Old Dominion University
ODU Digital Commons

Psychology Theses & Dissertations

Psychology

Spring 2019

Identifying the Strengths and Weaknesses of Over-the-Shoulder Attack Resistant Prototypical Graphical Authentication Schemes

Ashley Allison Cain
Old Dominion University

IDENTIFYING THE STRENGTHS AND WEAKNESSES OF OVER-THE-SHOULDER ATTACK RESISTANT PROTOTYPICAL GRAPHICAL AUTHENTICATION SCHEMES

by

Ashley Allison Cain
B.A. December 2009, University of California Santa Cruz
M.A. May 2016, San Jose State University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement of the Degree of

DOCTOR OF PHILOSOPHY

HUMAN FACOTORS PSYCHOLOGY

OLD DOMINION UNIVERSITY

May 2019

Approved by:

Jeremiah D. Still (Director)

Jing Chen (Member)

Cong Wang (Member)

Abstract

IDENTIFYING THE STRENGTHS AND WEAKNESSES OF OVER-THE-SHOULDER ATTACK RESISTANT PROTOTYPICAL GRAPHICAL AUTHENTICATION SCHEMES

Ashley Allison Cain
Old Dominion University, 2019
Director: Dr. Jeremiah D. Still

Authentication verifies users' identities to protect against costly attacks. Graphical authentication schemes utilize pictures as passcodes rather than strings of characters. Pictures have been found to be more memorable than the strings of characters used in alphanumeric passwords. However, graphical passcodes have been criticized for being susceptible to Over-the-Shoulder Attacks (OSA). To overcome this concern, many graphical schemes have been designed to be resistant to OSA. Security to this type of attack is accomplished by grouping targets among distractors, translating the selection of targets elsewhere, disguising targets, and using gaze-based input.

Prototypical examples of graphical schemes that use these strategies to bolster security against OSAs were directly compared in within-subjects runoffs in studies 1 and 2. The first aim of this research was to discover the current usability limitations of graphical schemes. The data suggested that error rates are a common issue among graphical passcodes attempting to resist OSAs. Studies 3 and 4 investigated the memorability of graphical passcodes when users need to remember multiple passcodes or longer passcodes. Longer passcodes provide advantages to security by protecting against brute force attacks, and multiple passcodes need to be investigated as users need to authenticate for numerous accounts. It was found that participants have strong item retention for passcodes of up to eight images and for up to eight accounts. Also these studies leveraged context to facilitate memorability. Context slightly improved the memorability of graphical passcodes when participants needed to remember credentials for eight accounts. These studies take steps toward understanding the readiness of graphical schemes as an authentication option.

ACKNOWLEDGMENTS

I want to thank Dr. Jeremiah Still for his guidance. His honest and constructive feedback has kept me on the right track throughout the creation of my dissertation. I am lucky to have an enthusiastic, communicative advisor who motivates his students to invest themselves in the research process. I would also like to thank my committee members, Dr. Jing Chen and Dr. Cong Wang, whose input has led to improvements in the dissertation research. Dr. Mary Still has also given me excellent advice. She seems to effortlessly point me in the right direction when I am feeling confused or blocked. Along the way, my Research Assistants, James Unverricht, Colin McDowell, Jen McClaren, Morgan Edwards, Paige Duplantis, and Lauren Tiller, have been a huge help coding data and running participants. They have also helped to make work fun.

My friends from graduate school have been awesome for their support when I want to complain about stresses and for their companionship when I want to celebrate successes. I want to thank my family for teaching me to value education and to be ambitious. My mom and dad, Pamela and Christopher Cain, have been wonderful fielding my daily Facetime calls. Pamela made sure to tell me she wanted to be mentioned by name in my acknowledgments. I want to thank my sister, Deirdre Cain, for being my twin and confidant, and my sister, Hillary Matheson, for being my second mom. Lastly, I want to let my husband, David Gardner, know that I appreciate how he has been there for me and kept my personal life fun during these years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF GRAPHS

CHAPTER 1: INTRODUCTION

Statement of the Problem

The frequency and cost of cybersecurity attacks, such as worm attacks, is increasing (Walters, 2014). The average cost of a breach in data has more than doubled since 2010 (Walters, 2014), with the cost reaching an average of $4 million per breach in 2016 (per a survey of 383 companies across 12 countries; IBM, 2016). For most attacks, a vulnerability in authentication must be exploited (Zviran & Haga, 1999). Authentication protects valuable or confidential information (e.g., company files, banking information, and health records) by requiring users to confirm their identity. A user is granted access to a network or system if they can confirm something they know (e.g., a password), something they have (e.g., a token), or something they are (e.g., a fingerprint; Cazier & Medlin, 2006).

Alphanumeric passwords, a knowledge-based scheme, are the most commonly used authentication scheme (Grawemeyer & Johnson, 2011; Zviran & Haga, 1999). Alphanumeric passwords have widespread use because they are effective, efficient, subjectively satisfactory, and learnable. These passwords offer security against attacks, such as guessing attacks or worm attacks, when they have a large dimensional space (Zviran & Haga, 1999). They should be long and complex (Barton & Barton, 1984; Choong & Greene, 2016). They should not contain common words, and they should contain numbers and symbols (Barton & Barton, 1984; Choong & Greene, 2016). To be secure, alphanumeric passwords, should not be written down, they should be different for every account, and they should be changed often (Barton & Barton, 1984). Users have difficulty applying the given rules for creating strong passwords, especially as guidance varies from system to system (Choong & Greene, 2016). In a recent study, Choong and Greene (2016) asked participants to classify passwords as to whether or not they comply with a given set of rules. Although “special characters,” “symbols,” and “non-alphanumeric characters” have the same meaning, participants interpreted rules differently depending on how the rules were explained. Even when users apply password rules, stronger security can lead to trade-offs with usability. End users deal with limitations of usability by using “workarounds” and not using the system as it was intended to be used (Grawemeyer & Johnson, 2011). Long, complex strings of characters, symbols, and numbers are hard to remember (Zviran & Haga, 1999). Memorability is further hindered by the considerable number of passwords users have and by the need to routinely change passwords (Still, Cain, & Schuster, 2017). The security of alphanumeric passcodes is often undermined by users when they write them down or share passwords with loved-ones to solve problems with memorability (Grawemeyer & Johnson, 2011; Kaye, 2011; Paans & Herschberg, 1987). When forced to use unfamiliar alphanumeric passwords to bolster security, users are 18 times more likely to write them down (Grawemeyer & Johnson, 2011). A third of users report sharing their email password with someone else (Kaye, 2011). Users also undermine security to aid memorability by reusing passwords (Grawemeyer & Johnson, 2011). Up to 50% of alphanumeric passwords are reused (Grawemeyer & Johnson, 2011), and they are typically reused for 1.7 to 3.4 websites (Wash, Rader, Berman, & Wellmer, 2016). Although usability and security compete when selecting alphanumeric passcodes, alternate approaches to authentication may be able to provide both usability and security.

Biometric authentication is becoming increasingly prevalent, and the marketplace is highly dependent on this type of authentication. With biometric authentication, users can authenticate using their personal characteristics, such as a fingerprint or face. The researchers do not believe biometric authentication is an all-encompassing solution because of its limitations. Firstly, the data from the biometric is stored in a database. When the database is eventually hacked, the passcode can never be changed and the user's identity is compromised for life (Jain & Nandakumar, 2012). Secondly, biometrics hold information about a user's health, which should be private and not shared with companies that are requiring authentication (Jain & Nandakumar, 2012). Lastly, biometrics are not secrets. Therefore, attackers always have the options of forcing a user to authenticate. In fact, police are using the strategy of forcing users to unlock their phones to find incriminating evidence (Waddel, 2017). The researchers focus on other solutions because of the shortcomings inherent in biometric authentication. Security requirements, such as whether it is desirable to have two-factor authentication, can guide which type of authentication to use.

Graphical authentication schemes are knowledge-based approaches that utilize pictures as passcodes rather than complex strings of characters. Graphical passcodes offer a solution to the problem of memorability that accompanies the alphanumeric scheme (Biddle, Chiasson, & Van Oorschot, 2012) and are not accompanied by the issues of privacy and security that accompany biometric authentication. The pictures used in graphical passcodes are more easily remembered than the strings of characters used in alphanumeric passwords because pictures allow for a greater depth of cognitive processing. The picture superiority effect explains that pictures are dual encoded both visually and semantically, whereas alphanumeric passcodes are only encoded semantically (Paivio, 1979). Further, pictures typically have more features than individual letters and numbers, thereby also facilitating retrieval.

Although a graphical passcode is generally more memorable than an alphanumeric password, there may be limits to the memorability of graphical passcodes when users have many passcodes to remember or longer passcodes to remember. These are important topics to investigate as users have many accounts that require authentication and because longer passcodes can improve security against brute force attacks. Also, although graphical schemes offer advantages for memory, they must also meet other usability needs to be considered as an alternative to the alphanumeric scheme. First, authentication systems must allow for quick access (Still et al., 2017) comparable to log in times for alphanumeric passwords (e.g., approximately 5 seconds; Wiedenbeck, Waters, Birget, Brodskiy, & Memon, 2005). Authentication is a secondary task that serves as a gateway to the primary goal, so authentication processes need to be quick. Login times for elaborate graphical schemes may not meet the need for quick access (Sreelatha, Shashi, Anirudh, Ahamer, & Kumar, 2011) because it may take users longer to search for or recognize images. Second, to be usable, appropriate actions should be apparent to a wide range of users so that logins are successful with little training (Still et al., 2017). Confusion when authenticating will frustrate users when they are blocked from their primary goal. High error rates or poor learnability over time may reflect a lack of transparency for the actions needed to authenticate. The same issues could arise due to a lack of accessibility (Behl, Bhat, Ubhaykar, Godbole, & Kulkarni, 2014). Measurements of login times, error rates, and learnability can be used to index the usability of novel graphical schemes, and subjective satisfaction can reflect users' reactions to these objective dimensions.

Graphical schemes need to meet requirements of security as well as usability. The most cited vectors for attacks against graphical schemes are intersection attacks and over-the-shoulder attacks (OSA). Intersection attacks happen when attackers go through authentication challenges, count the number of times that an image appears, and then attempt to log in using the most frequently used images. The more challenges there are in a login, the greater the risk of an intersection attack. To defend against intersection attacks, there should be the same distractor images at each login (English & Poet, 2012). Each target should be paired with a small set of the same distractors each time. In this way, the distractors will be seen as frequently as the targets. As a further countermeasure, if a distractor is selected, the scheme should show only dummy screens that include only distractors (English & Poet, 2012). Another good countermeasure is to lock out attackers after a certain number of failed attempts (English & Poet, 2012). This strategy prevents them from trying different passwords that they think might be likely.

Another prominent susceptibility has been Over-the-Shoulder Attacks (OSA). OSAs occur when a casual observer steals a passcode in a public place. Images associated with some graphical passcodes can be clearly observed on the screen. Just as users can quickly recognize and remember pictures in their passcodes, attackers may be able to casually peek at and reproduce the pictures. Casual attackers are characterized by having minimal resources, minimal knowledge, and as being opportunistic. OSAs frequently happen in public places, such as on public transportation (Eiband, Khamis, von Zezschwitz, Hussmann, & Alt, 2017). These invasions of privacy are not always malicious, but they can be, and they bring up negative feelings (Eiband et al., 2017). Vulnerability increases if an attacker has the opportunity to view a login more than once. Concern over OSA vulnerability has delayed broader deployment of these schemes. To overcome this concern, many graphical schemes have been designed to resist OSAs by allowing for non-direct selection of targets, by obscuring the appearance of the targets, or by obscuring target selection (Hayashi et al., 2008; Khot et al., 2012; Wiedenbeck et al., 2006). I have identified in the literature four strategies that defend against OSAs. 1. Grouping targets among distractors (Manjunath, Satheesh, Saranyadevi, & Nithya, 2014; Wiedenbeck et al., 2006). Users can select a group of images rather than directly selecting targets. 2. Translating targets to another location (De Luca, Hertzschuch, & Hussmann, 2010; Khot et al., 2012). Passcodes are also obscured when users translate targets to another location rather than directly clicking them. 3. Disguising targets (Cain & Still, 2016; Hayashi et al., 2008). Disguising targets, such as by degrading images, can interfere with an attacker's recognition of the passcodes. 4. Using gaze-based input (De Luca, Denzel, & Hussmann, 2009). Gaze-based input allows users to select targets using their eyes, which is difficult for an attacker to observe. These defenses keep out opportunistic attackers similarly to how a lock on a door functions. A locked door deters intruders who are casually looking for entry but does not keep out determined intruders. Similarly, OSA resistant graphical schemes deter casual attackers but do not protect against sophisticated attackers. I determined these classifications through a comprehensive literature search. I found articles using keywords such as graphical password, authentication, and over-the-shoulder attack. Many of the articles came from HCI journals and CHI. An excel spreadsheet was built to organize the schemes by their security strategy, and categories of defense strategies were identified by commonalities listed in the excel sheet. Lastly, I selected prototypical examples from each class of schemes that best represented the strategies. The prototypical examples were commonly cited in the literature and had been well evaluated in usability studies by their creators.

Prototypical Approaches

Grouping. Previous schemes have avoided the direct selection of targets and the observability that comes with it by allowing users to select a group of distractors with a target. When the user selects the group, it is unclear to an attacker which image is the target and which are the distractors. The S3PA scheme (Vachaspati, Chakravarthy, & Avadhani, 2013) presents symbols on an 8x8 grid. Passcodes consist of three symbols. Users authenticate by clicking inside of the triangular region formed by their targets. Schemes with the same premise were described by Rajavat, Gala, and Redekar (2015) and Zhao and Li (2007). Joshuva, Rani, and

John’s (2011) scheme present a background image, and points on the image compose a passcode. Users authenticate by clicking inside the region created by the target points. Manjunath and colleagues’ (2014) scheme presents a color wheel with symbols in each segment of the wheel. Passcodes consist of a color and a symbol. Users authenticate by turning the wheel so that their symbol lines up with their color. Qian, Song, Huang, and Lai’s (2013) scheme displays one large image encircled by a disk of smaller images. Passcodes consist of a small image and a point on the large image. To authenticate, users turn the disk until the target small image lines up with a target point on the large image. No experiment was performed to assess the usability or security of S3PA, Rajavat and colleagues’ (2015) scheme, Zhao and Li’s (2007) scheme, Joshuva and colleagues’ (2011) scheme, Manjunath and colleagues’ (2014) scheme, or Qian and colleagues’ (2013) scheme.

Sreelatha and colleagues’ (2011) scheme presents a grid of letters. Users’ passcodes consist of two letters on the grid. Users authenticate by selecting the letter that is at the intersection of their passcode letters, rather than clicking directly on their passcode. Behl and colleagues’ (2014) scheme similarly presents characters on a grid. Users’ passcodes consist of four characters. They authenticate by selecting the letter at the intersection of the four passcode characters. Sreelatha and colleagues (2011) reported login times of 29.95 seconds, and Behl and colleagues (2014) reported login times of five seconds for their schemes of finding the intersections of letters. Behl and colleagues (2014) found that participants successfully authenticated with their scheme at a rate of 80%.

Convex Hull Click (CHC; Wiedenbeck et al., 2006) displays an interface of icons. Some icons compose a passcode. Users authenticated by selecting an icon within the shape formed by their targets. Depending on the security required, users could repeat this process once or multiple times. CHC had login times of 71.66 seconds. Users successfully authenticated using CHC with success rates of 90.35%. CHC was found to be memorable. After a one-week delay, 14 out of 15 participants remembered their pass icons.

Translating to another location. Other schemes offer resistance by allowing users to transfer targets elsewhere rather than clicking directly on them. Similar to the grouping schemes, these schemes avoid the direct selection of targets. Van Oorschot and Wan’s (2009) scheme presents numbered images on a grid. Clickable numbers are also presented below the grid. Users authenticate by selecting the numbers below the grid associated with their target images. Rokade, Hasan, and Mahajan’s (2014) scheme displays a grid of images. Each image has a letter associated with it (e.g., “a” for apple). Passcodes consist of four images. Users authenticate by typing the four letters corresponding to their images. No experiment was performed to assess Van Oorschot and Wan’s (2009) scheme or Rokade and colleagues’ (2014) scheme.

GrIDsure (Brostoff, Inglesant, & Sasse, 2010) presents a 5x5 grid of images. There are four target images. Each image has a one-time, random number. Users authenticate by typing the numbers associated with their target images. Zangooei, Mansoori, and Welch’s (2012) scheme presents a grid of images. The grid is then replaced by a grid of one-time, random numbers and letters. Users authenticate by typing the numbers that were in the locations where their images had been. ColorPIN (De Luca et al., 2010) presents a grid of numbers with three letters below each. The numbers are in black, white, and red. The passcodes consist of a number and a color. Users authenticate by typing the number that was the target color for a target number. Login times were 8 seconds for Zangooei and colleagues’ (2012) scheme and 13.88 seconds for ColorPIN (De Luca et al., 2010). All participants could log in at least once in three attempts using ColorPin. Ninety-one percent of authentication attempts were successful for GrIDsure after three to four days, 97% after nine to ten days, and 27% after one year (Brostoff et al., 2010). Memorability was demonstrated for Zangooei and colleagues' scheme. Twenty-seven out of 30 participants remembered their passcode after a one-week delay. OSAs were explored for Zangooei and colleagues' scheme and ColorPin. A simulated attacker would be able to steal the passcode 4% of the time for ColorPin. When participants took on the role of attacker, none could identify the full passcode for Zangooei and colleagues' scheme.

SSSL (Perkovic, Cagalj, & Rakic, 2009) presents numbers on a 5x5 grid. Users authenticate by using arrows on the side of the grid to select their targets. Pressure-faces (Kim, Dunphy, Briggs, Hook, Nicholson, Nicholson, & Olivier, 2010) displays a 3x3 grid of faces on a tabletop, touch-sensitive interface. Each face had a pressure level associated with it, and the passcode consists of target faces and their pressures. On the interface, pressure sensitive bars extend from the grid. Users place their fingers on bars extending from the grid of faces to authenticate. Login times were eight seconds for SSSL and 11 seconds for Pressure-faces. Success rates for SSSL were 93%. Subjectively, 11 out of 15 participants reported that SSSL was easy to learn and use. As attackers, no participant could identify the full passcode for Pressure-faces.

CBFG (Liu, Qiu, Ma, Gao, & Ren, 2011b) displays a background image. The image is broken into numbered cells. Passcodes consist of points on the image. Users authenticate by selecting numbers on the side of the image corresponding to the cells of their target locations. PassMatrix (Sun, Chen, Yeh, & Cheng, 2016) also displays a background image divided into cells, and passcodes are points on the image. Users authenticate by using a scroll bar on the side to select cells containing their target image locations. Login times were 21.4 seconds for CBFG and 31.11 seconds for PassMatrix. Success rates were 92.3% for CBFG. Success rates were

86.67% for PassMatrix on day one and were 66.67% after two weeks. When participants took on the role of attacker, none could identify the full passcode for CBFG or PassMatrix.

Passblot (Gupta, Sahni, Sabbu, Varma, & Gangashetty, 2012) displays images of inkblots. Users authenticate by typing words they associate with the inkblots. What You See is What You Enter (WYSWYE; Khot et al., 2012) displays a 5x5 grid of images on the left side of the interface. On the right side is a blank, 4x4 grid. A passcode consists of four images. On the 5x5 grid, there is one row and one column that does not contain a target image. To authenticate, users mentally delete the row and column with no target, mentally shift the remaining cells together, and click the location on the blank grid. Login times were 23.74 seconds for Passblot and 35.5 seconds for WYSWYE. All participants could log in at least once in three attempts using WYSWYE. Success rates were 98.5% for Passblots on day one and 84.6% one week later. For WYSWYE, dimensions of satisfaction (e.g., preference and ease of use) were rated between 64 and 84%.

Disguising. Graphical schemes have been made resistant by disguising targets to interfere with attackers' recognition processes. R-Das (Chakrabarti, Landon, & Singhal, 2007) integrates rotation as a method for disguising Draw a Secret (DAS; Jermyn, Mayer, Monrose, Reiter, & Rubin, 1999). Using DAS, users draw a freehand doodle to authenticate. The rotation of R-Das introduces another dimension that an attacker must also detect. YAGP (Liu et al., 2011a) created a version of DAS that can be drawn smaller and anywhere on the screen to help hide the passcode. Zakaria, Griffiths, Brostoff, and Yan (2011) disguised DAS using decoy strokes, disappearing strokes, and a line snaking. Decoy strokes are additional lines drawn by the system. Disappearing strokes hide the passcode as soon as it is completely drawn. Line snaking hides the passcode while it is being drawn. Login times were 3.4 seconds for YAGP. Success rates were

76% for YGAP. Thirty-eight out of 42 participants could remember their passcode one week later. When participants took on the role of attacker, YGAP passcodes were stolen in 7% of attempts. For Zakaria and colleagues' decoy strokes, 77% of strokes were stolen, for disappearing strokes 40% were stolen, and for line snaking 50% were stolen.

Facelock (Jenkins, McLachlan, & Renaud, 2014) presents faces on a 3x3 grid similarly to Passface (RealUser, 2005). Users select target faces to authenticate. Different photographs of the same individuals are used at each login to confuse attackers. Incognito (Still & Bell, 2018) presents a 3x3 grid of numbers. As users log in to this PIN interface, the button that is being selected is highlighted. However, decoy buttons are also highlighted. Input from the mouse cursor or the gaze-guided mouse cursor is made invisible. Sasamoto and colleagues' (2008) scheme presents pictures that are distorted by removing detail but retaining general colors and shapes. Users could authenticate by using a tactile, rotating ball to select their targets. RSVP (Cain & Still, 2016) presents degraded images in rapid succession temporally rather than statically. Line drawings are degraded by removing lines for curvature and intersection to interfere with attackers' object recognition. Users authenticate by hitting the space bar or tapping the screen when they see their targets. Use Your Illusion (UYI; Hayashi et al., 2008) presents three subsequent, 3x3 grids of degraded images. Detail is removed, but color and shape are maintained. Passcodes are three images with one on each grid. Users authenticate by directly selecting their targets. Login times were over a minute for Sasamoto and colleagues' scheme, and login times were between 11.5 and 24.7 seconds for UYI. Participants needed an average of 1.42 attempts to log in using Incognito when using the mouse and 1.65 attempts when using eye-gaze. Most participants made one error or no errors for Sasamoto and colleagues' scheme. All participants correctly logged in at least once in three attempts using RSVP and UYI. When testing for UYI memorability, success rates dropped to 89% three weeks later. As attackers, 1.9% of attackers stole Facelock passcodes. Sasamoto and colleagues' scheme was resistant to OSA so long as the hand completely covered the rotating ball. No attacker identified the whole passcode for Incognito or RSVP.

Gaze-based input. OSAs can be protected against by having users authenticate using their eyes. Eye gaze is harder for an attacker to observe than mouse and touch input. Dunphy, Fitch, and Olivier's (2008) scheme allow users to select faces similarly to Passface (RealUser, 2005) using gaze input. Bulling and colleagues' (2012) scheme presents a background image. Users authenticate by looking at target locations on the image. EYE-Pass Shapes (De Luca et al., 2009) presents a 4x3 grid of dots. Users authenticate by making a pattern on the grid using their eyes.

Login times were 24.9 seconds for Dunphy and colleagues' (2008) scheme and were 5.8 seconds for EYE-Pass Shapes (De Luca et al., 2009). Participants often logged in on their first attempts using Dunphy and colleagues' scheme. Using EYE-Pass Shapes, all participants logged in at least once in three attempts. Three days later, participants made very few requests to remind them what their target faces were. 71% of participants could remember their pattern for EYE-Pass Shapes after a ten-day delay. When participants took on the role of attacker, they needed two or three attempts to identify passcodes for EYE-Pass Shapes. Subjectively, participants reported that Bulling and colleagues' (2012) scheme and EYE-Pass Shapes were not as easy to use as traditional PIN passcodes.

Needs Addressed by the Current Studies

Previous literature has offered many schemes for graphical authentication that are designed to be OSA resistant. Many schemes have been experimentally tested to determine their usability and security (Hayashi et al., 2008; Khot et al., 2012; Wiedenbeck et al., 2006). Schemes have also been compared with traditional PIN authentication (Bulling et al., 2012; De Luca et al., 2009). Because different methods were used by different experimenters to assess each scheme, comparisons among schemes are difficult. For example, different amounts of training are given before experimental trials, and there are different lengths of delays before measuring memorability. OSA measures may allow for one or multiple viewings of the passcode, they may allow for one or multiple attempts to identify a passcode, and they may be motivated by reward or not. Satisfaction was measured by a variety of methods.

Limited previous studies have directly compared graphical schemes. Schaub, Walch, Könings, and Weber (2013) compared five graphical schemes that allow for authentication on small touchscreen devices using strategies of recall and cued-recall. The schemes were also compared with the PIN scheme. UYI was the only scheme included in Schaub and colleagues' analysis that was designed to be resistant to OSA. They found that the graphical schemes had similar usability to the PIN scheme, and they were more resistant to OSA on small touch screens than the PIN scheme.

Johnson and Werner (2008) compared the memorability of four graphical passcodes and an alphanumeric password after 30 minutes and one week. The prototypes of graphical schemes the researchers included combined an image with a background, had grids of faces, had grids of images, and had one large image with click points. All of the graphical schemes were more memorable than the alphanumeric scheme.

The current studies provide a direct comparison of prototypical OSA resistant passcodes and an alphanumeric passcode. CHC (Wiedenbeck et al., 2006) represented graphical passcodes that are made resistant to OSA by allowing users to authenticate without clicking directly on the targets. WYSWYE (Khot et al., 2012) represented graphical passcodes that are made resistant to OSA by translating targets to another location. UYI (Hayashi et al.) represented a group of graphical passcodes that disguise targets. Eye-Pass Shapes (De Luca et al., 2009) represented passcodes that are entered using gaze to obstruct OSA.

To advance the study and eventual use of graphical schemes, it is important to consider how the OSA-resistant schemes compare to traditional passcodes (i.e., De Luca et al., 2010; Sasamoto, Christin, & Hayashi, 2008) and how they compare to each other (i.e., Bulling, Alt, & Schmidt, 2012; Cain & Still, 2016; De Luca et al., 2009; Liu, Gao, Wang, & Chang, 2011). Studies 1 and 2 directly compared the usability and security of prototypical OSA-resistant graphical schemes (grouping, translating to another location, disguising, and gaze-based input), a gaze-based scheme, and an alphanumeric scheme in a within-subjects design. Their effectiveness regarding usability - error rates, login times, learnability, memorability, acceptance, and satisfaction - and security - OSA resistance were explored. Findings showed how graphical schemes that use different strategies for security compare with each other and with the traditional alphanumeric scheme on usability requirements of memorability, quick access, accessibility, satisfaction, and security.

CHAPTER 2: STUDY 1: USABILITY AND SECURITY RUNOFF

Method

Participants. Twenty undergraduate students participated (females = 11). They were recruited through the SONA system and compensated with class research credit. One participant reported being left hand dominant. Ages ranged from 18 to 53 (*M* = 23.05, *SD* = 8.60). Reported computer use ranged from 3 to 15 hours a day (*M* = 7.2, *SD* = 3.28). All participants reported normal or corrected to normal vision.

Stimuli and Apparatus. Five prototypes of authentication schemes were created for this study. Four graphical schemes were based on Eye-Pass Shapes (a gaze-based scheme; De Luca et al., 2009), Convex-Hull Click (CHC; Wiedenbeck et al., 2006), Use Your Illusion (UYI; Hayashi et al., 2008), and What You See is Where You Enter (WYSWYE; Khot et al., 2012). These four schemes were compared to an alphanumeric scheme. CHC, UYI, WYSWYE, and alphanumeric prototypes were presented on a Windows desktop computer with a 24-inch monitor. Their presentation and data collection (selection locations and login times) was controlled using Paradigm©. The gaze-based scheme was presented on a Windows desktop computer with a 16-inch monitor.

Gaze-based scheme

The gaze-based prototype based on Eye-Pass Shapes consisted of a 3x4 dot configuration with an *Enter* button on the upper right (see figure 1). The grid was 513x379 pixels. Each dot had a radius of 50 pixels, and the selection area for each was a 120x120 pixel square. The interface was implemented using HTML. The internet browser was Firefox. An Eye Tribe © eye-tracker was used to control the mouse cursor. Java code made the mouse cursor invisible while over the grid of dots. Java code measured the time of every selection of the enter button and determined if it was correct. The response time and feedback of *correct* or *incorrect* was presented below the grid of dots. Dragger© made a selection every .7 seconds at the locations of the invisible mouse cursor. A jitter box of 22 pixels controlled minor movements of the mouse cursor. The passcode consisted of four dots in sequential order. Every participant used this same system-assigned passcode. A researcher recorded the time of all attempts to log in. This graphical scheme represents a collection of implementations offered by a variety of research groups (Arianezhad, Stebila, & Mozaffari, 2013; Bulling, Alt, & Schmidt, 2012; Dunphy, Fitch, & Olivier, 2008; Forget, Chiasson, & Biddle, 2010; Hoanca & Mock, 2006; Kumar, Garfinkel, Boneh, & Winograd, 2007).

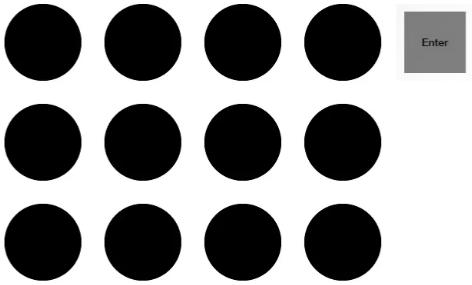

Figure 1. Prototype of a gaze-based graphical scheme

CHC. CHC consisted of icons on a 10x15 grid (see figure 2). The icons came from an online, open source database (http://www.fatcow.com/free-icons). The grid was 4138x1126 pixels. Each icon was 55x45 pixels. The passcode consisted of three system-assigned icons. Because there were three target icons, they would always form a triangle shape on the grid (see figure 3). Target icons were never located in a straight line. A correct login occurred when a participant selected one time anywhere inside the triangular region created by the three icons. They were told not to click directly on target icons and not to hover the mouse cursor over their target icons. The researcher provided verbal feedback of *correct* or *incorrect* after each authentication attempt. After each attempt, the icons were repositioned. This graphical scheme that was created represents a collection of implementations (Ankush & Husain, 2014; Behl et al., 2014; Chen, Ku, Yeh, Liao, 2013; Joshuva, Rani, & John, 2011; Kiran, Rao, & Rao, 2012; Li, Sreelatha et al., 2011; Sun, Lian, & Giusto, 2005; Manjunath et al., 2014; Rao & Yalamanchili, 2012; Tao, 2006; Vachaspati, Chakravarthy, & Avadhani, 2013; Zhao & Li, 2007).

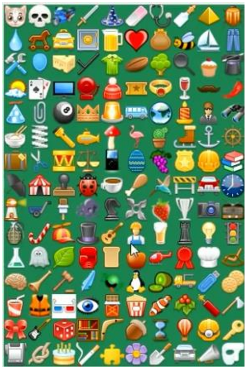

Figure 2. Prototype of CHC

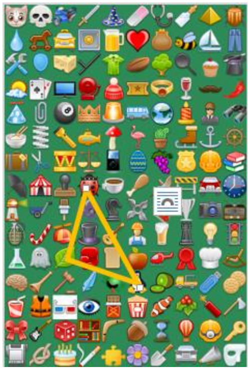

Figure 3. Three target icons form a region.

UYI. UYI was presented as images in a 3x3 grid that were degraded by removing detail but retaining general colors and shapes (see figure 4). The grid was 774x571 pixels. Each image was 213x175 pixels. A passcode consisted of three system-assigned images. A correct login occurred when a participant selected the degraded versions of each of their three targets on three subsequent grids. The researcher provided verbal feedback of *correct* or *incorrect* after each authentication attempt. After each attempt, the images were repositioned. This graphical scheme represents a collection of implementations offered by a variety of research groups (Cain & Still, 2016; Gao, Guo, Chen, Wang, & Liu, 2008; Ghori & Abbasi, 2013; Hui, Bashier, Hoe, Kwee, & Sayeed, 2014; Jenkins, McLachlan, & Renaud, 2014; Lin, Dunphy, Olivier, & Yan, 2007; Liu et al., 2011; Meng & Li, 2013; Nicholson, 2009; Sasamoto et al., 2008; Yakovlev & Arkhipov, 2015; Zakaria, Griffiths, Brostoff, & Yan, 2011).

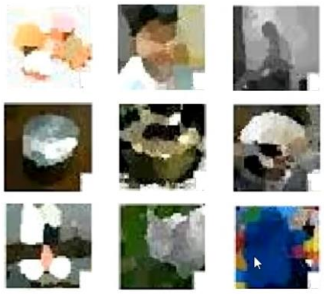

Figure 4. Prototype of UYI. Images are taken from Hayashi et al., 2008.

WYSWYE. The interface for WYSWYE showed a 5x5 grid of images on the right side of the screen. The grid of images was 715x549 pixels. Each image was 139x103 pixels. A blank 4x4 grid was on the left side (see figure 5). The blank grid was 578 by 459 pixels. The blank cells were 139x103 pixels. A passcode consisted of four system-assigned images. Participants had to perform mental operations before logging in. First, they had to mentally delete a row and column that did not contain a target on the 5x5 grid. Then, they would mentally shift the remaining cells together. They log in by clicking the locations of their four targets on the blank grid. The researcher provided verbal feedback of *correct* or *incorrect* after each authentication attempt. The images were repositioned for every attempt. This graphical scheme that was created represents a collection of implementations (Bianchi, Oakley, & Kim, 2016; Brostoff, Inglesant, & Sasse, 2010; De Luca et al., 2010; Gao, Liu, Dai, Wang, & Chang, 2009; Gupta, Sahni, Sabbu, Varma, & Gangashetty, 2012; Kawagoe, Sakaguchi, Sakon, & Huang, 2012; Kim, Dunphy, Briggs, Hook, Nicholson, Nicholson, & Olivier, 2010; Lashkari, Manaf, & Masrom, 2011; Perkovic, Cagalj, & Rakic, 2009; Zangooei, Mansoori, & Welch, 2012).

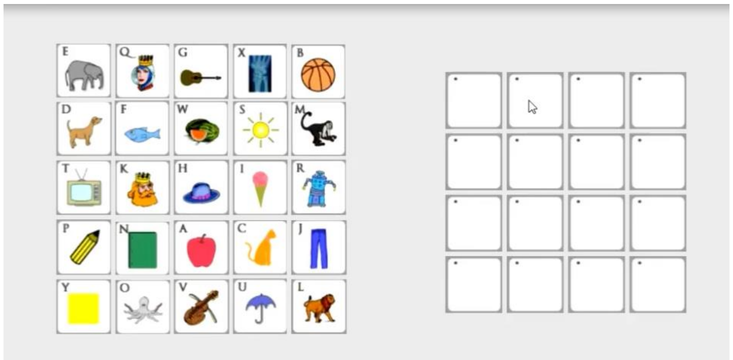

Figure 5. Prototype of WYSWYE. Images are taken from Khot et al., 2012.

Alphanumeric. The alphanumeric interface consisted of a box for text entry and an enter button. It was implemented with HTML and run in Firefox. Java code captured login times. It displayed them below the primary interface frame, and the researcher recorded them. The passcode used by all participants was “col2Wlan6.” This passcode met the rules for strong passwords including that it should not contain common words and it should contain numbers (Barton & Barton, 1984; Choong & Greene, 2016). The passcode was system-assigned because this ensured that participants did not reuse passwords from their other accounts, did not create passwords that contained personal information, and did not use dictionary words, which would result in weak passwords (Cazier & Medlin, 2006; Grawemeyer & Johnson, 2011, Zviran & Haga, 1999). Having the alphanumeric password being system-assigned also avoided giving it an unfair advantage over the graphical passwords, which were system-assigned and for which participants could not select images that were meaningful to them. Comparisons of the system-assigned alphanumeric password with new password schemes have been exemplified in previous literature (Zviran & Haga, 1990).

Procedure. Participants were run individually. They were seated in front of a desktop computer. The researcher explained that they would be authenticating using five different schemes, and participants signed a consent form after being allowed time to ask questions. The schemes were: gaze, CHC, UYI, WYSWYE, and alphanumeric. The order of the schemes was counterbalanced using a Latin Square design across participants. For each scheme, participants were given instructions, and they had one practice trial. After instructions, the experimenter would answer questions at any time. The experimenter would tell the participants whether they had correctly authenticated on every trial to provide them with feedback. Participants completed nine experimental trials of CHC, WYSWYE, and the alphanumeric scheme. Because three challenges are necessary to log in with UYI, participants logged in three times with UYI. Participants logged in four times with the gaze-based scheme.

After each set of experimental trials, participants took on the role of a casual attacker performing an OSA for each of the graphical schemes and the gaze-based scheme. They viewed a video of the researcher logging in one time. The video only showed the screen, including mouse movements as it appeared while the researcher logged in. Then they circled the passcode they thought they observed on an answer sheet. They viewed the same passcodes being entered two more times, and they made another attempt to identify the passcode on the answer sheet.

Results

Dependent variables were: error rates, learnability, and OSA performance. Bonferroni corrections were used for all post hoc comparisons.

Error rates. Error rates were calculated as the number of incorrect trials of the total experimental trials for each scheme. A repeated measures ANOVA (authentication scheme: CHC, UYI, WYSWYE, gaze-based, and alphanumeric) was conducted to explore error rates. Sphericity was violated, so a Greenhouse-Geisser correction was applied. Differences were found among error rates (see figure 6), $F(2.49, 37.29) = 9.02$, $p < .001$, partial $\eta^2 = .376$. Post hoc comparisons revealed that CHC ($M = .18$, $SD = .16$) was entered with fewer errors than the gaze-based scheme ($M = .42$, $SD = .20$), and alphanumeric passcodes ($M = .01$, $SD = .04$) were entered with fewer errors than passcodes for UYI ($M = .34$, $SD = .40$), WYSWYE ($M = .28$, $SD = .23$), and gaze-based, $p < .05$ for all comparisons. No error rate differences were found between gaze-based scheme and UYI, $p = 1.00$, or WYSWYE, $p = .55$, and no differences were found between CHC and UYI, $p = .32$, WYSWYE, $p = 1.00$, or alphanumeric passcodes, $p = .29$. No differences were found between UYI and WYSWYE, $p > .99$. Participants made more errors using the UYI, WYSWYE, and the gaze-based schemes compared to the alphanumeric scheme. Of the graphical schemes, CHC was the only one that did not have more errors than the alphanumeric scheme.

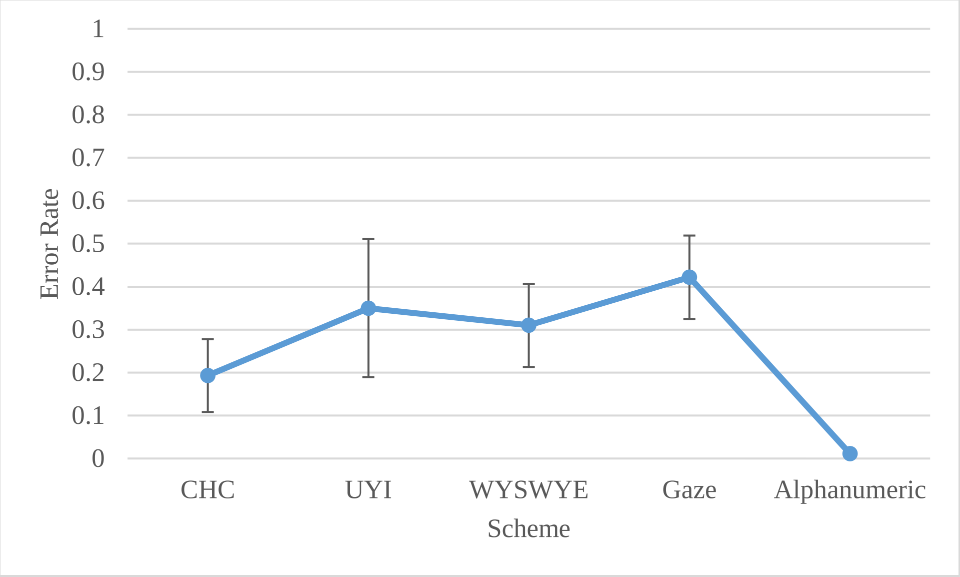

Figure 6. Error rates. Error bars represent 95% confidence intervals.

Login Times. Login times were calculated for correct, experimental trials. Login times were cleaned for CHC, UYI, WYSWYE, and alphanumeric schemes by removing outliers that were 2.5 standard deviations above or below an individual participant's mean for that scheme. No outliers were removed for CHC or UYI. Two outliers were removed for WYEWYE and three for alphanumeric. No outliers were removed for the gaze-based scheme. A repeated-measures ANOVA (authentication scheme: CHC, UYI, WYSWYE, gaze-based, and alphanumeric) was conducted to explore login times. Sphericity was violated, and a Greenhouse-Geisser correction was used. Differences were found among login times (see figure 7), *F*(2.21, 70.59) = 12.34, *p* < .001, partial η^2 = .278. Post hoc comparisons revealed that UYI (*M* = 20.22, *SD* = 10.73) had longer login times than CHC (*M* = 10.97, *SD* = 5.74), gaze-based (*M* = 11.98, *SD* = 6.87), and alphanumeric schemes (*M* = 6.81, *SD* = 3.24), and WYSWYE (*M* = 15.84, *SD*= 11.76) had longer login times than the alphanumeric scheme, *p* < .05 for all comparisons. No differences were found between login times for the gaze-based scheme and CHC, *p* > .99, WYSWYE, *p* = .62, or alphanumeric passcodes, *p* = .13. No differences were found between login times for CHC and WYSWYE, *p* = .19, or the alphanumeric schemes, *p* = .44. No differences were found between UYI and WYSWYE, *p* = .34. Alphanumeric passcodes were entered efficiently. However, the gaze-based scheme and CHC also had acceptable login times.

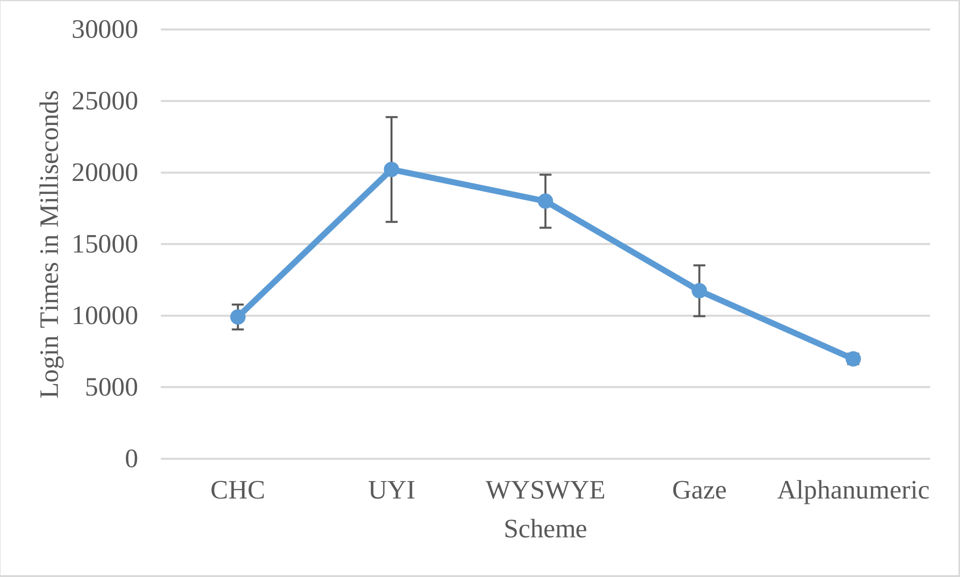

Figure 7. Login times. Error bars represent 95% confidence intervals.

The researchers also measured login times by considering that when a participant does not log in correctly, this would be added to their final login time for when they do login correctly. These login times were cleaned for CHC, UYI, WYSWYE, and alphanumeric schemes by removing outliers that were 2.5 standard deviations above or below an individual participant's mean for that scheme. No outliers were removed for CHC, UYI, WYEWYE, or the gaze-based scheme. One was removed for alphanumeric. A repeated-measures ANOVA (authentication scheme: CHC, UYI, WYSWYE, gaze-based, and alphanumeric) was conducted to explore these login times. Sphericity was violated, and a Greenhouse-Geisser correction was used. Differences were not found among login times when they were calculated in this way (see figure 8), $F(1.05, 19.98) = 3.19$, $p < .088$, partial $\eta^2 = .144$.

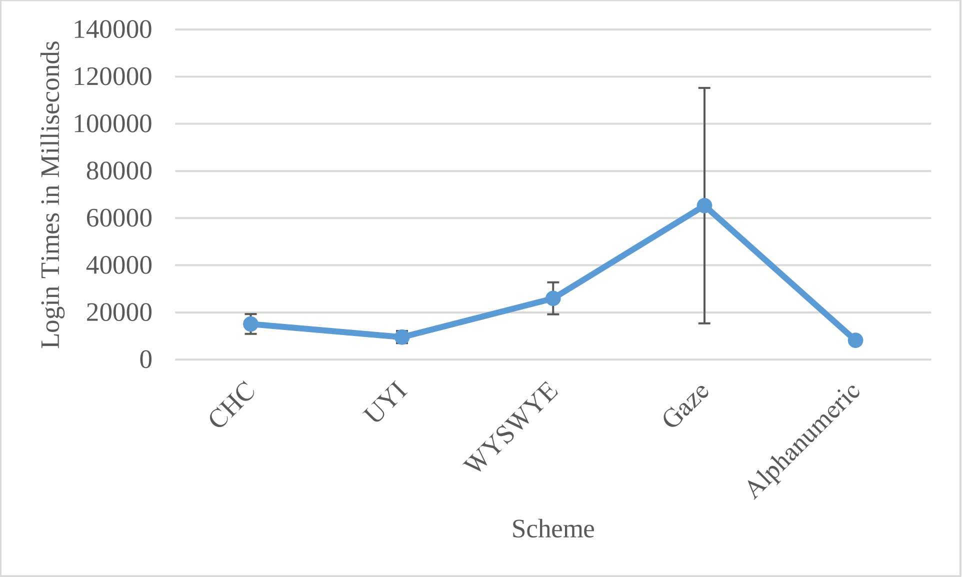

Figure 8. Login times calculated by summing incorrect attempts with next correct attempt. Error bars represent 95% confidence intervals.

Learnability. The correct attempts over time were measured to reflect learnability. The first three trials for each scheme including the practice trial composed the first rate for learnability, the second set of three interactions composed the second, and the third set of three interactions composed the third. The gaze-based scheme was not coded for learnability, because it had fewer trails than the other schemes. In order to include UYI in learnability and allow for equivalent practice among schemes, the challenges were considered individually instead of in sets of three. A 3 (number of attempts: first set, second set, and third set) x 4 (authentication scheme: CHC, UYI, WYSWYE, and alphanumeric) ANOVA was conducted to explore learnability. Sphericity was violated, so a Greenhouse-Geisser correction was employed. Main effects revealed differences among correct logins for the schemes and differences among correct logins over time, $p < .001$. The main effects were qualified by a two-way interaction between learnability and scheme, $F(3.65, 69.41) = 2.61$, $p = .021$, partial $\eta^2 = .121$. When using UYI, participant performance with UYI improved with practice, whereas it did not with the other schemes (see figure 9). Because there was an interaction, the researchers tested for simple effects. Post hoc comparisons revealed differences between the first set and second set and between the first set and third set, $p < .05$ for both comparisons. No differences were found between the second and third set of attempts, $p = .09$.

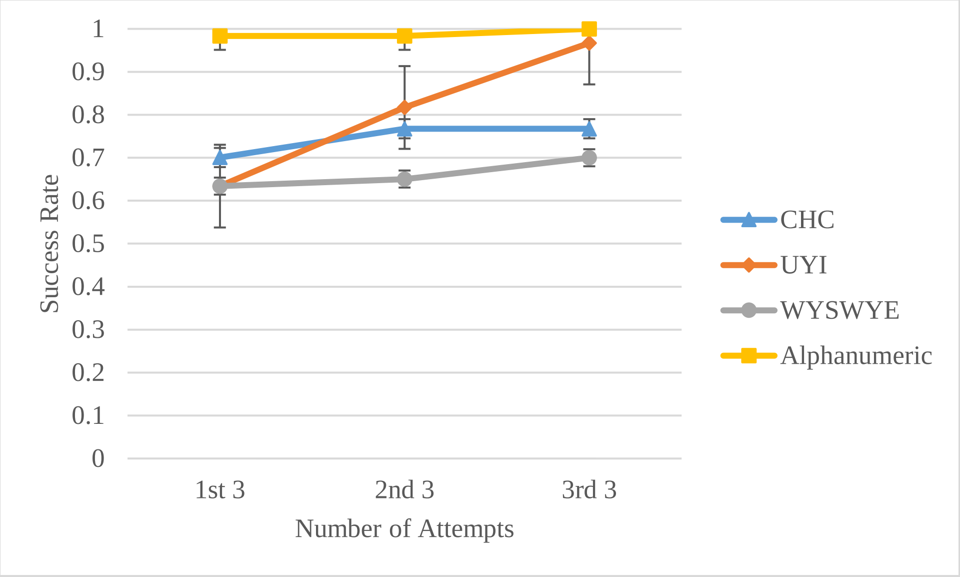

Figure 9. Learnability. Error bars represent 95% confidence intervals.

OSA performance. OSA were calculated for the first and second attempts to identify passcodes. The researchers calculated what percent of the passcode was identified for each attempt. For example, if one out of three images was identified for UYI, the OSA performance would be .33. A 2 (OSAs: one viewing and three viewings) x 4 (authentication scheme: CHC, UYI, WYSWYE, and gaze-based) ANOVA was conducted to explore OSA performance. The alphanumeric password was not tested for OSA resistance, because the password was made clearly visible in our prototype. Main effects revealed differences among OSA performances for the schemes and differences among OSA performances for the number of viewings, $p < .001$. The main effects were qualified by a two-way interaction between OSAs and scheme, $F(3, 51) = 10.38$, $p < .001$, partial $\eta^2 = .379$. UYI became vulnerable after three viewings while the other schemes did not become more vulnerable with additional viewings (see figure 10). Because of our research questions, simple effects were explored. Post hoc comparisons revealed differences between partial passcodes identified for the schemes after viewing a log in one and three times, *p* < .001, such that viewing videos additional times improved attack performance. Differences were found between attack performances on CHC (one observation: *M* = .06, *SD* = .13; three observations: *M* = .09, *SD* = .15) and UYI (one observation: *M* = .33, *SD* = .23; three observations: *M* =.80, *SD* = .20), CHC and WYSWYE (one observation: *M* = .32, *SD* = .22; three observations: *M* = .36, *SD* = .21), UYI and WYSWYE, UYI and the gaze-based scheme (one observation: *M* = .11, *SD* = .32; three observations: *M* = .17, *SD* = .38), and WYSWYE and the gaze-based scheme, *p* < .05 for all comparisons. No difference was found between attack performances for CHC and the gaze-based scheme, *p* = .001. All the schemes offered resistance to OSA. UYI was most vulnerable to attack followed by WYSWYE.

None of the participants were able to identify a full passcode after viewing a log in one time. However, participants identified partial passcodes; this was more likely for UYI and WYEWYE. Seventeen participants could not identify any correct icons for CHC on the first viewing, and three participants identified one correct icon. Three participants could not identify any correct images for WYEWYE, ten participants identified one image, five identified two, and two identified three. Four participants identified none for UYI, 12 identified one image, and four identified two. Two participants identified the gaze-based pattern by observing where the mouse entered and left the interface, which was a problem with our implementation rather than the scheme.

Given three viewings, no participant identified all of the targets for CHC or WYSWYE. Fifteen participants identified no correct icons for CHC, and five identified one. Two participants identified no correct images for WYSWYE, nine identified one, seven identified two, and two identified three. UYI was the most vulnerable after three viewings. Nine out of 20 participants identified the full passcode for UYI. One participant identified one image, and ten identified two.

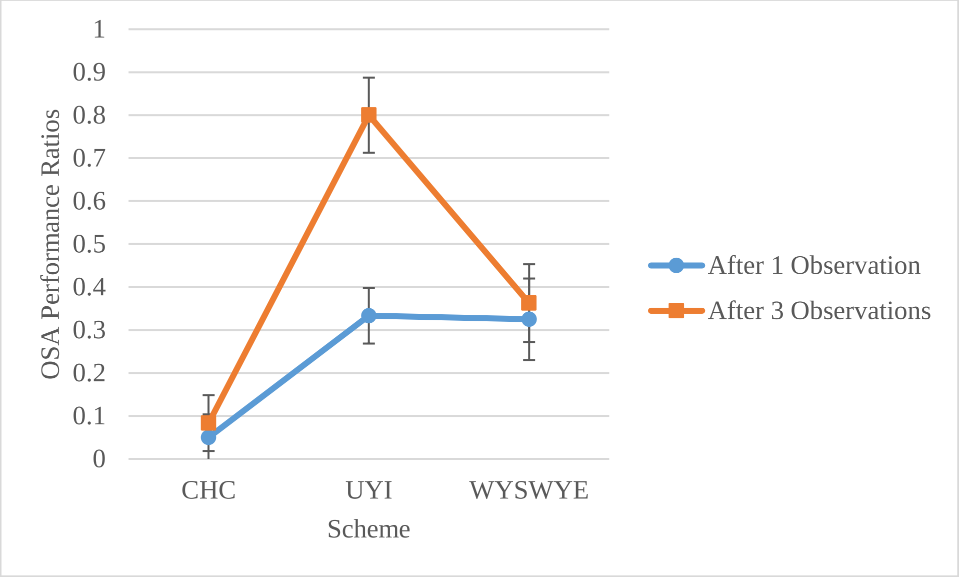

Figure 10. OSA performance. Error bars represent 95% confidence intervals.

Acceptability. Participants were asked whether they would accept added effort the OSA prevention requires for each scheme. 80% of participants accepted CHC, 45% accepted UYI, 50% accepted WYSWYE, and 68.75% accepted the gaze-based scheme.

STUDY 1: DISCUSSION

Alphanumeric passcodes had fast login times, which was expected. Login times were also appropriate for CHC and the gaze-based scheme. These schemes meet the usability requirement of providing quick access. When users are focused on their primary task of interacting with data on a device, these schemes with appropriate login times will not preoccupy users with the secondary task of authenticating. CHC may allow for different numbers of challenges, e.g., participants may click to authenticate once or on multiple subsequent grids of icons. Login times were low for CHC because participants completed one challenge to authenticate. The previous assessment of CHC had shown much slower login times of 71.66 seconds for multiple challenges (Wiedenbeck et al., 2006). Fast login times for CHC were consistent with Behl and colleagues' (2014) five second login times for their grouping scheme and were faster than Sreelatha and colleagues (2011) login times of 29.95 seconds. The gaze-based scheme also had appropriate login times because the technology enforces selections at a certain pace. Appropriate login times for the gaze-based scheme were consistent with previous literature (De Luca et al., 2009). WYSWYE and UYI had longer login times. UYI required three image selections on subsequent grids to authenticate, and WHSWYE required some mental transformations. Findings of long login times for WYSWYE and UYI were consistent with previous research (Hayashi et al., 2008; Khot et al., 2012). Shorter login times have been found for other schemes that use the same general strategies (Cain & Still, 2016; De Luca et al., 2010; Zangooei et al., 2012).

There was only a slight, non-significant improvement in errors for CHC and WYSWYE during study 1. However, learning was demonstrated for UYI. Once participants figured out what the degraded versions looked like through trial and error and feedback from the experimenter, participants improved for UYI. Learnability led to the low overall error rates for UYI. The quality of learnability that was present for UYI but not the other novel schemes is a shortcoming that needs to be addressed by changes in design rather than requiring training.

Because OSAs were measured using multiple metrics, e.g., one and three viewing and number of passcode items identified, the findings can more easily be situated with existing literature (Bošnjak & Brumen, 2018). The graphical approaches were found to be resistant to

OSA. No graphical passcode was stolen after one viewing. CHC, WHSWYE, and the gaze-based scheme continued to offer resistance after three viewings, after which no full passcode was stolen. However, UYI became vulnerable after three viewings. Just as participants could learn to identify degraded versions of targets and demonstrated learnability throughout the trials for UYI, the attackers were also able to learn the identity of degraded targets during additional viewings. UYI's vulnerability to OSA likely applies to other graphical schemes that involve the direct selection of static images. Findings that schemes that translate to another location are resistant to OSA were consistent with previous literature (Liu et al., 2011b; Sun et al., 2016), and previous literature has shown inconsistencies in the security toward OSAs provided by disguising targets (Zakaria et al., 2011).

There were high rates of acceptance for CHC and gaze-based schemes. There was lower acceptance for WYSWYE and UYI. Lower rates of acceptance for WYSWYE aligned with previous literature (Khot et al., 2012). Participants may have found it difficult to transform the images in WYSWYE and may have been dissatisfied by high error rates.

Error rates were quite high for the graphical passcodes but not for the alphanumeric scheme. The high error rates found in this within-subject runoff were consistent with Behl and colleagues' (2014) finding of 20% error rates for their grouping scheme, but they were higher than some previous assessments of graphical schemes using this strategy (Wiedenbeck et al., 2006). Error rates were also higher for the scheme that translated targets to another location compared to previous studies that showed successful logins (Gupta et al., 2012; Khot et al., 2012). High error rates likely came from a lack of familiarity and the additional cognitive effort often required with graphical approaches.

CHAPTER 3: STUDY 2: MEMORABILITY RUNOFF

Study 1 compared four OSA resistant graphical schemes and the alphanumeric scheme on dimensions of usability and security. Study 2 added the dimension of memorability by testing error rates and verbal memory for passcodes following a three week delay.

Method

Participants. Twenty undergraduate students participated in part one, and 18 returned for part two (females = 11). They were recruited through the SONA system and compensated with class research credit. Two participants reported being left hand dominant. Ages ranged for 18 to 42 (*M* = 21.67, *SD* = 5.60). Reported computer use ranged from 2.5 to 18 hours a day (*M* = 8.69, *SD* = 4.17). All participants reported normal or corrected to normal vision.

Stimuli, apparatus, and measures. Stimuli consisted of the same five authentication scheme prototypes that were used in Study 1. Study 2 included the System Usability Scale (SUS; Brooke, 1996) as a measure of satisfaction. This scale consists of ten items that participants rate on a five-point Likert scale.

Procedure. Participants were run individually. During part one of the study, participants were seated in front of a desktop computer, and they signed a consent form. The experimenter explained that they would be logging in to five authentication interfaces and that in three weeks they would be doing the same thing. They were told that when they come in for part two, they would use the same passcodes as during part one, but they would not be reminded what the passcodes are. The order in which they logged into each interface was counterbalanced using a Latin Square design. For each interface, participants were shown their passcode and asked to memorize it. Participants were given instructions and one practice trial. Throughout the trials, the experimenter would answer any questions asked but would only volunteer feedback about whether participants had correctly authenticated. Participants logged in using CHC, WYSWYE, and alphanumeric passcodes ten times. They logged in using UYI three times, with each login consisting of three identifications of targets. They logged in using the gaze-based scheme five times. Three weeks later, participants returned and followed the same procedure without being shown their passcodes. A three-week delay was used to represent infrequently used passwords that would be relatively difficult to remember.

Results

Dependent variables in study 2 were error rates and the percent of each passcode remembered.

Memorability. Error rates were calculated for time 1 and time 2 as the number of incorrect trials of the experimental trials. The researchers calculated error rates at time 1 and time 2 for CHC, UYI, WYSWYE, and the alphanumeric scheme. The gaze-based scheme was not included in this analysis because of missing data. A 2 (elapsed time: day one or three weeks later) x 4 (authentication scheme: CHC, UYI, WYSWYE, and alphanumeric) ANOVA was conducted to explore memorability as measured by error rates (see figure 11). Sphericity was violated, so a Greenhouse-Geisser correction was employed. Main effects revealed differences among error rates for the schemes and differences among error rates for elapsed time, $p < .001$. The main effects were qualified by a two-way interaction between elapsed time and scheme, $F(1.99, 27.88) = 35.97$, $p < .001$, partial $\eta^2 = .072$. Error rates for graphical passcodes did not differ by elapsed time. However, error rates for the alphanumeric passcode did (see figure 11). Simple effects were investigated. A Bonferroni correction was used for post hoc comparisons. Post hoc comparisons revealed that time 2 had more errors than time 1, $p < .001$. The alphanumeric scheme (time 1: $M = .05$, $SD = .20$; time 2: $M = 1.00$, SD = .00) had more errors than CHC (time 1: *M* = .18, *SD* = .12; time 2: *M* = .29, *SD* = .21), UYI (time 1: *M* = .23, *SD* = .21; time 2: *M* = .24, *SD* = .26), and WYSWYE (time 1: *M* = .30, *SD* = .19; time 2: *M* = .25, *SD* = .25), *p* < .05 for all comparisons. There were no differences found between CHC and UYI, *p* = 1.00, or WYWYE, *p* = 1.00. There were no differences found between WYSWYE and UYI, *p* = 1.00.

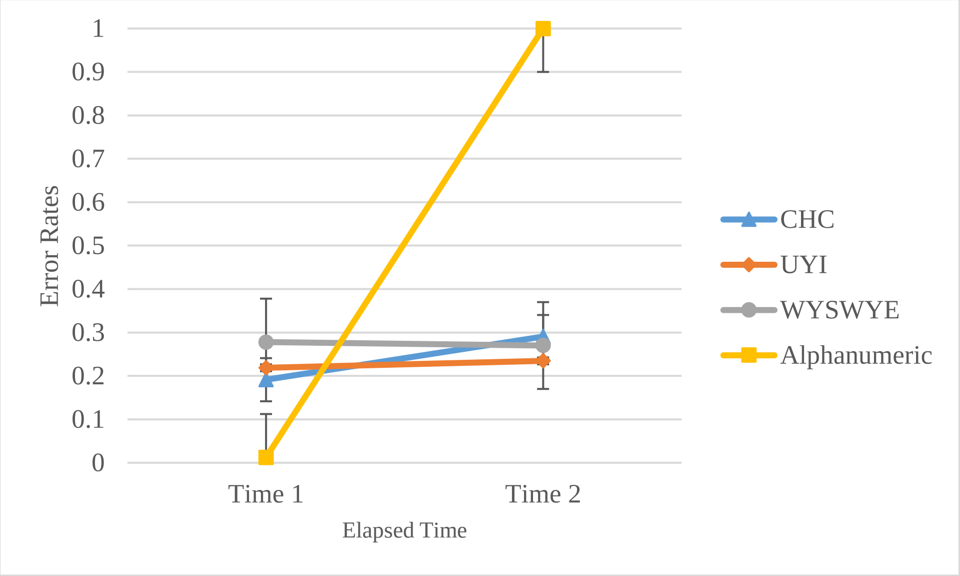

Figure 11. Memorability as reflected by login error rates. Error bars represent 95% confidence intervals.

For CHC, participants had three targets to remember after a three-week delay. Seven percent of participants remembered two icons when asked to vocalize their targets (see figure 12). Ninety-three percent remembered all three. UYI also consisted of three targets. 14% remembered one, 29% remembered two, and 40% remembered all three. WYSWYE consisted of four targets. Six percent of the participants remembered two, and 57% remembered all four. 82% of the participants remembered the gaze-based passcode. No participants remembered the alphanumeric password after a three-week delay. All other passcodes had fewer errors during verbal report than when participants entered them using the interfaces (see figure 13).

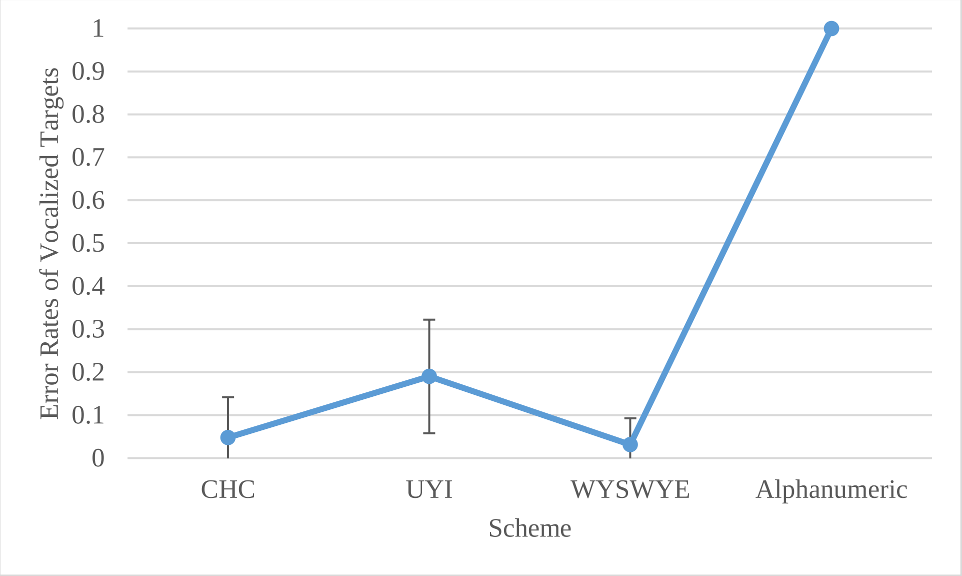

Figure 12. Memorability as reflected by login error rates. Error bars represent 95% confidence intervals.

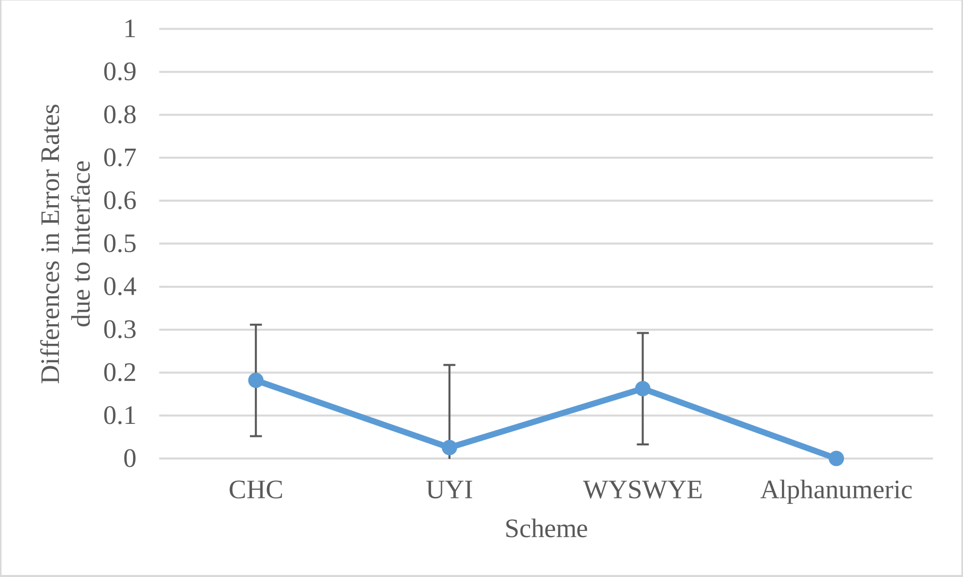

Figure 13. Increase in errors due to interacting with the interface. Error bars represent 95% confidence intervals.

Satisfaction. Satisfaction was measured by the SUS (Brooke, 1996) after the use of each scheme on time 1 and time 2. The SUS for the alphanumeric scheme was not measured at time two because no passcodes were successfully entered. Due to technical problems with the eye-tracker and missing data, the gaze-based scheme was also not included in this analysis. A 2 (elapsed time: day one and three weeks) x 3 (authentication scheme; CHC, UYI, WYSWYE) ANOVA was conducted to explore satisfaction (see figure 14). Sphericity was violated, and a Greenhouse-Geisser correction was used. There was no interaction between elapsed time and scheme, *F*(1.25, 18.80) = 0.09, *p* = .820, partial η^2 = .006, which indicated that satisfaction at time two did not depend on the scheme. Main effects revealed differences in satisfaction among schemes, *F*(1.85, 18.80) = 10.88, *p* < .001, partial η^2 = .420. There was no main effect for elapsed time, *p* = .217. Participants were most satisfied with CHC (time 1: *M* = 73.25, *SD* = 22.58; time

2: *M* = 81.88, *SD* = 11.35), followed by UYI (time 1: *M* = 64.53, *SD* = 16.34; time 2: *M* = 68.59, *SD* = 18.37), and then WYSWYE (time 1: *M* = 54.84, *SD* = 25.07; time 2: *M* = 61.78, *SD* = 20.35).

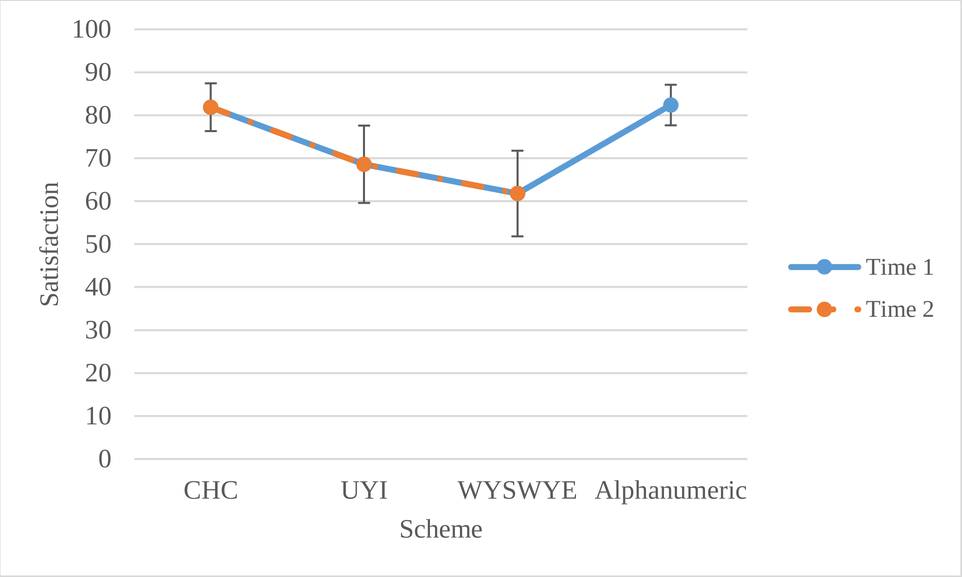

Figure 14. Satisfaction. Error bars represent 95% confidence intervals.

STUDY 2: DISCUSSION

Evidence was provided for the memorability of all four prototypes of graphical and gaze-based passcodes. Participants had similar rates of error on day one and three weeks later for these schemes. For CHC, WYSWYE, and the gaze-based scheme, most participants verbally remembered the whole passcode. Forty percent remembered the whole passcode for UYI. The memorability of the graphical schemes was impressive when compared with the alphanumeric scheme, for which no participant could enter the passcode correctly or verbally remember it three weeks later. Remembering the alphanumeric passcode may have been challenging because it was long and complex to be secure and because it was system-assigned. However, the graphical passcodes and gazed-based scheme were also system-assigned. The memorability for all of the schemes would likely have been higher if they had been user-chosen passcodes. However, user-chosen passcodes can compromise security because users have biases in their selections. For example, Pering, Sundar, Light, and Want (2003) allowed users to log in using photographs from their smartphone photo galleries. Participants who took on the role of the attacker were able to identify 100% of one of the user's passcodes. Impressively, the graphical passcodes were easily remembered, despite being system-assigned, likely due to the picture superiority effect. Muscle memory could have also aided the memorability of the gaze-based scheme. Memorability for UYI likely benefited from cued-recall. Being able to view the targets leverages cognitive abilities for memory (Al Ameen, 2016). Cued-recall would have aided memory for CHC and WYSWYE to a lesser extent than UYI because the targets were among many distractors. Findings that the grouping scheme and the scheme for translating to another location were memorable was consistent with previous literature (Brostoff et al., 2010, Wiedenbeck et al., 2006), and memorability for disguising targets and the gaze-based scheme was higher than in previous studies in which a drop in success rates were observed after three weeks (Hayashi et al., 2008) and after ten days (De Luca et al., 2009). In addition, the satisfaction measure showed that participants were more satisfied with CHC and the gaze-based scheme than UYI and WYSWYE.

GENERAL DISCUSSION

The current research provided a direct comparison of prototypical examples of graphical passcodes and a gaze-based scheme that were designed to thwart OSAs. These schemes were classified as providing resistance by grouping targets among distractors, translating targets to another location, disguising targets, and using gaze-input, and they were compared to the traditional alphanumeric scheme. The relative strengths of the authentication schemes in terms of memorability, quick access, learnability and successful entry, security, and satisfaction were determined.

All four prototypes of graphical and gaze-based passcodes were found to be memorable as demonstrated by similar error rates on day one and three weeks later. Memorability for these schemes contrasted with the alphanumeric scheme, for which no participant could enter the passcode correctly or verbally report it three weeks later.

The graphical approaches were resistant to OSA. No graphical passcode was stolen after one viewing. Partial passcodes were stolen, especially for UYI and WYSWYE, but a casual attacker cannot complete an attack with a partial passcode.

Studies 1 and 2 demonstrated some of the strengths of graphical passwords, but they also found limitations that need to be addressed by additional research. Although the graphical schemes offered memorability and resistance to OSA, error rates were high for these schemes compared to the familiar alphanumeric scheme. There is a need for strategies that can improve error rates for novel graphical schemes.

CHAPTER 4: EXAMINING THE LIMITS OF MEMORABILITY AND LEVERAGING CONTEXT

Subsequently, two studies investigated a novel intervention to improve encoding and retrieval of graphical passcodes by providing context. Previous studies on context-dependent memory have provided evidence that cues from the environment that are present during encoding can help with future retrieval (Godden & Baddeley, 1975; Smith, 1986). For example, in this classic study, participants learned a list of words while they were either underwater or on land, and they then recalled the list of words either underwater or on land. Retrieval was facilitated by having the same context for learning and recall (Godden & Baddeley, 1975). This basic finding of context-dependent memory has been demonstrated with a variety of contextual cues and testing conditions. It was also found for the water manipulation when participants were learning decompression tables, a tool used by divers (Martin & Aggleton, 1993).

Location serves as a strong contextual cue. Context-dependent memory has been demonstrated in a variety of situations. When participants were asked to perform eyewitness testimony, they were better at identifying a confederate when they returned to a fake crime scene (Smith & Vela, 1992). Smith (1986) also found context-dependent memory when participants better-recalled words learned either in a classroom or a cubicle when recall was in the same space as learning. The researchers found the same result when participants performed a forced-choice recognition task rather than recall, showing that the effect generalizes across memory tasks. Musicians learned music and were tested on recall of the piece (Mishra & Backlin, 2007). When context was provided by an atypical context (i.e., a lobby or a conference room) and when context was provided by an upright or grand piano, recall improved by consistencies between context during learning and testing. When learning an action sequence either verbally or through performing it in the context of either a basement or outdoors, recall was better when participants were in the same context as when they were learning (Sahakyan, 2010). Cats learned associations between tones or lights and shocks (Wickens, Tuber, & Wickens, 1983). They learned and were tested in either of two rooms with different décor. The learned associations were stronger when tested in the same context.

Sound can also serve as a contextual cue. Grant et al. (1998) had students learn material in either a noisy or silent room. Recall on short answer questions, and recognition on multiple choice questions were facilitated when the noise level at testing matched that at learning. When context was provided by ambient noise for recall and recognition of virtual objects in virtual environments, participants performed best on high fidelity ambient noise than low fidelity noise, and they performed worse on no noise (Davis, Scott, Pair, Hodges, & Oliverio, 1999). Participants learned information on a desktop computer (Stefanucci, O'Hargan, & Proffitt, 2007). Behind the computer, there was either nothing or there was a large screen depicting a scene accompanied by ambient noises. Recall of the information was better in the context condition. Context facilitating memory was demonstrated when the context is provided by a song (Balch, Bowman, & Mohler, 1992; Smith, 1985). After hearing a song during learning, recall was not facilitated when there was no song or when the tempo of the song was changed.

Bodily movements or kinesthetics can serve as a contextual cue. When recalling a list of words, chewing gum was found to act as a contextual cue facilitating later retrieval (Baker, Bezance, Zellaby, & Aggleton, 2004). Posture (i.e., sitting up or lying down) has been found to facilitate recall of nonsense syllables (Rand & Wapner, 1967). Exercise or rest facilitated retrieval of a list of words (Miles & Hardman, 1998).

When learning and recalling a passage, odor has even been found to support context-dependent memory (Pointer & Bond, 1998). Odor also facilitated a context effect for lists of words (Schab, 1990). The odor was particularly effective when it was novel or contextually inappropriate (Herz, 1997).

Context helps with memory for faces and words. Recognition memory is facilitated when faces are learned and context is provided by another face accompanying it (Watkins, Ho, & Tulving, 1976; Winograd & Rivers-Bulkeley, 1977), and recognition memory is facilitated when faces are learned and context is provided by a consistent description of the face (Baddeley, 1982). Context facilitating memory was demonstrated when participants learned a list of nonsense syllables by writing it with either their left or right hand (Nagge, 1935). After 24 hours they learned the list again. The dependent variable was how many times through the list it took to relearn it, which is called savings. There were more savings when relearning with the same hand as the original learning. Font facilitated a context effect for lists of words and nonsense words (Kirsner, 1973). Context facilitating memory was found when participants learned a list of words presented in one of two voices and then recognized the words among new words presented by either the original voice or the other voice (Craik & Kirsner, 1974; Palmeri, Goldinger, & Pisoni, 1993; Sheffert & Fowler, 1995). Not only did this context intervention increase accuracy, but it also increased the speed of responses (Palmeri et al., 1993). Context-dependent memory was demonstrated when participants learned two lists of words (Dallett & Wilcox, 1968). They were either both learned in lab space, both learned with participants’ heads in a box, or one list was learned in one context and the other in the other context. When participants were asked to recall the second list, it had fewer intrusions from the first list when it was learned in a different context, suggesting the participants had encoded the stimuli in association with its context. When participants learned a list of pairs of words and were asked to recognize them among new pairs, they performed better when the list was presented in the same screen location, foreground color, and background color during testing as during learning (Murnane & Phelps, 1993). When participants learned lists of words on one of two fonts or from one of two male voices, they performed better on recognition when the cue was the same during testing as during learning (Naveh-Benjamin & Craik, 1995). When participants heard a list of words in either a male or female voice and were tested on recognition, they performed better when they heard the list in the same voice during testing as during learning (Geiselman & Glenny, 1977). List learning for words was also improved when the context was provided by normal vision or vision restricted by goggles (Dolinsky, & Zabrucky, 1983). Context facilitating memory was demonstrated when participants learned nouns paired with adjectives (Light & Carter-Sobell, 1970). The adjectives either gave the same semantic meaning or a different meaning (e.g., official seal or performing seal). Participants had better recognition memorably for nouns paired with an adjective giving the same semantic meaning during testing as during learning. Cue words present during learning were also found to facilitate recall when they were weakly associated with the target words (Tulving & Osler, 1968). Also, when words on a list were learned in pairs, participants were more likely to remember each word if they also remembered its pair, and when digits were learned in pairs, participants were more likely to remember each digit if they remembered its pair (Horowitz & Prytulak, 1969).

Further, simply visualizing the study room can improve recall even if testing occurs in a different physical context (Smith, 1979; Smith, 1984). Context facilitating memory also occurred when participants were asked to visualize pairs of words interacting with a location but was not observed when participants visualized a location not interacting with the words (Winograd &

Lynn, 1979). When asked to visualize objects either in a separate frame or being on top of or concealed by each other, context improved recall when the objects were visualized as interacting (Neisser & Kerr, 1973). When asked to visualize nouns either in the same or separate scenes, participants did better on recall and recognition when they imagined them in the same scene (Begg, 1973; Begg, 1978). The same result was found for learning pairs of words (Robbins, Bray, Irvin, & Wise, 1974). Participants learned word triplets (Petersen, 1974). They were asked to either visualize them in a given location or not. During testing, they were either cued with the location or with one of the words. They performed best on recall when cued by the location. The more integrated a target is with a cue, the better the cue aids memory search.

Many previous studies have provided support for context facilitating memory. However, some have not (Godden & Baddeley, 1980; Johnson & Miles, 2008; Koens, Ten Cate, & Custers, 2003; Smith, Glenberg, & Bjork, 1978; Strand, 1970). Strand did not find context-dependent learning when participants were asked to leave the room they learned the words in before returning for recall. It was not found for test scores of college students using the contexts of a lecture hall or another room (Saufley, Otaka, & Bavaresco, 1985). Johnson and Miles did not find context-dependent memory when the context was provided by the flavor of chewing gum. Mishra and Backlin (2007) did not find it for recall of music when context was provided by typical locations (i.e., a practice room, a professor’s studio, and a stage) instead of an atypical location. Petrich and Chiesi (1976) did not find that context provided by background colors or colors of text facilitated learning lists of words. When participants learned faces with descriptions, and they recognized the faces either with the same description, a new description, or saw the same learned description with a new face, participants tended to identify faces with contexts from the studied list (Baddeley & Woodhead, 1982); participants recognized familiar contexts regardless of whether they were integrated with the correct face cue. When participants learned pairs of nouns and performed a recognition, a cued-recall, and a free recall test with the cue being provided by the first noun in each pair, there was no context effect (Fernandez & Glenberg, 1985). Even when the participants were instructed to create a sentence integrating the two nouns, there was still no effect of context. Lastly, although Emmerson (1986) found that water provided context-dependent memory for recognition memory, Godden and Baddeley did not find context-dependent memory for recognition of words learned either underwater or on land, and they did not find it for recognition of words paired with cue words. Researchers tend to agree that alignment of context between learning and testing environments facilitates retrieval (Smith & Vela, 2001). The context provides cues that help with memory search (Smith, 1994; Smith & Vela, 2001). However, there must be strong cues from the environment to establish contextual memory. It is uncertain how strong contextual cues need to be and whether this manipulation would translate to graphical authentication.

Context and elaboration have even been used to facilitate login performance by previous researchers (Bulling et al., 2012; Liu et al., 2011; Sun et al., 2016). Common graphical schemes that incorporate context include Cued-Click Points (Bulling et al., 2012), PassPoints (Wiedenbeck et al., 2005), and Background Draw a Secret (Dunphy & Yan, 2007). The Cued-Click Points and PassPoints schemes allow users to log in by clicking points on an image. Draw a Secret allows users to log in by drawing a doodle over an image. However, only two studies have investigated the effectiveness of providing context. Dunphy and Yan compared memorability for Draw a Secret with a background image and Draw a Secret with no background image in a sample of 42 participants after one week. Only one participant forgot the passcode in each condition. In a second study about the effectiveness of providing a background for Draw a

Secret, four out of seven participants remembered the Draw a Secret passcode after 15 days, and seven out of nine remembered the Background Draw a Secret passcode after the delay (Arya & Agarwal, 2011). This study provided tentative support for context in improving the retrieval of graphical passcodes.

Context is provided in CBFG (Liu, Qiu, Ma, Gao, & Ren, 2011b) and PassMatrix (Sun, Chen, Yeh, & Cheng, 2016) in the form of a background image, which is broken into numbered cells. Passcodes consist of points on the image. For CBFG, users authenticate by selecting numbers on the side of the image corresponding to the cells containing their targets, and for PassMatrix, users authenticate by using a scroll bar on the side to select cells containing their targets. Context was not manipulated for either of these schemes. Story (Davis, Monrose, & Reiter, 2004) is a graphical passcode scheme that presents images on a grid. Designers encouraged elaboration by asking participants to construct a story about the images in their passcode. The researchers found that their scheme provided for memorable passcodes. However, they did not manipulate whether the story intervention was given or not. Furthermore, 50% of participants reported that they had neglected to create a story. The “Comes from Draw a Secret” (Gao, Ren, Chang, Liu, & Aickelin, 2010) scheme instructs participants to create stories for their passcode images in which they connect key objects with a line within a grid. Passcodes for this scheme were found to be memorable 100% of the time. However, context was not manipulated as part of the experiment, and it was not verified that the participants indeed created stories. Context was provided by Li, Sun, Lian, and Giusto’s (2005) graphical scheme, although it was not manipulated as an independent variable. To log in users clicked a point on a photograph of a room, then they clicked an image on a grid, and then they selected a color on a grid. In another spatially focused approach, the Memory Palace scheme (Lu, Lee, Das, & Hong, 2016) allowed participants to log in by tracing a path through a set of 3-D rooms. This scheme provided memorable passcodes, but again context was not experimentally examined. These previous studies suggest that graphical passcodes' memorability can be boosted through context or elaboration. However, there is a need for research that determines a causal role for context in retrieving graphical passcodes.

If context can improve error rates of graphical schemes, this may not only help with improving interactions with these novel approaches; it may also help with potential limitations for memorability when users need to remember multiple graphical passcodes or longer graphical passcodes. As technology continues to advance, users are making use of more and more services. Each service requires a unique password (Barton & Barton, 1984). As a result, users have many passwords they need to remember, although most studies about graphical passcodes only have users remember one. Even studies that examine the use of multiple graphical passcodes have only required users to remember as many as four (Johnson & Werner, 2008; Schaub et al., 2013). An ecological study about alphanumeric password use found that users need to remember an average of eight passwords (Grawemeyer & Johnson, 2011). In their study, participants had trouble remembering their passwords and would frequently write them down or use common words or names. Although this research has shown that users have trouble remembering multiple alphanumeric passcodes, no study has determined the limits of memorability for graphical passcodes.

The third and fourth proposed studies filled a need by determining whether providing context in the form of a background image can facilitate retrieval of target images in a traditional grid-based graphical passcode scheme. It was hypothesized that context-dependent memory would increase the memorability of passcodes.

Furthermore, the proposed studies demonstrated the limits of memorability for graphical passcodes providing an indication of their applicability. Study 3 examined the number of passcodes that can be recalled after a 3 week delay (4, 8, or 12). Previous research has shown that humans have an almost limitless memory for pictures (Standing, 1973; Standing et al., 1970). Pictures are better remembered than words because of the picture superiority effect (Paivio, 1979), which describes a separate and stronger visual than verbal code. Bartram (1974) provided evidence for the picture superiority effect. Participants practiced naming objects. Objects were either the same each time, were viewed from a different angle each time, or were different objects with the same name. Naming latency was fastest for naming 2d and 3d objects, which relied on a visual code, and were slowest for naming different objects, which relied on a verbal code. Bruce (1982) also found evidence of a stronger, separate pictorial code. Participants performed yes-no recognition tests for pictures of unfamiliar faces. The faces were either the same as the learning phase, were rotated, had a different expression, or were rotated and had a different expression. Participants were fastest and more accurate for unchanged faces. Rotation or changed expressions were faster than rotation and changed expression. When participants learned familiar faces and recognized them from either the same view, from a changed angle, or with a changed expression, familiar faces were also recognized more slowly with changes, suggesting the pictorial code helped speed responses.

Shepard (1967) compared recognition memory for 600 words, sentences, and pictures. With no delay, participants remembered 90% of words, 88% of sentences, and 98% of words. After a week, recognition memory for pictures was the same as words and sentences without any delay. Participants were able to recognize 90% of a set of 2,560 pictures after a four-day delay, and they were able to recognize 99% of a set of 1000 pictures after a four-day delay (Standing et al., 1970). In alignment with findings of almost limitless memory for pictures in forced-choice recognition tasks (Standing et al., 1970), it was hypothesized that there would not be a decrement in memorability when participants need to remember multiple passcodes after a three week delay, as evidenced by a flat function.

CHAPTER 5: CAPACITY LIMITS FOR MULTIPLE GRAPHICAL PASSCODES

Method

Participants. Forty-two undergraduate students participated (females = 24). They were recruited through the SONA system and compensated with class research credit or were recruited with fliers and compensated with $20. Eight participants reported being left hand dominant. Ages ranged from 18 to 40 (*M* = 23.45, *SD* = 5.63). Reported computer use ranged from 3 to 18 hours a day (*M* = 8.43, *SD* = 4.44). All participants reported normal or corrected to normal vision.

Stimuli. Stimuli were presented in Paradigm©. Nine degraded images that were 112x90 pixels were displayed on a 3x3 grid. Images were degraded by removing detail but retaining general colors and shapes. This was accomplished by applying the oilify filter in Gimp© and using brushstroke 16. One image was the target on each grid. In the context condition, a scene was displayed containing the passcode targets during the first login (see figure 15). Scenes included were a porch, a park, a swimming pool, a store, a forest, a school, a mountain, an office, a living room, a river, a beach, and a movie theater. The scene was 800x532 pixels. Then on subsequent logins, the same scene was displayed without the passcode.

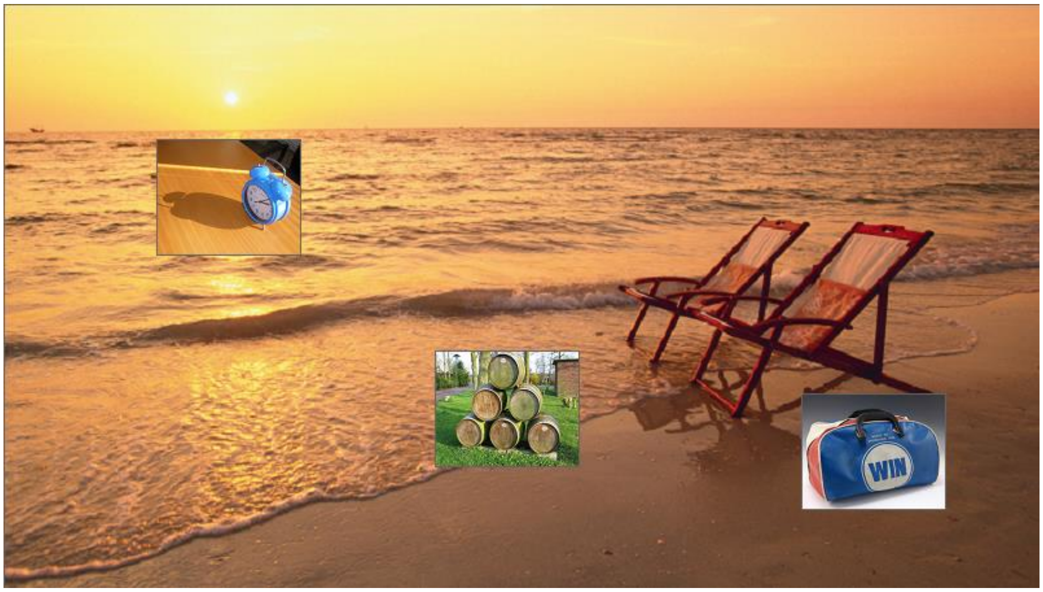

Figure 15. A passcode placed on a background scene.

Procedure. Participants sat in front of a computer and were presented with a passcode consisting of three, whole target images. Then in the context condition only, the participants saw their passcode on a background scene. For example, the three target images appeared on a beach. Then they logged in one time for practice followed by nine experimental trials. They selected the three degraded versions of their targets on three subsequent 3x3 grids. They repeated this procedure with either four, eight, or 12 passcodes. The number of passwords and the presence of context was counterbalanced between participants using a Latin Square design. For each participant, the software collected error rates. There were 12 total conditions, but each participant only experienced two conditions. There were 120 trials per condition for four passcodes, 240 for eight passcodes, and 360 for 12 passcodes. Therefore there were 240, 360, or 480 total trials per participant.

STIMULI DEVELOPMENT STUDY

In order to build the stimuli for studies 3 and 4, the researchers needed to determine the appropriate level of distortion for the images. Hayashi et al., 2008 performed a pilot study consisting of six participants to determine appropriate distortion levels when manipulating brush size in the oil painting filter. Participants viewed images at varying levels of distortion. First, they viewed the images starting with the highest level of distortion, and they were asked when they first recognize an image. Their answers to this question showed when an attacker would be able to steal a target. Second, they viewed the images starting from the least distortion, and they were asked when they could no longer recognize each image. Their answers to this question showed how well users would be able to recognize their targets. Hayashi and colleagues found that for an image that is 56x56 pixels, the appropriate level of distortion is brush size 8. At this level, users can recognize their targets and attackers cannot steal the targets. The current researchers collected a large set of images and replicated Hayashi and colleagues' pilot study with a larger participant sample.

Method

Participants. Twenty undergraduate students were recruited through SONA and compensated with class research credit. There were 16 females. Ages ranged from 18 to 30 (*M* = 20.10, *SD* = 3.3). Four participants reported left-handedness. English was reported as the first language of all participants except two who spoke Spanish before learning English. All participants had normal or corrected to normal vision.

Stimuli. Twenty images of objects were collected. The objects belonged to different classes. For example, there was only one image of a dog, rather than having images of different kinds of dogs. Each image was 112x90 pixels and was distorted using the oil painting filter in

Gimp©. Brush strokes were at levels of 10, 13, 16, 19, and 22. Brush strokes at level 16 for 112x90 pixels was equivalent to the appropriate brush stroke level of 8 for 56x56 pixels found by Hayashi et al., 2008). The images were printed and stacked in order of increasing distortion.

Procedure. Participants were first asked when they recognized images. They were shown each image at decreasing levels of distortion. They were then shown each whole image ask asked when they no longer recognize it. They were shown each image at increasing levels of distortion.

Results and Discussion

The average brush stroke at which participants could recognize an unfamiliar image was 3.46 (*SD* = 6.61). The average brush stroke at which participants could recognize a familiar image was 16.08 (*SD* = 4.47). A paired samples t-test revealed that differences between the responses to the two questions were significant, $t(399) = 45.42$, $p = <.001$, $d = 2.27$. As attackers, participants were not able to steal targets unless they were at very low levels of distortion. As users, participants could recognize images at brush stroke levels of 16. These findings confirm that the level of brush strokes used by Hayashi and colleagues (2008) is also appropriate for the images the researchers collected.

STIMULI ASSOCIATIONS STUDY

The objects and scenes were selected for studies 3 and 4 such that the objects were assumed to be unrelated to the scenes. The current study was performed to verify that the stimuli indeed were unrelated. If it is found that the stimuli are related to the images, this could present a confound because the context intervention is designed to be independent of the passcode that is assigned. If a contextual image is related to the passcode, this would provide potential attackers with clues about a passcode. To avoid this security risk, the stimuli association study was designed as a priming study that measures associations among stimuli using reaction times.

Reaction times from priming experiments can reflect how well two stimuli are related to each other (Neill, Lissner, & Beck, 1990; Tipper, MacQueen, & Brehaut, 1988). It is theorized that when two targets are associated, reaction times will be faster than for unrelated targets; underlying neural connectivity speeds reaction times (Bharucha & Stoeckig, 1986). For example, when participants heard two chords that were either related harmonically or unrelated, they identified whether the second chord was major or minor more quickly when it was related to the first chord than when it was unrelated (Bharucha & Stoeckig, 1986).

Method

Participants. Twenty undergraduate students were recruited through SONA and compensated with class research credit. There were 16 females. Ages ranged from 18 to 47 (*M* = 21.5, *SD* = 7.32). Two participants reported left-handedness. English was reported as the first language of all participants except three. All participants had normal or corrected to normal vision.

Stimuli. Twelve images of scenes and 36 images of objects, which were used in studies 3 and 4, were included. Because images from studies 3 and 4 were the experimental stimuli, ten more images of objects and scenes were added to compile the unrelated condition, and ten images of objects and scenes were added for the related condition. Objects and scenes that were related and unrelated to each other were determined using an associations database of semantic associations (Nelson, McEvoy, & Schreiber, 2004). The database consists of 5,019 words. 6,000 participants had responses with associations between the words. When they saw the first word, they responded with the first word that came to their minds that was related to it. The objects and scenes collected for this study were found on creative commons following the same procedure as in studies 3 and 4. The only difference between the stimuli for the related, unrelated, and previously used stimuli was the semantic associations (see figure 16 for examples of stimuli). An additional 30 images of objects and 60 images of scenes were added for non-critical trials. All scenes were 550x360 pixels, and all objects were 112x90 pixels.

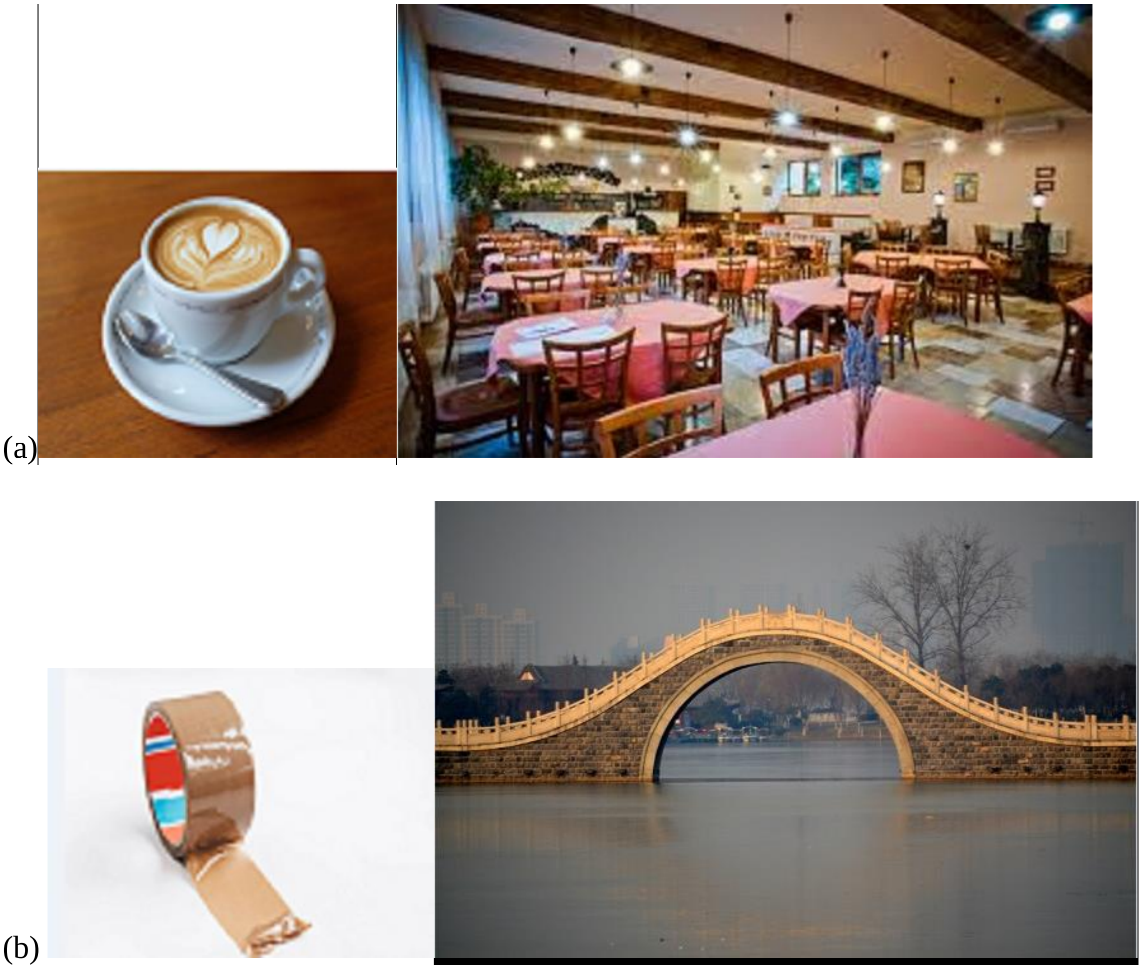

Figure 16. (a)A related object and scene. (b) An unrelated object and scene.

Procedure. Images were presented in Paradigm©. The software recorded correct responses and reaction times. The current study adopts a “same-different” methodology from

Neill and colleagues' previous priming study. In the previous study, participants saw a fixation cross, and then a string of letters briefly flashed on the screen. Participants indicated whether the first and last letters in the string were the same or different. In the current study, there were 60 total trials, 30 of which were "same," critical trials, and 30 of which were "different," non-critical trials. Participants first saw a fixation cross and hit the space bar when they were ready. Then they saw an image of a scene for 150 milliseconds, the same time used in a previous same-different, priming experiment (Tipper et al., 1988). The scene was followed by an object for 150 milliseconds, and then a scene for 150 milliseconds. Then on a blank screen, participants typed y if the scenes were the same and n if they were different. Of the 30 "same" trials, ten images were from studies 3 and 4, ten were related, and ten were unrelated. Trials were presented in a random order. Because there were more than ten object and scenes in studies 3 and 4, the images from the previous studies were randomly distributed in the ten trials.

Results

The hypothesis was investigated that the stimuli coming from studies 3 and 4 would be different from the related stimuli and that the unrelated stimuli would also be different from the related stimuli. A repeated-measures ANOVA (type of stimuli: related, unrelated, from studies 3 and 4) was run to investigate the association among stimuli and background scenes. Data were cleaned by removing outliers beyond 2.5 standard deviations from the mean at the participant level. Sphericity was violated, so a Greenhouse-Geisser correction was employed. There were no differences among the stimuli conditions, $F(1.04, 19.87) = 0.07$, $p = .385$, partial $\eta^2 = .041$, meaning the relatedness of objects to scenes was non-conclusive.

Study 3: Results

Memorability as Measured by Success Rates by Login. Success rates by login were calculated as the number of correct logins of the total logins. Each login consists of three trials. A 2 (delay: immediate or three weeks) x 2 (context: yes or no) x 3 (number of passcodes: 4, 8, or 12) split plot ANOVA was conducted to explore the impact on success rates by login**.** No interaction was found among delay, context, and the number of passcodes, $p > .05$. There was an interaction between delay and context (see figures 17 and 18), $F(1, 36) = 8.51$, $p < .001$, partial $\eta^2 = .321$, such that there was a larger difference between time 1 and 2 for the no context condition compared to the context condition. However, this difference was observed because of better performance at time 1 for the no context condition rather than better performance at time 2 for the context condition. A difference was found between the delay conditions, $F(1, 36) = 0.05$, $p = .819$, partial $\eta^2 = .001$, such that participants performed better during time 1 ($M = .95$; $SD =$ .16) than time 2 (M =.62; $SD = .27$). There was no difference between context and no context, $F(1, 36) = 3.45$, $p = .071$, partial $\eta^2 = .088$. A post hoc power analysis showed that that same size of 42 would provide for a power of .015 for an effect size of this size. There was an interaction found between delay and number of passcodes (see figure 17 and 18), $F(2, 36) = 8.51$, $p < .001$, partial $\eta^2 = .321$. Post hoc comparisons revealed that, collapsing across delay and context, there was no difference between 4 and 8 passcodes, 8 and 12 passcodes, or 4 and 12 passcodes, $p > .05$ for all. Differences were not observed between 4 ($M = .97$, $SD = .06$) and 12 passcodes ($M = .98$, $SD = .03$) at time 1 but were observed between 4 ($M = .79$, $SD = .21$) and 12 passcodes ($M = .47$, $SD = .27$) after a delay. A linear regression analysis was performed to check if there was a flat trend between error rates and the number of passcodes. It was found that there was a relationship between number of passcodes and error rates at time 2, $F(1, 41) = 13.28$, $p < .001$.

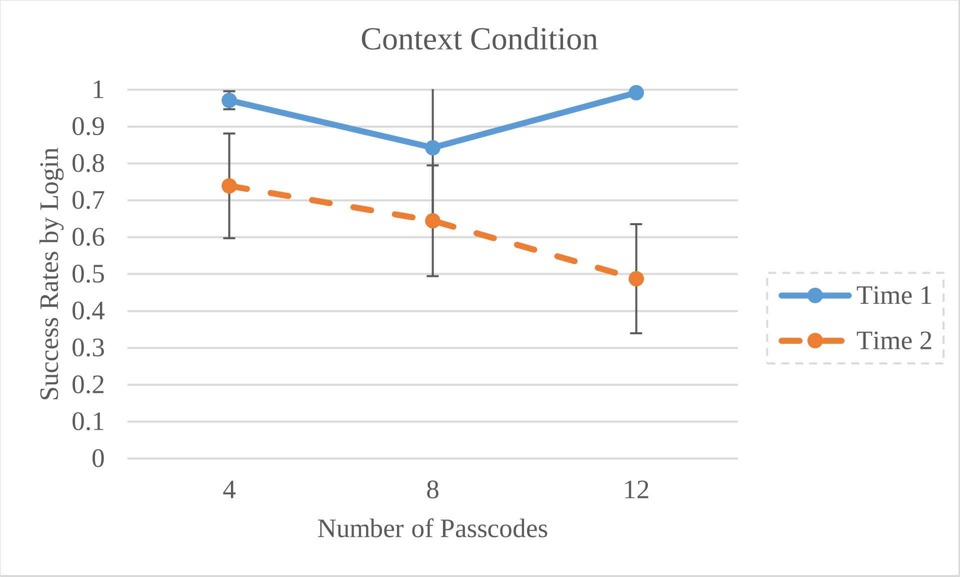

Figure 17. Success rates by login. Error bars represent 95% confidence intervals.

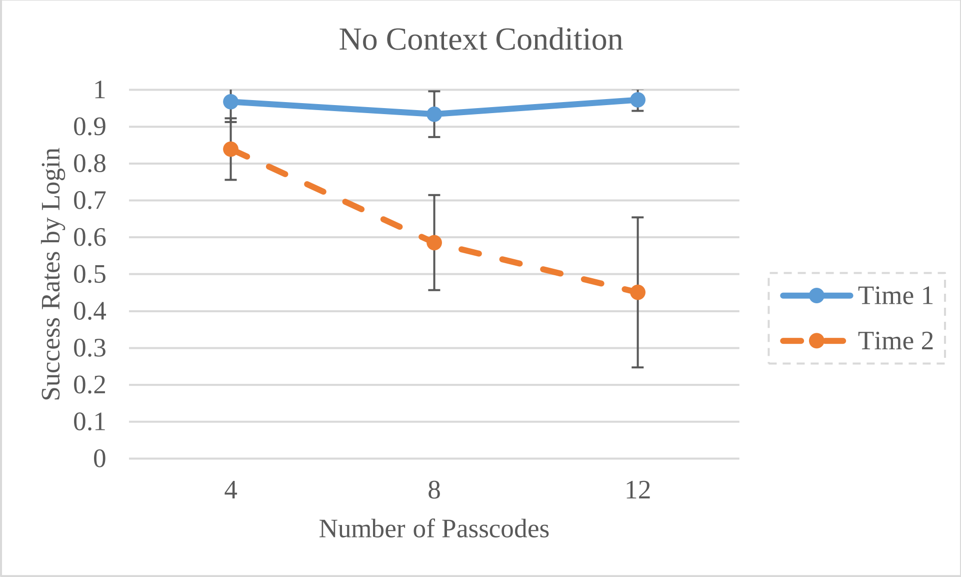

Figure 18. Success rates by login. Error bars represent 95% confidence intervals.

Memorability as Measured by Success Rates by Trial. Success rates by trial were calculated as the number of correct trials of the total trials. The data were cleaned by removing outliers that were more than 2.5 standard deviations above and below the mean for all participants. One participant was removed. A 2 (delay: immediate or three weeks) x 2 (context: yes or no) x 3 (number of passcodes: 4, 8, or 12) split plot ANOVA was conducted to explore the impact on success rates by trial**.** There was no interaction found among delay, context, and number of passcodes and no interaction found between delay and context, $p > .05$ for all. There was an interaction found between delay and number of passcodes (see figure 19), $F(2, 35) = 10.64$, $p < .001$, partial $\eta^2 = .378$, such that performance declined at time 2 but only for 12 passcodes. During time 1, participants performed similarly regardless of the number of passcodes (*M* for 4 passcodes = .99; *SD* = .02; *M* for 8 passcodes = .95; *SD* = .13; *M* for 12 passcodes = .99; *SD* = .01). After three weeks, participants still performed similarly for 4 and 8 passcodes (*M* for 4 passcodes = .92; *SD* = .07; *M* for 8 passcodes = .82; *SD* = .13), but performance dropped for 12 passcodes (*M* for 12 passcodes = .72; *SD* = .17). A difference was found between the delay conditions, $F(1, 35) = 72.87$, $p < .001$, partial $\eta^2 = .676$, such that participants performed better during time 1 (*M* = .99; *SD* = .02) than time 2 (M =.83; *SD* = .15). There was no difference between context and no context, $F(1, 35) = 0.36$, $p = .551$, partial $\eta^2 = .010$. A linear regression analysis found that there was a relationship between number of passcodes and error rates at time 2, $F(1, 40) = 19.35$, $p < .001$.

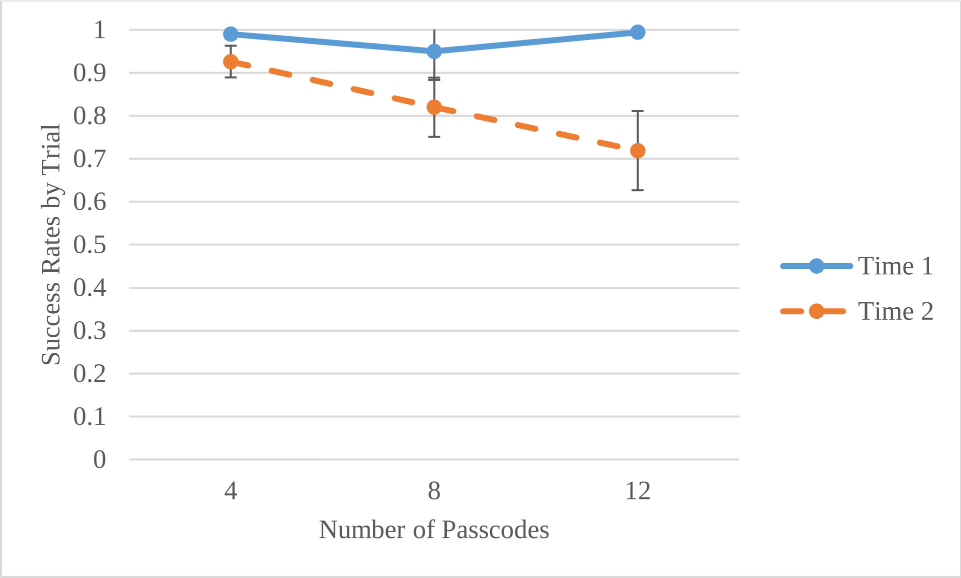

Figure 19. Success rates by trial. Error bars represent 95% confidence intervals.

Memorability as Measured by Login Time per Login. Login times per login were calculated for correct, experimental logins. The data were cleaned by removing outliers that were more than 2.5 standard deviations above and below the mean at the participant level. A 2 (delay: immediate or three weeks) x 2 (context: yes or no) x 3 (number of passcodes: 4, 8, or 12) split plot ANOVA was conducted to explore the impact on login times per login**.** There was no interaction found among delay, context, and number of passcodes, between delay and context, or between delay and number of passcodes, *p* > .05 for all. A difference was found between the delay conditions (see figure 20), *F*(1, 36) = 15.44, *p* < .001, partial η^2 = .300, such that participants performed faster during time 1 (*M* = 5.31; *SD* = 1.94) than time 2 (M = 6.25; *SD* = 1.88). There was no difference between context and no context, *F*(1, 36) = 0.26, *p* = .616, partial η^2 = .007. Post hoc comparisons revealed that collapsing across delay and context, there was no difference between 4 and 8 passcodes, 8 and 12 passcodes, or 4 and 12 passcodes, *p* > .05 for all.

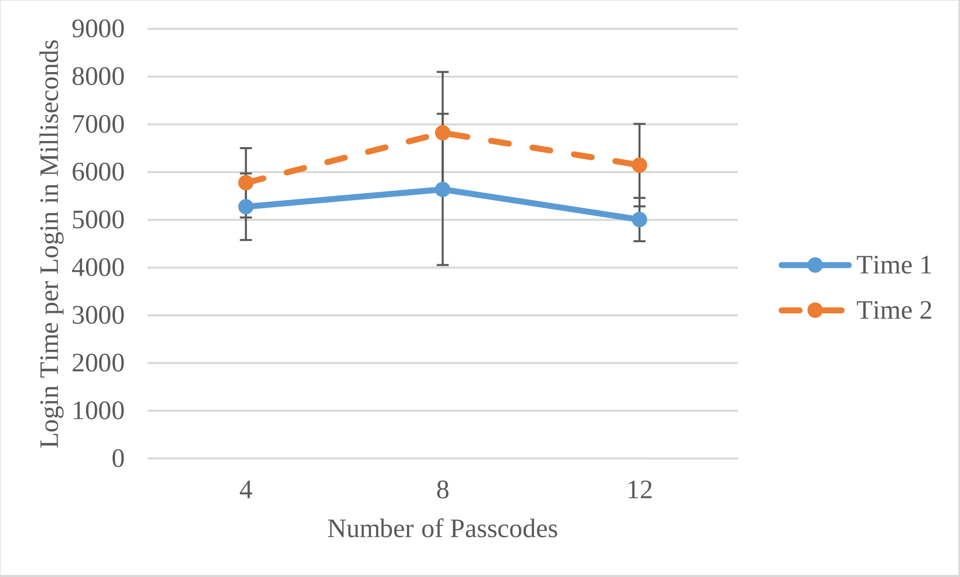

Figure 20. Login time per login. Error bars represent 95% confidence intervals.

Memorability as Measured by Login Time per Trial. Login times per trial were calculated for correct, experimental trials. The data were cleaned by removing outliers that were more than 2.5 standard deviations above and below the mean at the participant level. A 2 (delay: immediate or three weeks) x 2 (context: yes or no) x 3 (number of passcodes: 4, 8, or 12) split plot ANOVA was conducted to explore the impact on login times per trial**.** There was a three way interaction found among delay, context, and number of passcodes, *F*(2, 36) = 4.27, *p* = .022, partial η^2 = .192. Login times per trial were slowed for the 12 passcode condition, but only at time 2 and when there was context provided. There was no interaction found between delay and context or between delay and number of passcodes, *p* > .05 for both. A difference was found between the delay conditions (see figure 21 and 22), *F*(1, 36) = 39.31, *p* < .001, partial η^2 = .522, such that participants performed faster during time 1 (*M* = 1.71; *SD* = 4.07) than time 2 (*M* = 2.08; *SD* = 5.13). There was no difference between context and no context, *F*(1, 36) = 0.30, *p* =

.585, partial $\eta^2 = .008$. Post hoc comparisons revealed that, collapsing across delay, there was no difference between 4 and 8 passcodes, 8 and 12 passcodes, or 4 and 12 passcodes, $p > .05$ for all.

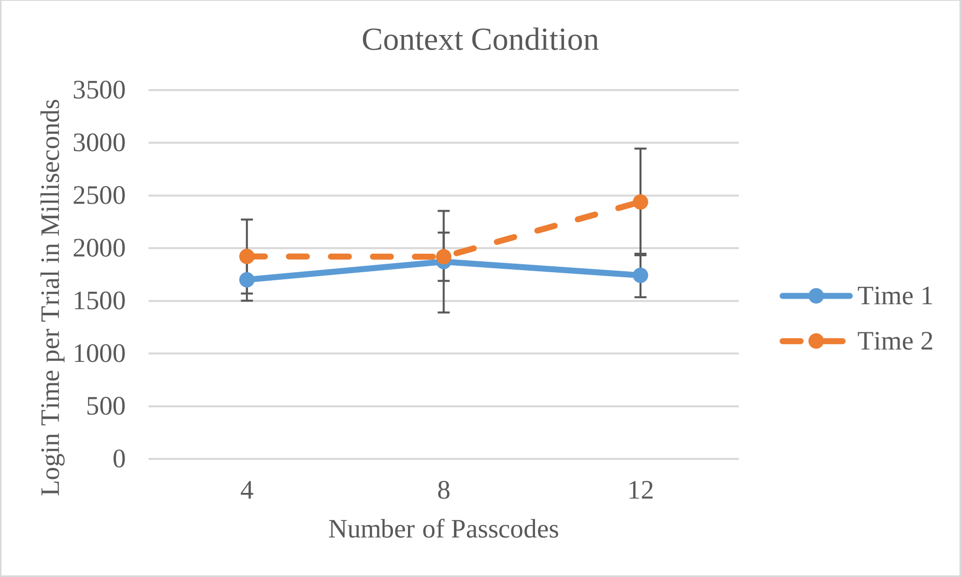

Figure 21. Login time per trial. Error bars represent 95% confidence intervals.

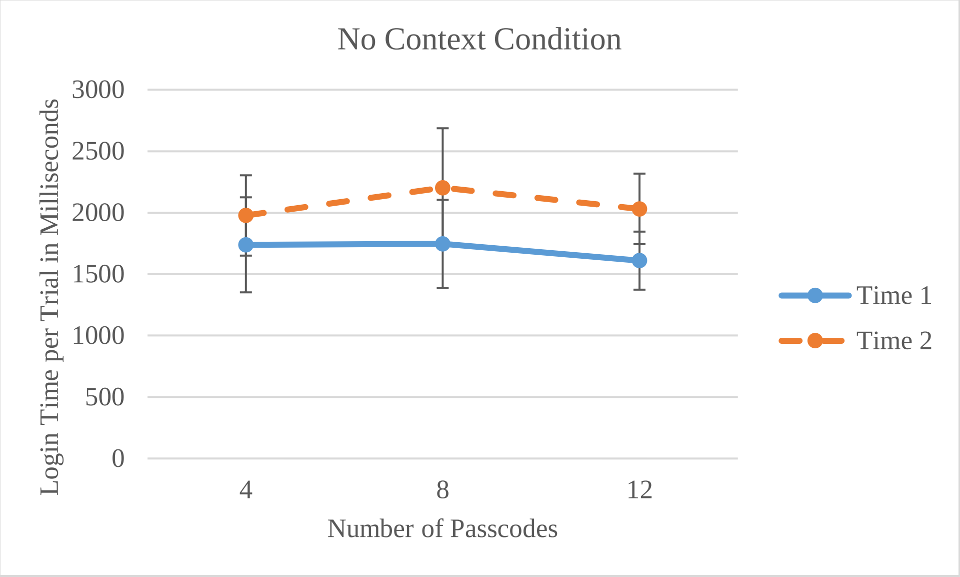

Figure 22. Login time per trial. Error bars represent 95% confidence intervals.

Memorability as Measured by Login Time per Login Including Incorrect Attempts. Login times were calculated for correct experimental logins including the time for the previous incorrect attempts. The data were cleaned by removing outliers that were more than 2.5 standard deviations above and below the mean at the participant level. A 2 (delay: immediate or three weeks) x 2 (context: yes or no) x 3 (number of passcodes: 4, 8, or 12) split plot ANOVA was conducted to explore the impact on login times**.** The researchers collapsed across number of attempts. There was no interaction found among delay, context, and number of passcodes, between delay and context, or between delay and number of passcodes, *p* > .05 for all. There was no difference between delay conditions, p > .05, or between context and no context, *F*(1, 36) = 0.87, *p* = .357, partial η^2 = .024. Post hoc comparisons revealed that collapsing across delay and context, there was no difference between 4 and 8 passcodes, 8 and 12 passcodes, or 4 and 12 passcodes (see figure 23), *p* > .05 for all. There was a bimodal distribution of login times for 8 passcodes. Participants either remembered their passcode and logged in quickly, or they had many failed attempts and logged in slowly.

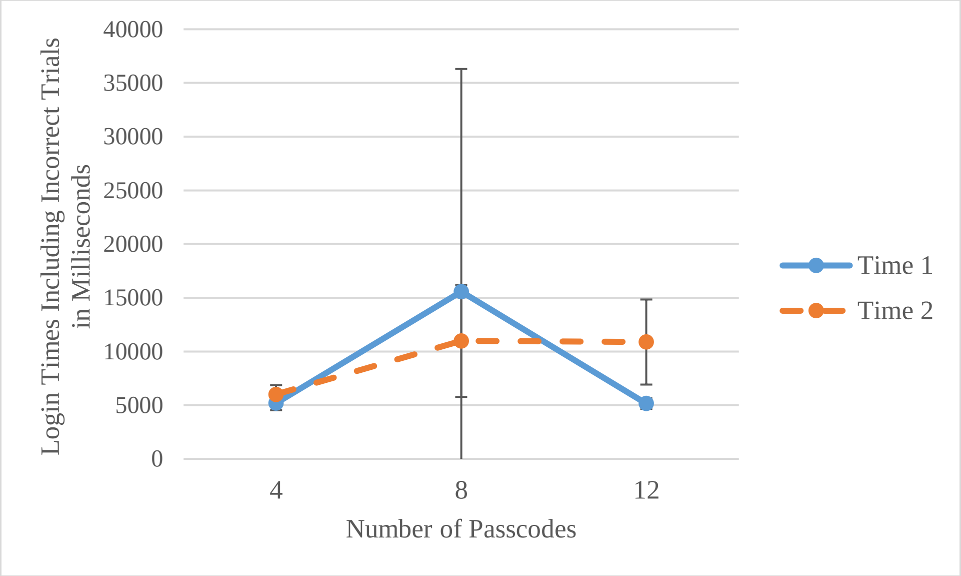

Figure 23. Login time per login including incorrect attempts. Error bars represent 95% confidence intervals.

CHAPTER 6: STUDY 4: CAPACITY LIMITS FOR LONGER GRAPHICAL PASSCODES

In addition to memorability issues associated with multiple graphical passcodes, the proposed research also assessed limitations for remembering long passwords. In many authentication systems, there is a threat of brute force attacks. The best way to defend against a brute force attack is to use a large password space (Suo, Zhu, & Owen, 2005). For example, alphanumeric passwords have the password space of 94^N, where 94 is the number of available characters on a keyboard, and N is the length of a password (Suo et al., 2005). Just as alphanumeric passwords are encouraged to be long, the security of graphical passcodes also benefits from the length. There are well-established rules for making strong alphanumeric passwords by making them long and complex. No previous studies have examined the impact of similarly making graphical passcodes more secure by making them longer. Lengthy alphanumeric passwords are difficult to remember, and so they have a limited dimensional space. However, graphical passcodes allow for almost limitless recognition memory (Standing, 1973; Standing et al., 1970) producing a larger dimensional space that offers greater resistance to brute force attacks.

The fourth proposed study attempted to leverage context-dependent memory to improve memorability similarly to the goal of study 3. Furthermore, study 4 determined the length of a passcode that can be remembered after the delay (3, 8, or 13 images). It was hypothesized that there would not be a decrement in memorability when participants need to remember longer passcodes after a three week delay, as evidenced by a flat function.

Method

Study 4 is the same as study 3 except that participants logged in with only one passcode, but it will have either three, eight, or 13 targets in it. The only scene included was a forest.

Participants. Forty-two undergraduate students participated (females = 22). They were recruited through the SONA system and compensated with class research credit or were recruited with fliers and compensated with \$20. Two participants reported being left hand dominant. Ages ranged from 18 to 40 (*M* = 22.76, *SD* = 4.93). Reported computer use ranged from 2 to 15 hours a day (*M* = 7.19, *SD* = 3.84). All participants reported normal or corrected to normal vision.

Results

Memorability was measured as overall error rates at time 1 and three weeks later as well as login times at time 1 and 2. A 3 (number of attempts: first three, second three, or third three) x 2 (delay: immediate or three weeks) x 2 (context or no context) x 3 (length of passcode: 3, 8, or 13) split plot ANOVA was conducted to explore the impact on error rates. Delay and number of attempts are a within-subjects variables, and context and length are between-subjects variables. Bonferroni corrections were used for all post hoc comparisons.

Memorability as Measured by Success Rates by Login. Success rates by login were calculated in the same way as in study 3. The researchers conducted a 2 (delay: immediate or three weeks) x 2 (context: yes or no) x 3 (length of passcode: 3, 8, or 13) split plot ANOVA was conducted to explore the impact on error rates. There were no differences found among delay, context, and number of passcodes, there was no interaction between delay and context, between delay and length of passcode, or between context and length of passcode, $p > .05$ for all (see figure 24). A difference was not found between the delay conditions, $p > .05$, between context and no context, $F(1, 36) = 3.45$, $p = .071$, partial $\eta^2 = .088$. A post hoc power analysis showed that that same size of 42 would provide for a power of .015 for an effect size of this size for context and no context. Post hoc comparisons revealed that, collapsing across delay, there was no difference between performance for passcodes of length 3 (*M* = .95, *SD* = .12) and 8 (*M* = .75, *SD* = .38), *p* = .18. Differences were observed between performance on passcodes of length 3 and 13 (*M* = .19, *SD* = .36) and between performance on passcodes of length 8 and 13, *p* < .001 for both. A linear regression analysis was performed to check if there was a flat trend between error rates and the length of passcodes. It was found that there was a linear relationship between the length of passcodes and error rates at time 2, *F*(1, 41) = 35.42, *p* < .001.

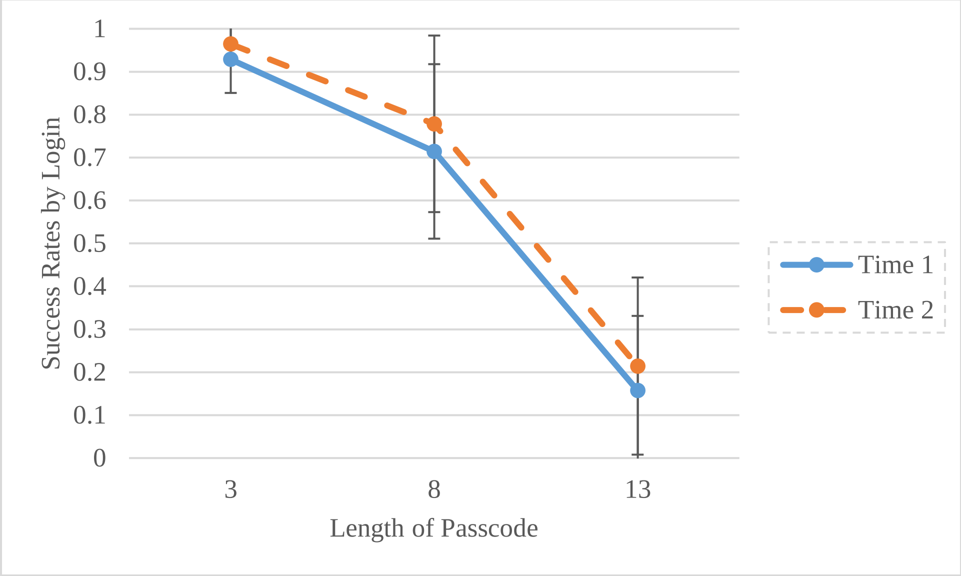

Figure 24. Success rates by login. Error bars represent 95% confidence intervals.

Memorability as Measured by Success Rates by Trial. Success rates by trial were calculated the same as in study 3. The data were cleaned by removing outliers that were more than 2.5 standard deviations above and below the mean for all participants. One participant was removed. The researchers conducted a 2 (delay: immediate or three weeks) x 2 (context: yes or no) x 3 (length of passcode: 3, 8, or 13) split plot ANOVA was conducted to explore the impact on error rates. There were no differences found among delay, context, and number of passcodes, there was no interaction between delay and context, between delay and length of passcode, or between context and length of passcode, $p > .05$ for all (see figure 25). A difference was not found between the delay conditions, $p > .05$, between context and no context, $F(1, 35) = 0.01$, $p = .928$, partial $\eta^2 = .000$. Post hoc comparisons revealed that, collapsing across delay and context, there was no difference between performance for passcodes length 3 ($M = .98$, $SD = .04$) and 8 ($M = .88$, $SD = .17$) or length 8 and 13 ($M = .82$, $SD = .14$), $p > .05$ for both. There was a difference between performance for passcodes length 3 and 13, $p = .005$. A linear regression analysis found that there was a relationship between number of passcodes and error rates at time 2, $F(1, 40) = 12.63$, $p < .001$.

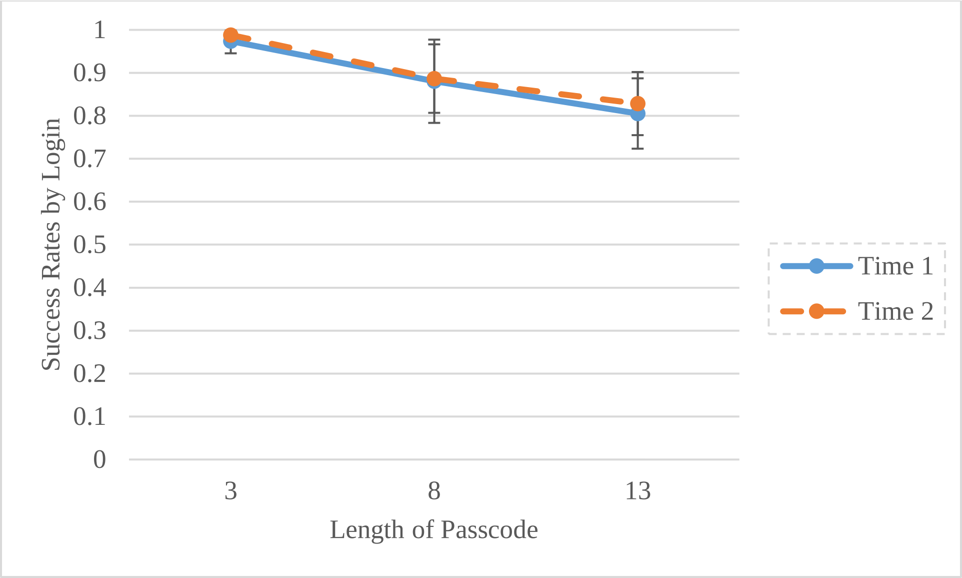

Figure 25. Success rates by trial. Error bars represent 95% confidence intervals.

Memorability as Measured by Login Time per Login. Login times per login were calculated in the same way as in study 3. Fourteen participants could not be included in this analysis because they did not correctly login at least one time. As the majority of unsuccessful logins were in the condition of passcodes of length 13, the analysis could not be completed for this dependent variable.

Memorability as Measured by Login Time per Trial. Login times per trial were calculated in the same way as in study 3. The data were cleaned by removing outliers that were more than 2.5 standard deviations above and below the mean at the participant level. A 2 (delay: immediate or three weeks) x 2 (context: yes or no) x 3 (length of passcode: 3, 8, or 13) split plot ANOVA was conducted to explore the impact on login times per trial**.** There were no differences found among delay, context, and length of passcode, and there was no interaction found between delay and context, between delay and number of passcodes, or between context and length of passcode, *p* > .05 for all. A difference was found between the delay conditions (see figure 26), *F*(1, 36) = 19.34, *p* < .001, partial η^2 = .349, such that participants performed faster during time 1 (*M* = 2.88; *SD* = 1.74) than time 2 (*M* = 2.01; *SD* = 6.33). There was no difference between context and no context, *F*(1, 36) = 1.13, *p* = .295, partial η^2 = .030. Post hoc comparisons revealed that collapsing across delay and context, there was no difference between 3 and 13 or 8 and 13 passcodes, *p* > .05 for both. Participants performed faster for 3 passcodes (*M* = 1.78; *SD* = 4.56) than 8 (*M* = 3.07; *SD* = 1.94), *p* = .006. There was a bimodal distribution resulting in large confidence intervals for passcodes of length 8.

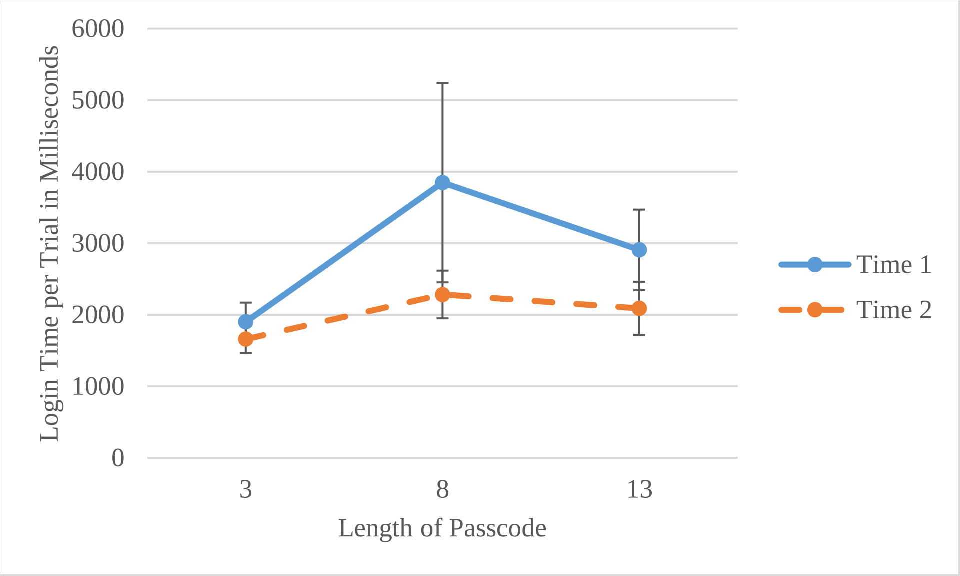

Figure 26. Login time per trial. Error bars represent 95% confidence intervals.

Memorability as Measured by Login Time per Login Including Incorrect Attempts. Login times were calculated for correct experimental logins including the time for the previous incorrect attempts. Thirteen participants could not be included in this analysis because they did not correctly login at least one time. As the majority of unsuccessful logins were in the condition of passcodes of length 13, the analysis could not be completed for this dependent variable.

DISCUSSION

There was a smaller difference found between time 1 and 2 when context was provided for the for the eight passcode condition when compared to the no context condition when error rates were measured by login, which would partially support the hypothesis that context would improve memorability for multiple graphical passcodes. However, this difference was observed because of differences between performance at time 1 rather than at time 2. The prediction that context would improve memorability was based on previous literature that consistent context facilitates retrieval (Godden & Baddeley, 1975). Context did not improve memorability for the four or 12 passcode conditions, and it did not improve memorability when errors were coded by trial. Future research should continue to explore avenues for improving memorability. For example, a background image may not have been a strong enough context intervention, but possibly a more immersive visual or auditory experience at login could improve memorability for more than eight passcodes.

When participants needed to remember longer passcodes, I expected that at time 2 rates for memorability would be improved for the context condition compared to the no context condition. However, the context condition did not facilitate memorability, and memorability was not demonstrated in either condition. In previous studies that demonstrated context facilitating memory, the context interventions were strong. For example, participants were in different settings, such as on land or submerged underwater (Godden & Baddeley, 1975) or in a basement or outdoors (Sahakyan, 2010). A strong context intervention such as these was not possible in the application of a graphical passcode because the context needed to be presented on a computer monitor. It was likely that the context intervention was not strong enough to improve memorability, even though some previous studies have found a context effect for seemingly weak context interventions (e.g., listening to music (Balch, Bowman, & Mohler, 1992; Smith, 1985) or chewing gum (Baker, Bezance, Zellaby, & Aggleton, 2004)). A background scene on a computer motor has been found to provide context effects (Stefanucci et al., 2007). However, in that instance, there was a very large monitor accompanied by ambient noises. Context has not been found for other studies that use smaller manipulations, such as background colors (Petrich & Chiesi, 1976) or the flavor of chewing gum (Johnson & Miles, 2008). It is also possible that the context intervention was not successful because it confused or distracted participants from their primary goal. Future research could consider whether a stronger context condition could be provided or whether the purpose of the context could be clarified so as not to confuse participants. Future research could potentially make the context intervention stronger by including only targets and distractors that are associated with the background scenes to allow for a greater depth of processing. If the distractors as well as the targets were associated with the scenes, this would not present a security weakness because the targets would not stand out as the only associated images.

The weak context intervention that was used for studies 3 and 4 was selected because this research builds on previous studies in the domain of authentication, which have incorporated context but have not tested the effectiveness of the manipulation. These studies have used context interventions such as asking participants to think of stories that include their passcode (Davis et al., 2004; Gao et al., 2010) or allowing participants to authenticate by selecting points on background images (Bulling et al., 2012; Wiedenbeck et al., 2005). Because studies 3 and 4 used similarly subtle context interventions, a contribution was made to the literature that the effectiveness of context interventions in the domain of authentication cannot be assumed.

I predicted that there would not be differences in error rates after the delay regardless of the number of passcodes that participants need to remember. This prediction was based on previous findings of participants' almost limitless memory for pictures (Standing et al., 1970). When success rates were coded by login, there was no difference between 4, 8 and 12 passcodes after the delay. This measure is less sensitive than measuring success rates by trial because a login requires that all three trials be successful. When success rates were measured by trial, there was no difference in performance on 4 or 8 passcodes after the delay. Performance dropped only for 12 passcodes, the most challenging condition, although they were still at an acceptable 72% (Behl et al., 2014; Liu et al., 2011a). Memory for the passcodes may not have been as limitless as picture memory has been found to be in previous studies because the images were degraded. The applied finding of this study is noteworthy because it has demonstrated that participants can remember eight, system-assigned passcodes after a three-week delay. Therefore, graphical authentication may be an effective solution for the problem of memorability that accompanies the proliferation of services for which users need to identify themselves. However, 12 passcodes were difficult to remember after three weeks. It may be that having the same scheme, e.g., Use Your Illusion (UYI), of graphical authentication for 12 accounts may be too much. Future research should investigate whether having different types of graphical schemes can meet the need to remember passcodes for more than eight accounts. For example, possibly participants can remember six passcodes consisting of degraded images for UYI and concurrently six passcodes consisting of icons for CHC successfully.

I predicted that there would not be differences in error rates after the delay regardless of the length of passcodes that participants needed to remember. When success rates were coded by login, there was no difference between length three and eight or eight and 13, but performance dropped from length three to 13. When success rates were measured by trial, there was no difference in performance for length three and eight, but there was a difference between eight and 13 and between three and 13. As in study 3, memory for degraded images has been found not to be limitless, demonstrated by the most challenging condition. These findings are noteworthy because, while longer passcodes are desirable in high-security situations to increase bit strength, passcodes of length 13 have been found to be detrimental to memorability. To increase security, passcodes could be lengthened from three to eight without problems with performance. This memorability experiment challenged participants to remember a system-assigned passcodes after three weeks. It is desirable to use system-assigned graphical passcodes to avoid user biases. However, future research could investigate whether it is possible that a password that is used more often than every three weeks could allow for passcodes longer than eight.

The researchers also measured login times because faster login times may demonstrate ease of retrieval from memory. There were no differences among login time among the numbers of passcodes, and login times were only found to be faster at time 1 than after the delay. These findings may suggest that four, eight, and 12 passcodes were received with similar ease, but these findings are inconclusive. When participants logged in with longer passcodes, login times could not be analyzed per login because most participants could not successfully login for passcodes of length 13. When analyzed per trial, shorter passcodes were found to be retrieved from memory more quickly than longer passcodes, suggesting that when security is less important, shorter passcodes will be more usable.

Based on the findings about the number and length of passcodes, it was noted that participants had better memory three weeks later for multiple short passcodes than one longer passcode. Previous studies that have examined long-term memory for visual objects have also found that memory is better for a smaller set size (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle, Brady, Alvarez, & Oliva, 2010a; Konkle, Brady, Alvarez, & Oliva, 2010b; Zhao & Turk-Browne, 2011). Konkle and colleagues showed participants a set of 2,800 pictures. Each picture depicted an item, such as a hat or a butterfly. Pictures were presented for three seconds, each separated by a mask. Within the set of pictures, there were either one, two, four, eight, or 16 exemplars of each item, meaning there were sets of, for example, one or multiple hats. After a ten-minute break, participants performed a forced-choice recognition task. The performance was better for smaller set sizes of exemplars.

Similarly, when participants saw a set of 2,500 pictures and performed a forced-choice recognition task, participants were found to perform better for pictures that were of different object classes than the same object class (Brady et al., 2008). Better memory for smaller set sizes extends to scenes as well as objects (Konkle et al., 2010b). Participants viewed 3,000 scenes and performed a forced-choice recognition task. During the presentation, scenes were either unrelated or from a set of four, 16, or 64 related scenes. Participants performed better for smaller set sizes. Participants have also performed better for smaller set sizes when reporting numerosity (Zhao & Turk-Browne, 2011). They viewed objects and indicated whether each was natural or artificial. Each object was either unrelated to the other objects or belonged within a set of exemplars ranging from one to ten. After a break, participants reported how many times they had seen each class of object. Numerosity reports were more accurate with smaller sets of exemplars. These previous studies employ exemplars from existing, natural categories. Future research should investigate memory for sets that are newly formed, similar to images used in UYI and other graphical passcodes.

Set size has also been examined for contextual or associative learning. Voss (2009) specifically examined set size and associative long-term memory. The experimenter learned left-right button associations with a set of 4,980 pictures. He learned 30 associations in the first session, and he added 30 pictures in each subsequent session. Four hours after each learning session, he would test himself. He performed better with a smaller set of associations (e.g., 85% correct for 1,000 pictures) compared to a larger set (e.g., 65% correct for about 5,000 pictures).

Previous studies that have examined set size and long-term memory only observed limitations in long-term memory for pictures for huge databases of pictures. The current studies observed limitations for much smaller picture sets. However, this is not surprising because in the current studies 3 and 4, participants were tasked with remembering pictures that were degraded. Loftus, Kaufman, Nishimoto, and Ruthruff (1992) demonstrated that when images are degraded, it should be expected to observe greater limitations in long-term memory. Participants saw images of scenes such as farmland and then performed a recognition task. Each image was viewed for between 55 and 400 milliseconds. A subset of the images was degraded by removing contrast. Participant either viewed degraded images both at learning and at testing or at neither learning nor testing. Participants performed better when images were consistently either degraded or clean at learning and testing, and they performed better overall for recognizing clean than degraded images. Similar to findings of studies 3 and 4 of the current research, participants present greater limitations in long-term memory when there is less encoding specificity, such as when they are presented with a whole image for UYI and asked to recognize a degraded image, and when the features that are encoded during learning are not present when recognizing degraded images.

In addition, the memory limitation observed in studies 3 and 4 can be accounted for by the design of these experiments compared with previous studies that have found almost limitless memory for pictures. In previous studies, participants performed forced-choice recognition tasks, selecting between a picture that was previously learned and a new picture (Standing et al., 1970). The learned picture could be relatively easily identified because it produced a familiarity signal, while the new picture did not. Studies 3 and 4 tasked participants with identifying a learned target from a set of eight distractors. All the of distractors were recycled for each login because if the distractors changed, this would produce a risk of intersection attack. Therefore the targets and the distractors would both produce familiarity signals, and it was likely more challenging for the participants to remember their targets than in previous studies.

CHAPTER 7: CONCLUSIONS

Our goal is to further develop usable security theory from a human-centered perspective while operating under the necessary security constraints. The first two studies benchmarked the relative strengths and weaknesses among schemes that use common strategies to provide OSA resistance. Prototypical schemes were selected that represented commonly used strategies to provide OSA resistance, including grouping targets among distractors, translating targets to another location, disguising targets, and using gaze-based input. These schemes were measured by a variety of dependent variables, and they were compared with the alphanumeric approach. Dependent measures included error rates, learnability, login times, memorability, acceptance, satisfaction, and OSA performance. OSA performance was measured given one viewing of a login, three viewings, and the percent of passcodes identified.

Numerous dependent variables were used to benchmark the schemes to produce generalizability to a wide range of previous literature. Previous literature tends to use just a few dependent variables. By including many measures (e.g., by measuring OSA performance in multiple ways), findings can more easily be compared with specific measures from previous studies.

Findings showed that prototypical graphical passcodes offer resistance to OSA. It has been a concern that graphical schemes would be vulnerable to OSAs because pictures may be clearly visible to both users and attackers. However, schemes have been successfully designed to resist this type of attack, making graphical schemes a viable alternative to the alphanumeric approach. Findings also showed that graphical passcodes were memorable. The passcodes used in studies 1 and 2 were system-assigned, and performance did not drop after a three-week delay for the graphical passcodes but did for the alphanumeric approach. These findings verify that graphical passcodes offer a valuable advantage of memorability over the current widely used approach (Brostoff et al., 2010; Wiedenbeck et al., 2006). This advantage for memorability is especially important because users are needing to remember login credentials for an increasing number of services.

While finding demonstrated security against OSAs and advantages for memorability, study 1 showed that error rates for graphical passcodes are still a potential weakness. Error rates were between 18 and 42%, which aligns with previous literature. Improvements in error rates could make graphical passcodes a more appealing alternative to the alphanumeric scheme. It is necessary that users of novel graphical passcodes be able to verify their identities and not be blocked from their primary tasks, which would lead to frustration.

In subsequent studies investigating larger numbers of passcodes and longer passcodes, the researchers attempted to mitigate the discovered shortcoming of error rates by providing context in the form of a background scene to each passcode. By leveraging context, the researchers aimed to offer a solution that can be applied to any graphical passcode and make graphical passcodes a more feasible alternative.

Also in these subsequent studies, the limits of memorability of graphical passcodes when there are a greater number of passcodes and when passcodes have a greater length were determined. It is essential to determine how well the advantages of recognition memory for pictures holds when it is considered that participants need to remember passcodes for many different accounts and when it is considered that longer passcodes are needed to defend against brute force attacks. In study 3 participants logged in with multiple graphical passcodes, and in study 4 participants logged in with longer graphical passcodes. In both studies, participants returned and logged in again three weeks later.

When participants logged in with multiple graphical passcodes, context appeared to improved memorability for eight passcodes, based on a smaller difference in error rates at time 1 and 2 for this condition. However, the difference was due to time1 and not to time 2, suggesting context did not improve memorability. Context did not improve memorability when participants needed to remember longer graphical passcodes in study 4. It is likely that a stronger context intervention will need to be used if future researchers pursue this avenue.

Findings also showed that participants perform well when remembering passcodes of length eight and can remember eight graphical passcodes after a three-week delay, but performance dropped for passcodes of length 13 and for 12 passcodes. These findings show that graphical schemes could potentially help with the hurdle of helping users to remember login credentials for many accounts and more sensitive information, and future research should address the challenge of improving memorability when participants need to remember more than eight passcodes or passcodes longer than eight pictures.

This research benchmarked common graphical schemes, identified strengths and shortcomings, examined the limits of memorability, and translated basic research about context to the domain of graphical authentication. And, aimed to increase graphical passcodes readiness as an authentication option.

TIPS FOR PRACTITIONERS

The following tips for practitioners include general guidance that comes from previous literature and our findings:

- Choose graphical schemas that use the strategy of grouping, such as CHC, to promote better success rates.

- To promote learnability, choose a graphical scheme that uses the strategy of disguising the passcode, such as UYI.
- For the best learnability, choose the familiar alphanumeric scheme.
- For faster login times, choose a graphical scheme that uses the strategy of disguising, such as UYI.
- For better security against OSAs, choose a graphical scheme that uses the strategy of grouping, such as CHC, or translating to another location, such as WYSWYE.
- When numerous graphical passcodes are needed for different accounts, limit the number of passcodes to eight.
- When longer graphical passcodes are needed for security, limit the length to eight.
- When users are remembering multiple graphical passcodes, context in the form of a background image can improve memorability.

REFERENCES

Al Ameen, M. N. (2016). *The impact of cues and user interaction on the memorability of system-assigned random passwords* (Doctoral dissertation). Retrieved from UTA Libraries.

Ankush, D. A., & Husain, S. S. (2014). Authentication Scheme for Shoulder surfing using Graphical and Pair Based scheme. *International Journal*, *2*, 161-166.

Arianezhad, M., Stebila, D., & Mozaffari, B. (2013). Usability and security of gaze-based graphical grid passwords. In *International Conference on Financial Cryptography and Data Security* (pp. 17-33). Springer, Berlin, Heidelberg.

Arya, Y. D. S., & Agarwal, G. (2011). Impact of Background Images on the DAS (Draw-A-Secret) Graphical Password Authentication Scheme. *IJCA Special Issue on "Network Security and Cryptography" NSC* (pp. 47-50).

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review*, *89*, 708-729.

Baddeley, A. D., & Woodhead, M. (1982). Depth of processing, context, and face recognition. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *36*, 148-164.

Baker, J. R., Bezance, J. B., Zellaby, E., & Aggleton, J. P. (2004). Chewing gum can produce context-dependent effects upon memory. *Appetite, 43*, 207-210.

Balch, W. R., Bowman, K., & Mohler, L. A. (1992). Music-dependent memory in immediate and delayed word recall. *Memory & Cognition*, *20*, 21-28.

Barton, B. F., & Barton, M. S. (1984). User-friendly password methods for computer-mediated information systems. *Computers & Security*, *3*, 186-195.

Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology*, *6*, 325-356.

Begg, I. (1973). Imagery and integration in the recall of words. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *27*, 159-167.

Begg, I. (1978). Imagery and organization in memory: Instructional effects. *Memory & Cognition*, *6*, 174-183.

Behl, U., Bhat, D., Ubhaykar, N., Godbole, V., & Kulkarni, S. (2014). Multi-level scalable textual-graphical password authentication scheme for web based applications. *REV Journal on Electronics and Communications*, *3*, 166-124.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 403-410.

Bianchi, A., Oakley, I., & Kim, H. (2016). PassBYOP: Bring your own picture for securing graphical passwords. *IEEE Transactions on Human-Machine Systems*, *46*, 380-389.

Biddle, R., Chiasson, S., & Van Oorschot, P. C. (2012). Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys (CSUR)*, *44*, 1-25.

Bošnjak, L., & Brumen, B. (2018, June). Improving the Evaluation of Shoulder Surfing Attacks. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics* (pp. 19-20).

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*, 14325-14329.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*, 4-7.

Brostoff, S., Inglesant, P., & Sasse, M. A. (2010). Evaluating the usability and security of a graphical one-time PIN system. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp. 88-97). British Computer Society.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*, 105-116.

Bulling, A., Alt, F., & Schmidt, A. (2012). Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3011-3020.

Cain, A. A., & Still, J. D. (2016). A Rapid Serial Visual Presentation Method for Graphical Authentication. *Advances in Human Factors in Cybersecurity*, 3-11.

Cain, A. A., Werner, S., & Still, J. D. (2017). Graphical Authentication Resistance to Over-the-Shoulder-Attacks. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2416-2422).

Cazier, J. A., & Medlin, B. D. (2006). Password security: An empirical investigation into e-commerce passwords and their crack times. *Information Systems Security*, *15*(6), 45-55.

Chakrabarti, S., Landon, G. V., & Singhal, M. (2007). Graphical passwords: Drawing a secret with rotation as a new degree of freedom. *Proceedings of the Fourth IASTED Asian Conference on Communication Systems and Networks*, 561-173.

Chen, Y. L., Ku, W. C., Yeh, Y. C., & Liao, D. M. (2013). A simple text-based shoulder surfing resistant graphical password scheme. In *2013 IEEE International Symposium on Next-Generation Electronics (ISNE),* (pp. 161-164).

Choong, Y. Y., & Greene, K. K. (2016). What’s a Special Character Anyway? Effects of Ambiguous Terminology in Password Rules. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*, 760-764.

Craik, F. I., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *The Quarterly Journal of Experimental Psychology*, *26*, 274-284.

Dallett, K., & Wilcox, S. G. (1968). Contextual stimuli and proactive inhibition. *Journal of Experimental Psychology*, *78*, 475-480.

Davis, D., Monrose, F., & Reiter, M. K. (2004). On User Choice in Graphical Password Schemes. In *USENIX Security Symposium* (Vol. 13, pp. 11-11).

Davis, E. T., Scott, K., Pair, J., Hodges, L. F., & Oliverio, J. (1999). Can audio enhance visual perception and performance in a virtual environment? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 43, pp. 1197-1201).

De Luca, A., Denzel, M., & Hussmann, H. (2009). Look into my eyes! Can you guess my password? *Proceedings of the 5th Symposium on Usable Privacy and Security*, 7-19.

De Luca, A., Hertzschuch, K., & Hussmann, H. (2010). ColorPIN: Securing PIN entry through indirect input. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1103-1106.

Dolinsky, R., & Zabrucky, K. (1983). Effects of environmental context changes on memory. *Bulletin of the Psychonomic Society*, *21*, 423-426.

Dosono, B., Hayes, J., & Wang, Y. (2015). “I’m Stuck!”: A Contextual Inquiry of People with Visual Impairments in Authentication. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)* (pp. 151-168).

Dourish, P., Anderson, K., & Nafus, D. (2007). Cultural mobilities: Diversity and agency in urban computing. In *IFIP Conference on Human-Computer Interaction* (pp. 100-113). Springer Berlin Heidelberg.

Dunphy, P., Fitch, A., & Olivier, P. (2008). Gaze-contingent passwords at the ATM.

In *4th Conference on Communication by Gaze Interaction (COGAIN)* (pp. 59-62).

Dunphy, P., & Yan, J. (2007). Do background images improve draw a secret graphical passwords? In *Proceedings of the 14th ACM conference on Computer and communications security* (pp. 36-47).

Eiband, M., Khamis, M., von Zezschwitz, E., Hussmann, H., & Alt, F. (2017). Understanding shoulder surfing in the wild: Stories from users and observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 4254-4265).

Emmerson, P. G. (1986). Effects of environmental context on recognition memory in an unusual environment. *Perceptual and Motor Skills, 63*, 1047-1050.

Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition, 13*, 333-345.

Forget, A., Chiasson, S., & Biddle, R. (2010). Shoulder-surfing resistance with eye-gaze entry in cued-recall graphical passwords. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1107-1110).

Fuglerud, K.S. and Dale, Ø. (2011). Secure and inclusive authentication with a talking mobile one-time-password client. *Security & Privacy, IEEE*, Vol. 9, pp. 27-34.

Gao, H., Guo, X., Chen, X., Wang, L., & Liu, X. (2008). Yagp: Yet another graphical password strategy. In *Computer Security Applications Conference, 2008.* (pp. 121-129).

Gao, H., Liu, X., Dai, R., Wang, S., & Chang, X. (2009). Analysis and evaluation of the Color login graphical password scheme. In *Fifth International Conference on Image and Graphics, 2009* (pp. 722-727).

Gao, H., Ren, Z., Chang, X., Liu, X., & Aickelin, U. (2010). A new graphical password scheme resistant to shoulder-surfing. In *2010 International Conference on Cyberworlds (CW),* (pp. 194-199).

Geiselman, R. E., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition*, *5*, 499-504.

Ghori, F., & Abbasi, K. (2013). Secure User Authentication Using Graphical Passwords. *Journal of Independent Studies and Research*, *11*, 34-40.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology*, *66*, 325-331.

Godden, D., & Baddeley, A. (1980). When does context influence recognition memory? *British journal of Psychology*, *71*, 99-104.

Grant, H. M., Bredahl, L. C., Clay, J., Ferrie, J., Groves, J. E., McDorman, T. A., & Dark, V. J. (1998). Context-dependent memory for meaningful material: Information for students. *Applied Cognitive Psychology*, *12*, 617-623.

Grawemeyer, B. & Johnson, H. (2011). Using and managing multiple passwords: A week to a view. *Interacting with Computers*, *23*, 256-267.

Gupta, S., Sahni, S., Sabbu, P., Varma, S., & Gangashetty, S. V. (2012). Passblot: A highly scalable graphical one time password system. *International Journal of Network Security & Its Applications*, *4*, 201-216.

Hayashi, E., Dhamija, R., Christin, N., & Perrig, A. (2008). Use your illusion: Secure authentication usable anywhere. *Proceedings of the 4th symposium on Usable privacy and security*, 35-45.

Herz, R. S. (1997). The effects of cue distinctiveness on odor-based context-dependent memory. *Memory & Cognition*, *25*, 375-380.

Hoanca, B., & Mock, K. (2006). Secure graphical password system for high traffic public areas. In *Proceedings of the 2006 symposium on Eye tracking research & applications* (pp. 35-35).

Holman, J., Lazar, J., Feng, J. H., & D'Arcy, J. (2007). Developing usable CAPTCHAs for blind users. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility* (pp. 245-246).

Horowitz, L. M., & Prytulak, L. S. (1969). Redintegrative memory. *Psychological Review, 76,* 519-531.

Hui, L. T., Bashier, H. K., Hoe, L. S., Kwee, W. K., & Sayeed, M. S. (2014). A Hybrid Graphical Password Scheme for High-End System. *Australian Journal of Basic and Applied Sciences*, *8*, 23-29.

IBM (2016). 2016 cost of data breach study: Global analysis. Ponemon Institute LLC.

Jain, A. K., & Nandakumar, K. (2012). Biometric Authentication: System Security and User Privacy. *IEEE Computer, 45*, 87-92.

Jenkins, R., McLachlan, J. L., & Renaud, K. (2014). Facelock: Familiarity-based graphical authentication. *PeerJ*, *2*, e444, 1-24.

Jermyn, I., Mayer, A. J., Monrose, F., Reiter, M. K., & Rubin, A. D. (1999). The design and analysis of graphical passwords. *Usenix Security*, 1-14.

Johnson, A. J., & Miles, C. (2008). Chewing gum and context-dependent memory: The independent roles of chewing gum and mint flavour. *British Journal of Psychology*, *99*, 293-306.

Johnson, K., & Werner, S. (2008). Graphical user authentication: A comparative evaluation of composite scene authentication vs. three competing graphical passcode systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*, 542-546.

Joshuva, M., Rani, T. S., & John, M. S. (2011). Implementing CHC to counter shoulder surfing attack in PassPoint–style graphical passwords. *International journal of advanced networking and applications*, *2*, 906-910.

Kawagoe, K., Sakaguchi, S., Sakon, Y., & Huang, H. H. (2012). Tag association based graphical password using image feature matching. In *International Conference on Database Systems for Advanced Applications* (pp. 282-286). Springer, Berlin, Heidelberg.

Kaye, J. J. (2011). Self-reported password sharing strategies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2619-2622.

Khot, R. A., Kumaraguru, P., & Srinathan, K. (2012). WYSWYE: Shoulder surfing defense for recognition based graphical passwords. *Proceedings of the 24th Australian Computer-Human Interaction Conference*, 285-294.

Kim, D., Dunphy, P., Briggs, P., Hook, J., Nicholson, J. W., Nicholson, J., & Olivier, P. (2010). Multi-touch authentication on tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1093-1102).

Kiran, T. S. R., Rao, K. S., & Rao, M. K. (2012). A novel graphical password scheme resistant to peeping attack. *IJCSIT) International Journal of Computer Science and Information Technologies*, *3*, 5051-5054.

Kirsner, K. (1973). An analysis of the visual component in recognition memory for verbal stimuli. *Memory & Cognition*, *1*, 449-453.

Koens, F., Ten Cate, O. T. J., & Custers, E. J. (2003). Context-dependent memory in a meaningful environment for medical education: in the classroom and at the bedside. *Advances in Health Sciences Education*, *8*, 155-165.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*, 558-578.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*, 1551-1556.

Kuber, R. and Sharma, S. (2012). Developing an extension to an existing tactile authentication mechanism to support non-visual interaction. In *Proceedings of the Conference on Human-Computer Interaction*, pp. 190-198.

Kumar, M., Garfinkel, T., Boneh, D., & Winograd, T. (2007). Reducing shoulder-surfing by using gaze-based password entry. In *Proceedings of the 3rd symposium on Usable privacy and security* (pp. 13-19).

Lashkari, A. H., Manaf, A. A., & Masrom, M. (2011). A secure recognition based graphical password by watermarking. In *11th International Conference on Computer and Information Technology* (pp. 164-170).

Lazar, J., Feng, J., Brooks, T., Melamed, G., Wentz, B., Holman, J., Olalere, A., & Ekedebe, N. (2012, May). The SoundsRight CAPTCHA: an improved approach to audio human interaction proofs for blind users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2267-2276).

Li, Z., Sun, Q., Lian, Y., & Giusto, D. D. (2005). An association-based graphical password design resistant to shoulder-surfing attack. In *IEEE International Conference on Multimedia and Expo, 2005*. (pp. 245-248).

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of verbal learning and verbal behavior*, *9,* 1-11.

Lin, D., Dunphy, P., Olivier, P., & Yan, J. (2007). Graphical passwords & qualitative spatial relations. In *Proceedings of the 3rd symposium on Usable privacy and security* (pp. 161-162).

Liu, X. Y., Gao, H. C., Wang, L. M., & Chang, X. L. (2011). An enhanced drawing reproduction graphical password strategy. *Journal of Computer Science and Technology*, *26*, 988-999.

Liu, X., Qiu, J., Ma, L., Gao, H., & Ren, Z. (2011b). A novel cued-recall graphical password scheme. *Image and Graphics (ICIG), 2011 Sixth International Conference on*, 949-956.

Loftus, G. R., Kaufman, L., Nishimoto, T., & Ruthruff, E. (1992). Effects of visual degradation on eye-fixation duration, perceptual processing, and long-term visual memory. In *Eye Movements and Visual Cognition* (pp. 203-226). Springer, New York, NY.

Lu, D., Lee, T., Das, S., & Hong, J. I. (2016). Examining Visual-Spatial Paths for Mobile Authentication. In *WAY@ SOUPS* (pp. 1-2).

Manjunath, G., Satheesh, K., Saranyadevi, C., & Nithya, M. (2014). Text-Based Shoulder Surfing Resistant Graphical Password Scheme. *International Journal of Computer Science & Information Technologies*, *5*, 2277-2280.

Martin, K. M., & Aggleton, J. P. (1993). Contextual effects on the ability of divers to use decompression tables. *Applied Cognitive Psychology*, *7,* 311-316.

Meng, Y., & Li, W. (2013). Enhancing click-draw based graphical passwords using multi-touch on mobile phones. In *IFIP International Information Security Conference* (pp. 55-68). Springer, Berlin, Heidelberg.

Meyer, A. and Rose, D.H. (2000). Universal Design for Individual Differences. *Educational Leadership*, *58*, 39-43.

Miles, C., & Hardman, E. (1998). State-dependent memory produced by aerobic exercise. *Ergonomics, 41*, 20-28.

Mishra, J., & Backlin, W. M. (2007). The effects of altering environmental and instrumental context on the performance of memorized music. *Psychology of Music*, *35,* 453-472.

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 882-894.

Nagge, J. W. (1935). An experimental test of the theory of associative interference. *Journal of Experimental Psychology*, *18*, 663-682.

Naveh-Benjamin, M., & Craik, F. I. (1995). Memory for context and its use in item memory: Comparisons of younger and older persons. *Psychology and Aging*, *10*, 284-293.

Neill, W. T., Lissner, L. S., & Beck, J. L. (1990). Negative priming insame-different matching: Further evidence for a central locus of inhibition. *Perception & Psychophysics*, *48,* 398-400.

Neisser, U., & Kerr, N. (1973). Spatial and mnemonic properties of visual images. *Cognitive Psychology*, *5,* 138-150.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402-407.

Nicholson, J. (2009). Design of a Multi-Touch shoulder surfing resilient graphical password. *B. Sc in Information Systems, Newcastle University, Newcastle*.

Nicholson, J., Coventry, L. and Briggs, P. (2013). Age-related performance issues for PIN and face-based authentication systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 323-332.

Olalere, A. and Lazar, J. (2011). Accessibility of US federal government home pages: Section 508 compliance and site accessibility statements. *Government Information Quarterly, 28,* 303-309.

Paans, R., & Herschberg, I. S. (1987). Computer security: The long road ahead. *Computers & Security*, *6*, 403-416.

Paivio, A. (1979). *Imagery and Verbal Processes*. London, Ontario: Psychology Press.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19,* 309-328.

Pering, T., Sundar, M., Light, J., & Want, R. (2003). Photographic authentication through untrusted terminals. *IEEE Pervasive Computing*, 30-36.

Perkovic, T., Cagalj, M., & Rakic, N. (2009). SSSL: shoulder surfing safe login. In *17th International Conference on Software, Telecommunications & Computer Networks, 2009* (pp. 270-275).

Petersen, R. C. (1974). Imagery and cued recall: Concreteness or context? *Journal of experimental psychology*, *102,* 841-844.

Petrich, J. A., & Chiesi, H. L. (1976). The locus of color-context changes, encoding instructions, and their effect on retroactive inhibition. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 190-199.

Pointer, S. C., & Bond, N. W. (1998). Context-dependent memory: Colour versus odour. *Chemical Senses*, *23*, 359-362.

Qian, C., Song, X., Huang, Y., & Lai, X. (2013). A graphical password scheme against snapshot remote monitoring and shoulder-surfing with its application in one-time password. *2013 International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013)* (pp. 608-615).

Rajavat, R., Gala, B., & Redekar, A. (2015). Textual and graphical password authentication scheme resistant to shoulder surfing. *International Journal of Computer Applications, 114*, 26-30.

Rand, G., & Wapner, S. (1967). Postural status as a factor in memory. *Journal of Verbal Learning and Verbal Behavior*, *6*, 268-271.

Rao, K., & Yalamanchili, S. (2012). Novel shoulder-surfing resistant authentication schemes using text-graphical passwords. *International Journal of Information and Network Security*, *1*, 163-170.

RealUser, "www.realuser.com," last accessed in June 2005.

Robbins, D., Bray, J. F., Irvin, J. R., & Wise, P. S. (1974). Memorial strategy and imagery: An interaction between instructions and rated imagery. *Journal of Experimental Psychology, 102*, 706-709.

Rokade, A. H., Hasan, Z. U., & Mahajan, S. A. (2014). User authentication by secured graphical password implementation. *International Journal of Innovative Research in Science & Engineering (IJIRSE),* 1-8.

Sahakyan, L. (2010). Environmental context change affects memory for performed actions. *Quarterly Journal of Experimental Psychology*, *63*, 425-433.

Sasamoto, H., Christin, N., & Hayashi, E. (2008). Undercover: Authentication usable in front of prying eyes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 183-192.

Saufley, W. H., Otaka, S. R., & Bavaresco, J. L. (1985). Context effects: Classroom tests and context independence. *Memory & Cognition*, *13*, 522-528.

Schab, F. R. (1990). Odors and the remembrance of things past. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 648-655.

Schaub, F., Walch, M., Könings, B., & Weber, M. (2013). Exploring the design space of graphical passwords on smartphones. *Proceedings of the Ninth Symposium on Usable Privacy and security*, 11-31.

Sheffert, S. M., & Fowler, C. A. (1995). The effects of voice and visible speaker change on memory for spoken words. *Journal of Memory and Language*, *34*, 665-685.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of verbal Learning and Verbal Behavior*, *6*, 156-163.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 460-471.

Smith, S. M. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory & Cognition*, *12*, 477-482.

Smith, S. M. (1985). Background music and context-dependent memory. *The American Journal of Psychology*, 591-603.

Smith, S. M. (1986). Environmental context-dependent recognition memory using a short-term memory task for input. *Memory & Cognition*, *14*, 347-354.

Smith, S. M. (1994). Theoretical principles of context-dependent memory. *Theoretical aspects of memory*, *2*, 168-195.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342-353.

Smith, S. M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology*, *6*, 125-139.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203-220.

Sreelatha, M., Shashi, M., Anirudh, M., Ahamer, M. S., & Kumar, V. M. (2011). Authentication schemes for session passwords using color and images. *International Journal of Network Security & Its Applications*, *3*, 111-119.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology*, *25*, 207-222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*, 73-74.

Stefanucci, J. K., O'Hargan, S. P., & Proffitt, D. R. (2007). Augmenting context-dependent memory. *Journal of Cognitive Engineering and Decision Making*, *1*, 391-404.

Still, J. D., & Bell, J. (2018). Incognito: Shoulder-surfing resistant selection method. *Journal of Information Security and Applications*, *40*, 1-8.

Still, J. D., Cain, A., & Schuster, D. (2017). Human-centered authentication guidelines. *Information & Computer Security*, *25*, 437-453.

Strand, B. Z. (1970). Change of context and retroactive inhibition. *Journal of Verbal Learning* *and Verbal Behavior*, *9*, 202-206.

Sun, H. M., Chen, S. T., Yeh, J. H., & Cheng, C. Y. (2016). A shoulder surfing resistant graphical authentication system. *IEEE Transactions on Dependable and Secure Computing*, 1-14.

Suo, X., Zhu, Y., & Owen, G. S. (2005). Graphical passwords: A survey. In *Computer security applications conference, 21st annual* (pp. 10-20).

Tao, H. (2006). *Pass-Go, a new graphical password scheme* (Doctoral dissertation, University of Ottawa (Canada)).

Tipper, S. P., MacQueen, G. M., & Brehaut, J. C. (1988). Negative priming between response modalities: Evidence for the central locus of inhibition in selective attention. *Perception & Psychophysics*, *43*, 45-52.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of experimental psychology*, *77*, 593-601.

Vachaspati, P. S. V., Chakravarthy, A. S. N., & Avadhani, P. S. (2013). A Novel Soft Computing Authentication Scheme for Textual and Graphical Passwords. *International Journal of Computer Applications*, *71*, 42-54.

Van Oorschot, P. C., & Wan, T. (2009). TwoStep: An authentication method combining text and graphical passwords. *International Conference on E-Technologies*, (pp. 233-239).

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and aging*, *11*, 164-178.

Voss, J. L. (2009). Long-term associative memory capacity in man. *Psychonomic Bulletin & Review*, *16*, 1076-1081.

Waddel, K. (2017). Can cops force you to unlock your phone with your face? *The Atlantic*. Retrieved from https://www.theatlantic.com/technology/archive/2017/09/can-cops-force-you-to-unlock-your-phone-with-your-face/539694/

Walters, R. (2014). Cyber attacks on U.S. companies since November 2014. *The Heritage Foundation.* Retrieved from http://www.heritage.org/cybersecurity/report/cyber-attacks-us-companies-november-2014

Walters, R. (2014). Cyber attacks on U.S. companies since November 2014. *The Heritage Foundation.* Retrieved from http://www.heritage.org/cybersecurity/report/cyber-attacks-us-companies-november-2014

Wash, R., Rader, E., Berman, R., & Wellmer, Z. (2016). Understanding password choices: How frequently entered passwords are re-used across websites. *Symposium on Usable Privacy and Security (SOUPS)* (pp. 175-190).

Watkins, M. J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behavior*, *15*, 505-517.

Wickens, C., Tuber, D. S., & Wickens, D. D. (1983). Memory for the conditioned response: The proactive effect of preexposure to potential conditioning stimuli and context change. *Journal of Experimental Psychology: General*, *112*, 41-57.

Wiedenbeck, S., Waters, J., Birget, J. C., Brodskiy, A., & Memon, N. (2005). PassPoints: Design and longitudinal evaluation of a graphical password system. *International journal of human-computer studies, 63*, 102-127.

Wiedenbeck, S., Waters, J., Sobrado, L., & Birget, J. C. (2006). Design and evaluation of a shoulder-surfing resistant graphical password scheme. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 177-184.

Winograd, E., & Lynn, D. S. (1979). Role of contextual imagery in associative recall. *Memory & Cognition*, *7*, 29-34.

Winograd, E., & Rivers-Bulkeley, N. T. (1977). Effects of changing context on remembering faces. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 397-405.

Yakovlev, V. A., & Arkhipov, V. V. (2015). User authentication based on the chess graphical password scheme resistant to shoulder surfing. *Automatic Control and Computer Sciences*, *49*, 803-812.

Zakaria, N. H., Griffiths, D., Brostoff, S., & Yan, J. (2011). Shoulder surfing defense for recall-based graphical passwords. *Proceedings of the Seventh Symposium on Usable Privacy and Security*, 6-18.

Zangooei, T., Mansoori, M., & Welch, I. (2012). A hybrid recognition and recall based approach in graphical passwords. *Proceedings of the 24th Australian Computer-Human Interaction Conference*, 665-673.

Zhao, H., & Li, X. (2007). S3PAS: A scalable shoulder-surfing resistant textual-graphical password authentication scheme. In *21st International Conference on Advanced Information Networking and Applications Workshops, 2007, AINAW'07*, 2, 467-472.

Zhao, J., & Turk-Browne, N. B. (2011). Incidental encoding of numerosity in visual long-term memory. *Visual Cognition, 19*, 928-955.

Zviran, M., & Haga, W. J. (1990). Cognitive passwords: The key to easy access control. *Computers & Security, 9,* 723-736.

Zviran, M., & Haga, W. J. (1999). Password security: An empirical study. *Journal of Management Information Systems*, *15*, 161-185.

VITAE

Ashley Allison Cain

250 Mills Godwin Life Sciences Bldg

Norfolk, VA 23529

Ashley is a Ph.D. student in human factors psychology at Old Dominion University where she focuses on the human side of cybersecurity. She received her master’s degree from San Jose State University in experimental and research psychology in 2015.