on exams and assignments.<sup>1</sup> With a projector, the experimenter showed the students how GraphicV scheme works. Then the students completed their registration with the GraphicV scheme in a lab-setting at the presence of the experimenter, so that they could receive help in case of any technical difficulty with using the system. However, we found that registration process with our scheme was well-understood by the students and they did not face any technical difficulty during registration.

Each student in the class was given a username. To protect against unauthorized access, the usernames of the students were pre-stored in the system so that only students in this class could create accounts, one per username. Table 6.1 shows the results for registration.

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. Except one participant, no other participant had taken any courses on usable security or human-computer interaction, nor had they participated in a password-related user study.

In this field study, the users could log in at anytime from anywhere using their desktop or laptop computers. During authentication, we started counting login time after entering the username. A successful attempt required the user to correctly enter both her username and GraphicV password. An unsuccessful attempt refers only to

---

<sup>1</sup>Grades were posted in a file containing all students' grades and anonymized by replacing names with a code given to each student.

sessions where the username was correct but the GraphicV password was incorrect. In this case, we found that the students never failed to correctly enter their username.

To reset a password, the participant had to send an email to the experimenter from her *.edu* email account, and in response she would receive a link through email that would lead her through the registration process to assign a new GraphicV password. Two participants reset their password within the first few days of the study. Thereafter, no participant reset her GraphicV password during the study.

**Ecological validity.** In our field study, the participants were generally young and university educated. So, our findings may not generalize to the entire population of Web users, however, given that the participants were performing a real life task that mattered in their specific situation, the task had reasonable ecological validity.

It is a common approach in password studies to examine the usability of a scheme through analyzing the login performance of university students [82, 15, 22, 69, 16, 50, 21, 51, 83, 57, 32], however, we note that it is important to do further studies for people from different age groups and backgrounds. Now that our results for field study show promise, we would examine the usability of GraphicV for senior users in future work. We also plan to evaluate this scheme for people with cognitive limitations to have a greater understanding on the usability.

### 6.3 Results and Analysis

In this section, we present our results and analyze the login performance of users while using GraphicV scheme in a real-life setting.

Table 6.1. *Field Study*: Registration time (seconds). SD: Standard Deviation

Mean	Median	SD
265	241	110

81

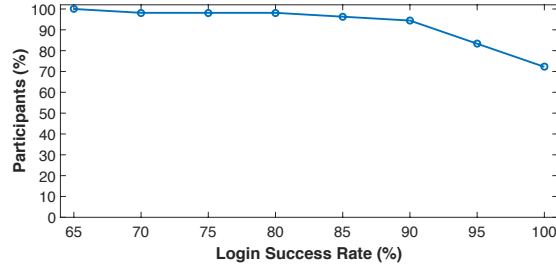


Figure 6.2. *Field Study*: Login success rate (54 participants)..

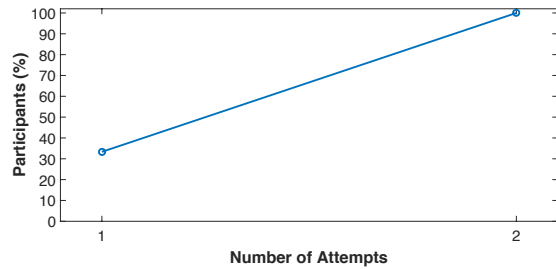


Figure 6.3. *Field Study*: Mean number of attempts..

### 6.3.1 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 (minimum: 6, maximum: 61). Figure 6.1 shows the number of logins by the participants, where a  $(x, y)$  point represents the percentage of participants ( $y\%$ ) who conducted at least  $x$  logins, either successful or unsuccessful.

We measured the average login performance of each participant in her login sessions (see Figure 6.2, 6.3, and 6.4) and calculated the overall login performances for all of the 54 participants over 1349 login sessions (see Table 6.2). The overall login

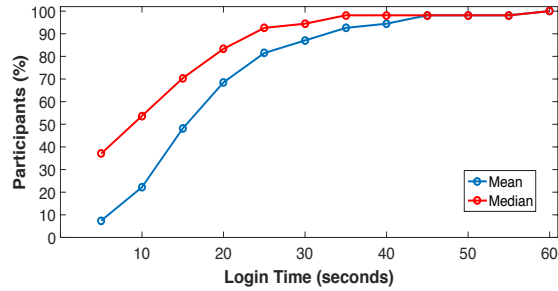


Figure 6.4. *Field Study*: Login time..

success rate was 98%. Users required 1.1 attempts (on average) per successful login, while the average login time was 15 seconds, with a median of 9 seconds.

To illustrate the login performance in more detail, Figures 6.2, 6.3, and 6.4 show empirical cumulative distribution functions (ECDFs) of login performance statistics taken over the users in our study (in Figure 6.2, the x-axis is shown with increasing success rates and thus appears reversed).

Figure 6.2 shows login success rates among participants. GraphicV proved sufficiently memorable for nearly all our participants. 72% of participants had a 100% login success rate, 83% had at least a 95% success rate, and 94% had at least an 90% success rate. The minimum success rate for any participant was 65%. Figure 6.3 shows the average number of attempts per successful login among the participants, where 100% of participants logged in successfully within two attempts on average.

Table 6.2. *Field Study*: Overall login performance [Login Sessions: 1349, Login Success Rate: 98%]. SD: Standard Deviation.

	Mean	Median	SD
No. of Attempts	1.1	1	0.4
Login Time (sec.)	15	9	24



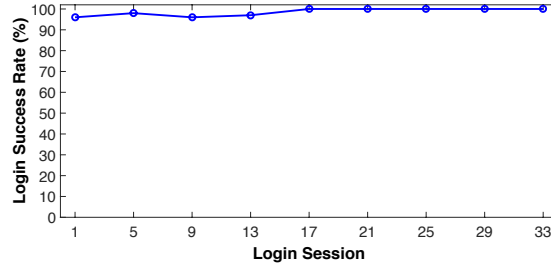


Figure 6.5. *Field Study*: The change in login success rate over the sessions..

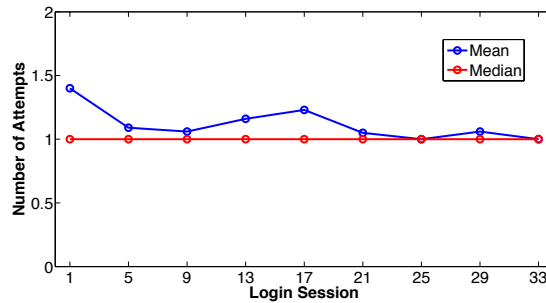


Figure 6.6. *Field Study*: The change in mean number of attempts over the sessions..

The performance for login time was more mixed. Figure 6.4 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.

### 6.3.2 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<sup>rd</sup> login session, since using the higher values of  $x$  would make for a rather small sample size (e.g., 7 users for  $x = 41$ ). The results are shown in Figure 6.5, 6.6, and 6.7.

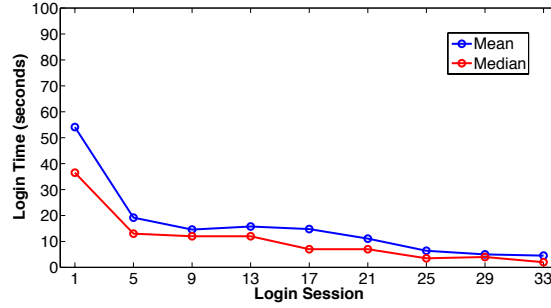


Figure 6.7. *Field Study*: The change in login time over the sessions..

Table 6.3. *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

We note that the sample size changes (shrinks) for each successive value of  $x > 5$  (see Table 6.3). 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 her login success rate ( $r = 0.37$ ) or number of attempts for successful login ( $r = -0.22$ ).

A given  $(x, y)$  point in Figure 6.5, 6.6, and 6.7 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 (see Figure 6.1). Table 6.3 represents the number of participants in each of  $x^{th}$  login sessions.

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 a little impact of training effect on these metrics of login performance. Our results show that

Table 6.4. *Field Study*: Significance tests (Wilcoxon-Mann-Whitney tests) for login time between pairs of login sessions. We consider results comparing two conditions to be significantly different when we find  $p < 0.05$

Session	5	9	13	17	21	25	29	33
1	W=2313, <b>p&lt;0.01</b>	W=2328, <b>p&lt;0.01</b>	W=1930, <b>p&lt;0.01</b>	W=1754, <b>p&lt;0.01</b>	W=1651, <b>p&lt;0.01</b>	W=1119, <b>p&lt;0.01</b>	W=774, <b>p&lt;0.01</b>	W=667, <b>p&lt;0.01</b>
5	-	W=1442, $p=0.44$	W=1226, $p=0.52$	W=1265, <b>p&lt;0.05</b>	W=1187, <b>p&lt;0.05</b>	W=847, <b>p&lt;0.05</b>	W=612, <b>p&lt;0.05</b>	W=556, <b>p&lt;0.05</b>
9	-	-	W=1089, $p=0.91$	W=1156, $p=0.08$	W=1095, <b>p&lt;0.05</b>	W=768, <b>p&lt;0.05</b>	W=565, <b>p&lt;0.05</b>	W=518, <b>p&lt;0.05</b>
13	-	-	-	W=983, $p=0.11$	W=908, $p=0.06$	W=677, <b>p&lt;0.05</b>	W=492, <b>p&lt;0.05</b>	W=448, <b>p&lt;0.05</b>
17	-	-	-	-	W=648, $p=0.98$	W=497, $p=0.22$	W=372, $p=0.08$	W=347, <b>p&lt;0.05</b>
21	-	-	-	-	-	W=447, $p=0.21$	W=335, $p=0.08$	W=315, <b>p&lt;0.05</b>
25	-	-	-	-	-	-	W=182, $p=0.61$	W=184, $p=0.15$
29	-	-	-	-	-	-	-	W=108, $p=0.63$

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<sup>th</sup> session (see Figure 6.5). The mean number of attempts for successful login was less than 1.4 in any login session while the median number of attempts for successful login was 1 (see 6.6).

Training effect was most prominent over login time as shown in Figure 6.7. The median login time was 37 seconds in the first login session, which decreased to 13

Table 6.5. Questionnaire responses for the usability of GraphicV scheme. Scores are out of 10. \* indicates that scale was reversed. *Med*: Median, *Mo*: Mode

Questions	<i>Med</i>	<i>Mo</i>
I could easily sign up with this scheme	8	10
Logging in using this scheme was easy	8	10
Passwords in this scheme are easy to remember	8	10
Passwords in GraphicV scheme is more secure than traditional textual password	8	8

seconds in the 5<sup>th</sup> login session and reduced to 7 seconds in the 17<sup>th</sup> login session, a 81% reduction.

As shown in Figure 6.7, mean and median login times decreased over the login sessions, where we find an exception for mean login time at the 13<sup>th</sup> login session. The results of significance tests for login time between pairs of login sessions are shown in Table 6.4, which suggest 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.

### 6.3.3 User feedback

In the post-experiment survey, participants answered a 10-point likert scale questionnaire, where a higher score indicates a more positive result for GraphicV scheme (see Table 6.5). The majority of participants in field study were satisfied with the GraphicV scheme in terms of ease of registration (Mode: 10, Median: 8), ease of login (Mode: 10, Median: 8), and memorability (Mode: 10, Median: 8). We note that user feedback should be taken with caution because of possible social acceptability bias, although participants were notified that their response would remain anonymous and would not affect their compensation.

## 6.4 Discussion

As noted by Biddle et al. [19], 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 scheme with an overall login success rate of 98%.

### 6.4.0.1 Training effect

Since the prior field studies on system-assigned graphical 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 effect played an important role in the improvement of login time with more login sessions.

As reported in Chapter 5, the login success rate for GraphicV scheme was 96.2% in the short-term lab study while the median login time was 41 seconds. In our field study, the average login success rate for GraphicV was found 96% in the first login session, which is in line with that in the lab study reported in Chapter 5<sup>2</sup>. Our field study found an overall improvement in login performance with more login sessions, including a 81% reduction in median login time to just 7 seconds by the 17<sup>th</sup> login session.

---

<sup>2</sup>We note that direct comparison between different studies should be taken with caution.

#### 6.4.0.2 Lockout rules

Lockout rules [60] 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 to this issue, as we found that 100% of participants made at most two attempts on average to authenticate successfully. Thus, GraphicV scheme is amenable to reasonable lockout rules.

### 6.5 Conclusion

The results from the lab studies showed that a cued-recognition-based scheme (e.g., GraphicV/ObjectSV) offering users with spatial cues and verbal cues for recognizing object images performed best in terms of memorability. So, we conducted a field study for a further understanding on the usability of this scheme in a real-life scenario. The memorability for GraphicV scheme was quite satisfactory in a real-life setting with an average login success rate of 98%, while the login time significantly improved with more login sessions because of training effect. Users responded with positive opinion about the usability of this scheme in a real-life scenario.

## CHAPTER 7

### Conclusion and Research Directions

In this chapter, we summarize our primary contributions and provide a direction for future research.

We argue that existing password systems fail to fully address users' cognitive limitations or leverage humans' cognitive strengths. Thus, despite a large body of research, it remains a critical challenge to build an authentication system that offers both high memorability and guessing resilience. We address this challenge in this thesis, in which we aim to design a novel authentication scheme that successfully addresses the usability-security tension in online user authentication. To achieve the goal, we argue that passwords should be randomly assigned by the system to offer resilience against password reuse and predictable patterns, and then we leverage the theories from cognitive psychology to improve the password memorability.

In the first study, we propose a novel password scheme: CuedR, which helps us to explore the efficacy of combining graphical, verbal, and spatial cues to improve the memorability of system-assigned random passwords. We evaluated CuedR scheme with a single-password lab study that showed 100% login success rate one week after registration. Although the login time was relatively high, our primary findings indicated high memorability for CuedR, suggesting that cued-recognition would be an important direction in password research to address the usability-security tension in authentication.

The first study showed promise for offering users with multiple cues to gain satisfactory login success rate for system-assigned recognition-based passwords, however,

it did not examine the impact of individual memory cue. So, in the second study, we aimed to better understand the individual impact of cues and user interaction on the memorability of system-assigned recognition-based graphical passwords, and designed seven different study conditions to achieve this goal. In a study with 56 participants, we had a 98% login success rate for a scheme offering spatial and verbal cues (ObjectSV), while a scheme based on user interaction had a 95% login success rate for face images (FaceSUI) and a 93% login success rate for object images (ObjectSUI). Our analysis show that verbal cues and user interaction made an important contribution to gain significantly higher login success rate as compared to the control conditions representing existing graphical password schemes.

Since the combination of spatial and verbal cues performed best in the second study in providing satisfactory memorability for system-assigned recognition-based graphical passwords, in the third study, we aimed to understand the impact of combining spatial and verbal cues for system-assigned recognition-based textual passwords. We 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 spatial and verbal cues (TextV), which was significantly higher than that for the Control condition providing only spatial cues. 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. To note, the GraphicV scheme in this study is similar to the ObjectSV scheme in the second study.

The results from the lab studies showed that a cued-recognition-based scheme (e.g., GraphicV/ObjectSV) offering users with spatial and verbal cues for recognizing object images performed best in terms of memorability. So, we conducted a field study



for a further understanding on the usability of this scheme in a real-life scenario. The memorability for GraphicV scheme was quite satisfactory in a real-life setting with an average login success rate of 98%, while the login time significantly improved with more login sessions because of training effect. Users responded with positive opinion about the usability of this scheme in a real-life scenario.

In our studies, the participants were university-educated and most of them were young, which represent a large fraction of frequent Web users but may not generalize to the entire population. To gain higher ecological validity, we would examine the usability of GraphicV scheme for senior users in a future work.

We would also explore the usability of our proposed scheme for people with cognitive limitations. It is important to note that people with learning disability might find it difficult to create a secure password and find an appropriate way to memorize that password. In this case, in our proposed scheme, passwords are randomly assigned by the system so that users do not require to worry about security during password-creation, and the scheme (e.g., GraphicV) offers users with multiple memory cues to help them memorizing the system-assigned random passwords. So, it would be interesting to examine the usability of this scheme for people with cognitive limitations in a future work.

## REFERENCES

- [1] M. Honan, “How Apple and Amazon security flaws led to my epic hacking,” *Wired*, Aug 06, 2012.
- [2] R. Shay, I. Ion, R. W. Reeder, and S. Consolvo, ““my religious aunt asked why i was trying to sell her Viagra”: Experiences with account hijacking,” in *CHI*, 2014.
- [3] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wangz, “The tangled web of password reuse,” in *NDSS*, 2014.
- [4] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, and L. F. Cranor, “Encountering stronger password requirements: User attitudes and behaviors,” in *SOUPS*, 2010.
- [5] D. Florencio and C. Herley, “Where do security policies come from?” in *SOUPS*, 2010.
- [6] R. W. Proctor, M.-C. Lien, K.-P. L. Vu, E. E. Schultz, and G. Salvendy, “Improving computer security for authentication of users: Influence of proactive password restrictions,” *Behavior Research Methods, Instruments, and Computers*, vol. 34(2), 2002.
- [7] E. Tulving and M. Watkins, “Continuity between recall and recognition,” *American Journal of Psych*, vol. 86(4), 1973.
- [8] J. R. Anderson and G. H. Bower, “Recognition and recall processes in free recall,” *Psychological Review*, vol. 79(2), 1972.

- [9] W. A. Wickelgren and D. A. Norman, “Strength models and serial position in short-term recognition memory,” *Journal of Mathematical Psychology*, vol. 3, 1966.
- [10] A. Paivio, *Mind and Its Evolution: A Dual Coding Theoretical Approach*. Lawrence Erlbaum: Mahwah, N.J., 2006.
- [11] D. L. Nelson, V. S. Reed, and C. L. McEvoy, “Learning to order pictures and words: A model of sensory and semantic encoding,” *Journal of Experimental Psychology: Human Learning and Memory*, vol. 3(5), 1977.
- [12] “Passfaces corporation,” The science behind Passfaces. White paper, [http://www.passfaces.com/enterprise/resources/white\\_papers.htm](http://www.passfaces.com/enterprise/resources/white_papers.htm).
- [13] D. Davis, F. Monrose, and M. Reiter, “On user choice in graphical password schemes,” in *USENIX Security*, 2004.
- [14] R. Shay, P. G. Kelley, S. Komanduri, M. L. Mazurek, B. Ur, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor, “Correct horse battery staple: Exploring the usability of system-assigned passphrases,” in *SOUPS*, 2012.
- [15] N. Wright, A. S. Patrick, and R. Biddle, “Do you see your password? Applying recognition to textual passwords,” in *SOUPS*, 2012.
- [16] K. Everitt, T. Bragin, J. Fogarty, and T. Kohno, “A comprehensive study of frequency, interference, and training of multiple graphical passwords,” in *CHI*, 2009.
- [17] C. R. Atinkson and M. R. Shiffrin, “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.
- [18] M. Knopf, A. Mack, S. Lenel, and S. Ferrante, “Memory for action events: Findings in neurological patients,” *Scandinavian Journal of Psychology*, vol. 46, 2005.

- [19] R. Biddle, S. Chiasson, and P. van Oorschot, “Graphical passwords: Learning from the first twelve years,” *ACM Computing Surveys*, vol. 44(4), 2012.
- [20] M. N. Al-Ameen, K. Fatema, M. Wright, and S. Scielzo, “Leveraging real-life facts to make random passwords more memorable,” in *ESORICS*, 2015.
- [21] —, “The impact of cues and user interaction on the memorability of system-assigned recognition-based graphical passwords,” in *Symposium on Usable Privacy and Security (SOUPS)*, 2015.
- [22] M. N. Al-Ameen, M. Wright, and S. Scielzo, “Towards making random passwords memorable: Leveraging users’ cognitive ability through multiple cues,” in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2015.
- [23] M. N. Al-Ameen and M. Wright, “Multiple-password interference in the geopass user authentication scheme,” in *USEC*, 2015.
- [24] R. Akavipat, M. N. Al-Ameen, A. Kapadia, Z. Rahman, R. Schlegel, and M. Wright, “Reds: A framework for reputation-enhanced dhTs,” *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 2, pp. 321–331, 2014.
- [25] M. N. Al-Ameen and M. Wright, “Design and evaluation of perseas, a sybil-resistant dhT,” in *Proceedings of the 9th ACM symposium on Information, computer and communications security*. ACM, 2014, pp. 75–86.
- [26] —, “Persea: a sybil-resistant social dhT,” in *Proceedings of the third ACM conference on Data and application security and privacy*. ACM, 2013, pp. 169–172.
- [27] M. N. Al-Ameen, C. Gatz, and M. Wright, “Sda-2h: Understanding the value of background cover against statistical disclosure,” *JOURNAL OF NETWORKS*, vol. 7, no. 12, 2012.
- [28] M. N. Al-Ameen, M. Shahriar, and A. Billah, “Time and space efficient algorithm for consumer’s priority product management,” in *Computer and Information*

- Technology (ICCIT), 2012 15th International Conference on.* IEEE, 2012, pp. 4–8.
- [29] M. N. Al-Ameen, C. Gatz, and M. Wright, “Sda-2h: Understanding the value of background cover against statistical disclosure,” in *Computer and Information Technology (ICCIT), 2011 14th International Conference on.* IEEE, 2011.
- [30] M. N. Al-Ameen, M. M. Hasan, and A. Hamid, “Making findbugs more powerful,” in *Software Engineering and Service Science (ICSESS), 2011 IEEE 2nd International Conference on.* IEEE, 2011, pp. 705–708.
- [31] M. N. Al-Ameen and M. Wright, “ipersea: The improved persea with sybil detection mechanism,” 2014.
- [32] —, “A comprehensive study of the GeoPass user authentication scheme,” arXiv:1408.2852 [cs.HC], Tech. Rep., 2014.
- [33] M. N. Al-Ameen, S. M. T. Haque, and M. Wright, “Q-A: Towards the solution of usability-security tension in user authentication,” arXiv:1407.7277 [cs.HC], Tech. Rep., 2014.
- [34] M. N. Al-Ameen, “An intelligent fire alert system using wireless mobile communication,” arXiv:1308.0372, Tech. Rep., 2013.
- [35] J. Campbell, W. Ma, and D. Kleeman, “Impact of restrictive composition policy on user password choices,” *Behaviour & Information Technology*, vol. 30, no. 3, pp. 379–388, 2011.
- [36] B. Ur, F. Noma, J. Bees, S. M. Segreti, R. Shay, L. Bauer, N. Christin, and L. F. Cranor, “‘ i added’ ! ’at the end to make it secure”: Observing password creation in the lab,” in *SOUPS*, 2015.
- [37] B. Ur, J. Bees, S. Segreti, L. Bauer, N. Christin, L. Cranor, and A. Deepak, “Do users perceptions of password security match reality?” in *CHI*, 2016.

- [38] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor, “Can long passwords be secure and usable?” in *CHI*, 2014.
- [39] R. Shay, L. Bauer, N. Christin, L. F. Cranor, A. Forget, S. Komanduri, M. L. Mazurek, W. Melicher, S. M. Segreti, and B. Ur, “A spoonful of sugar?: The impact of guidance and feedback on password-creation behavior,” in *CHI*, 2015.
- [40] C. Kuo, S. Romanosky, and L. F. Cranor, “Human selection of mnemonic phrase-based passwords,” in *SOUPS*, 2006.
- [41] A. Forget, S. Chiasson, P. van Oorschot, and R. Biddle, “Improving text passwords through persuasion,” in *SOUPS*, 2008.
- [42] A. Forget, “A world with many authentication schemes,” Ph.D. dissertation, Carleton University, 2012.
- [43] S. Furnell, I. Papadopoulos, and P. Dowland, “A long-term trial of alternative user authentication technologies,” *Information Management and Computer Security*, vol. 12(2), 2004.
- [44] M. Just and D. Aspinall, “Personal choice and challenge questions a security and usability assessment,” in *SOUPS*, 2009.
- [45] S. Schechter, A. J. B. Brush, and S. Egelman, “It’s no secret: Measuring the security and reliability of authentication via ‘secret’ questions,” in *IEEE S&P*, 2009.
- [46] A. Rabkin, “Personal knowledge questions for fallback authentication: Security questions in the era of Facebook,” in *SOUPS*, 2008.
- [47] I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin, “The design and analysis of graphical passwords,” in *USENIX Security*, 1999.
- [48] D. Nali and J. Thorpe, “Analyzing user choice in graphical passwords,” School of Computer Science, Carleton University, Tech. Rep. TR-04-01, 2004.

- [49] P. Dunphy and J. Yan, “Do background images improve “Draw a Secret” graphical passwords?” in *CCS*, 2007.
- [50] S. Chiasson, R. Biddle, and P. C. van Oorschot, “A second look at the usability of click-based graphical passwords,” in *SOUPS*, 2007.
- [51] S. Chiasson, P. C. van Oorschot, and R. Biddle, “Graphical password authentication using cued click points,” in *ESORICS*, 2007.
- [52] A. E. Dirik, N. Memon, and J.-C. Birget, “Modeling user choice in the passpoints graphical password scheme,” in *SOUPS*, 2007.
- [53] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle, “Persuasion for stronger passwords: Motivation and pilot study,” in *PT*, 2008.
- [54] S. Chiasson, E. Stobert, R. Biddle, and P. van Oorschot, “Persuasive cued click-points: design, implementation, and evaluation of a knowledge-based authentication mechanism,” *IEEE TDSC*, vol. 9, 2012.
- [55] S. Chiasson, A. Forget, R. Biddle, and P. van Oorschot, “User interface design affects security: Patterns in click-based graphical passwords,” *International Journal of Information Security*, vol. 8(6), 2009.
- [56] S. Chiasson, A. Forget, E. Stobert, P. C. van Oorschot, and R. Biddle, “Multiple password interference in text and click-based graphical passwords,” in *CCS*, 2009.
- [57] J. Thorpe, B. MacRae, and A. Salehi-Abari, “Usability and security evaluation of GeoPass: A geographic location-password scheme,” in *SOUPS*, 2013.
- [58] H. Sun, Y. Chen, C. Fang, and S. Chang, “A map based graphical-password authentication scheme,” in *ASIACCS*, 2012.
- [59] J. Spitzer, C. Shingh, and D. Schweitzer, “A security class project in graphical passwords,” *Journal of Computing Sciences in Colleges*, vol. 26 (2), p. 7, 2010.
- [60] D. Florencio, C. Herley, and B. Coskun, “Do strong Web passwords accomplish anything?” in *HotSec*, 2007.

- [61] D. Silver, S. Jana, D. Boneh, E. Chen, and C. Jackson, "Password managers: Attacks and defenses," in *USENIX Security*, Aug. 2014.
- [62] Z. Li, W. He, D. Akhawe, and D. Song, "The emperor's new password manager: Security analysis of Web-based password managers," in *USENIX Security*, Aug. 2014.
- [63] F. Tari, A. Ozok, and S. Holden, "A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords," in *SOUPS*, 2006.
- [64] N. Provos and D. Mazieres, "A future-adaptable password scheme," in *USENIX ATC*, 1999.
- [65] B. Kaliski, "RFC 2898: PKCS #5: Password-based cryptography specification version 2.0," Sep. 2000. [Online]. Available: <https://www.ietf.org/rfc/rfc2898.txt>
- [66] E. Tulving and D. M. Thompson, "Encoding specificity and retrieval processes in episodic memory," *Psychological Review*, vol. 80(5), 1973.
- [67] K. Renaud, "A visuo-biometric authentication mechanism for older users," in *British HCI*, 2005.
- [68] J. Yan, A. Blackwell, R. Anderson, and A. Grant, "Password memorability and security: Empirical results," *IEEE Security and Privacy*, vol. 2 (5), p. 25, 2004.
- [69] M. Hlywa, R. Biddle, and A. S. Patrick, "Facing the facts about image type in recognition-based graphical passwords," in *ACSAC*, 2011.
- [70] J. Bonneau, "The science of guessing: Analyzing an anonymized corpus of 70 million passwords," in *IEEE S&P*, 2012.
- [71] E. Hayashi and J. I. Hong, "A diary study of password usage in daily life," in *CHI*, 2011.
- [72] J. Nicholson, L. Coventry, and P. Briggs, "Age-related performance issues for PIN and face-based authentication systems," in *CHI*, 2013.



- [73] M. Peters, “Revised Vandenberg and Kuse mental rotations tests: Forms MRT-A to MRT-D,” Department of Psychology, University of Guelph, Tech. Rep., 1995.
- [74] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, “Authentication using graphical passwords: Effects of tolerance and image choice,” in *SOUPS*, 2005.
- [75] F. Schaub, M. Walch, B. Konings, and M. Weber, “Exploring the design space of graphical passwords on smartphones,” in *SOUPS*, 2013.
- [76] J. Bonneau and S. Preibusch, “The password thicket: Technical and market failures in human authentication on the Web,” in *WEIS*, 2010.
- [77] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini, “The distributed human neural system for face perception,” *Trends in Cognitive Science*, vol. 4, p. 223, 2000.
- [78] D. A. Minnebusch, B. Suchan, O. Koster, and I. Daum, “A bilateral occipitotemporal network mediates face perception,” *Behavioural Brain Research*, vol. 198 (1), p. 179, 2009.
- [79] M. N. Al-Ameen, M. Wright, and S. Scielzo, “Towards making random passwords memorable: Leveraging users’ cognitive ability through multiple cues,” arXiv:1503.02314 [cs.HC], Tech. Rep., 2015.
- [80] S. Fahl, M. Harbach, Y. Acar, and M. Smith, “On the ecological validity of a password study,” in *SOUPS*, 2013.
- [81] J. Robertson, “Stats: We’re doing it wrong,” <http://cacm.acm.org/blogs/blog-cacm/107125-stats-were-doing-it-wrong/fulltext>, April 2011.
- [82] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, “Passpoints: Design and longitudinal evaluation of a graphical password system,” *International Journal of Human-Computer Studies*, vol. 63(1-2), 2005.
- [83] S. Chiasson, R. B. A. Forget, and P. C. van Oorschot, “Influencing users towards better passwords: Persuasive cued click-points,” in *British HCI*, 2008.