Exploring the Memorability of Multiple Recognition-Based Graphical Passwords and Their Resistance to Guessability Attacks

Soumyadeb Chowdhury

Submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy

School of Computing Science
College of Science and Engineering
University of Glasgow

March 2015

Declaration

I declare that this thesis was composed by me and that the work contained therein is my own, except where explicitly stated otherwise in the text.

Soumyadeb Chowdhury

The thesis is dedicated to my grandparents Mrs Roma Mitra, Mr B.K. Mitra, Mrs Deepti Chowdhury and Mr S.K. Chowdhury.

Abstract

Most users find it difficult to remember traditional text-based passwords. In order to cope with multiple passwords, users tend to adopt unsafe mechanisms like writing down the passwords or sharing them with others. Recognition-based graphical authentication systems (RBGSs) have been proposed as one potential solution to minimize the above problems. But, most prior works in the field of RBGSs make the unrealistic assumption of studying a single password. It is also an untested assumption that RBGS passwords are resistant to being written down or verbally communicated.

The main aim of the research reported in this thesis is to examine the memorability of multiple image passwords and their guessability using written descriptions (provided by the respective account holders). In this context, the thesis presents four user studies. The first user study (US1) examined the usability of multiple RBGS passwords with four different image types: Mikon, doodle, art and everyday objects (e.g. images of food, buildings, sports etc.). The results obtained in US1 demonstrated that subjects found it difficult to remember four RBGS passwords (of the same image type) and the memorability of the passwords deteriorated over time. The results of another usability study (US2) conducted using the same four image types (as in US1) demonstrated that the memorability of the multiple RBGS passwords created by employing a mnemonic strategy do not improve even when compared to the existing multiple password studies and US1. In the context of the guessability, a user study (GS1) examined the guessability of RBGS passwords (created in US1), using the textual descriptions given by the respective account holders. Another study (GS2) examined the guessability of RBGS passwords (created in US2), using descriptions given by the respective account holders. The results obtained from both the studies showed that RBGS passwords can be guessed using the password descriptions in the experimental set-up used.

Additionally, this thesis presents a novel Passhint authentication system (PHAS).The results of a usability study (US3) demonstrated that the memorability of multiple PHAS passwords is better than in existing Graphical authentication systems (GASs). Although the registration time is high, authentication time for the successful attempts is either equivalent to or less than the time reported for previous GASs. The guessability study (GS3) showed that the art passwords are the least guessable, followed by Mikon, doodle and objects in that order. This thesis offers these initial studies as a proof of principle to conduct large scale field studies in the future with PHAS. Based on the review of the existing literature, this thesis identifies the need for a general set of principles to design usability experiments that would allow systematic evaluation and comparison of different authentication systems.

From the empirical studies (US1, US2 and US3) reported in this thesis, we found that multiple RBGS passwords are difficult to remember, and the memorability of such passwords can be increased using the novel PHAS. We also recommend using the art images as the passwords in PHAS, because they are found to be the least guessable using the written descriptions in the empirical studies (GS1, GS2 and GS3) reported in this thesis.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Dr. Ron Poet who gave me the opportunity to conduct this research under his supervision. I am thankful for his advice and support. He gave me the freedom and the flexibility to explore different avenues, and has always encouraged pursuing the research endeavors. I also extend my gratitude to Dr. Lewis Mackenzie for accepting to supervise me. I am grateful for his suggestions which had indeed helped me to improvise on my work. I would like to thank Dr. Lewis for all the interesting and challenging discussions, which helped to shape this research in a coherent way. This thesis would have not been possible without the guidance of my supervisors. I would also like to extend my thanks to Dr. Karen Renaud for sharing the doodle and Mikon image collections (and her numerous publications in the field), which also contributed to the success of this research immensely. I also appreciate the comments and suggestions made by Professor Steve Furnell (Plymouth University) and Dr. Iadh Ounis, which helped to improve the structure and readability of the thesis. I extend my thanks to Dr. David Manlove for his advice, efforts and patience, while convening my PhD examination.

I am grateful to Scottish Informatics and Computer Science Alliance (SICSA) for funding my PhD research and thus making my dream come true. I would like to extend my thanks to the support team and all the staff members in the School of Computing Science, for providing me with all the necessary help throughout my PhD studies. Special thanks to my colleagues Niaz Morshed Chowdhury, Md. Sadek Ferdous and their respective families for their enormous, love, support and care. I extend my gratitude to Ms. Heather Lambie (Adviser of Studies) for promptly helping out with all the administrative issues. Many thanks, to Ms. Katherine Henderson for her encouraging and motivating words. Thanks to all the College of Science and Engineering staff (especially, Mrs. Helen Border) for their cooperation and help.

I extend my gratitude to Dr. Balvinder Shukla, Vice Chancellor Amity University, India, for helping me gather the participants to conduct some of the user studies reported in this thesis. The research produced in this thesis would have not been possible without her co-operation. I would like to thank all the participants who gave their time and efforts to take part in my experiments. Their cooperation and feedback were keys to the success of this research.

Finally and most importantly I offer my deepest gratefulness and owe a large amount of credit to my family for their unconditional love, support, friendly encouragement and understanding. I am thankful to my father for all his suggestions and pep talks throughout the PhD journey. I am thankful to my mother for always being there for me, 24/7, despite the distance and supporting me throughout the ups and downs psychologically. I would like to take the opportunity to thank Sepideh, for her continuous love, support, encouragement and understanding.

Finally, I owe gratitude to my grandparents for their constant love, support and encouragement throughout my life. I cannot thank them enough.

Table of Contents

List of Tables

List of Figures

Appendix Figures

Abbreviations

- GAS: Graphical Authentication System
- GS1: Guessability Study reported in Chapter 4 (textual descriptions only)
- GS2: Guessability Study reported in Chapter 5 (textual descriptions + sketches)
- GS3: Guessability Study reported in Chapter 6 (PHAS)
- PHAS: Passhint Authentication System
- RBGS: Recognition-Based Graphical Authentication System
- Sec: Seconds
- SD: Standard Deviation
- SE: Standard Error
- US1: Usability Study reported in Chapter 3
- US2: Usability Study of story passwords reported in Chapter 5
- US3: Usability Study of PHAS reported in Chapter 6

Chapter 1
Introduction

This chapter introduces the context and motivation of this thesis. The chapter also presents a brief overview of authentication mechanisms, followed by the thesis statement and research objectives. The chapter concludes by presenting the structure of the thesis and the main contributions of the research reported in this thesis.

1.1 Context

Human-Computer Interaction and Security, also referred to as *Usable Security*, is a relatively new area in the field of Computing Science combining: *Human-Computer Interaction (HCI)* and *Computer Security*. Human factors are often considered as the *"weakest link"* in computer security systems (Sasse et al., 2001). In this context, existing studies reported in (Adams & Sasse, 1999; Florencio & Herley, 2007) have shown that since users can remember only a limited number of passwords, they tend to write them down or use the same passwords for different accounts. Such practices would compromise the security of the authentication system. Hence, the area of Usable Security was identified by the HCI and security practitioners to improve the usability of the secure systems. In this context, Patrick et al. (2003) had also pointed out authentication, security operations and developing secure systems, as the three major areas, where HCI is important.

According to Hewett et al. (1996), *"HCI is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and the study of major phenomena surrounding them"*. According to Ross (1999), *"Computer Security is a discipline concerned with the ability of a system to protect information and system resources with respect to confidentiality and integrity"*. This is often associated with: confidentiality; integrity; authentication; non repudiation; availability; access control and privacy. Hence, Usable Security focuses on various human factors in computer security, which primarily includes the impact of user behaviour on the security of a system and the effect of interaction design of a secured system on the users. In this context, Carnor & Garfinkel (2005) have described the aim of Usable Security as *"designing secure systems that people can use."*

This thesis focuses on one particular aspect of security, namely *user authentication*. The most common form of authentication uses traditional text-based passwords, which are plagued with a number of usability and security problems – for example, the increase in the number of such passwords typically makes it difficult to remember them and users tend to employ unsafe coping strategies which make them insecure to use (Florencio & Herley, 2007). When users resort to unsafe coping strategies to aid memorability of such passwords, the decrease in security cannot be simply addressed by strengthening, in isolation, the underlying technical security of the system. In the view of such problems, alternative authentication mechanisms such as biometrics (Coventry, 2005) and token-based authentication have been recommended. However, some of the attractive characteristics of the traditional passwords over alternative mechanisms, which make them popular among the service providers and users, are: low cost to deploy compared to the aforementioned alternatives, which will incur additional expenses; avoiding privacy issues incurred by the use of biometrics; advantage of being portable, i.e. the users do not need to carry physical tokens. In the context of improving the memorability of multiple passwords, and dissuading users from unsafe coping practices, graphical passwords have been proposed as a possible alternative to the traditional text passwords in the recent years (Biddle at al., 2009). However, as we will discuss further in Sections 2.5 and 2.6, most of the prior work in this area have focused on the unrealistic context of remembering a single graphical password, and have not examined the guessability of such passwords using descriptions written by the respective account holders. In this thesis, we advance the research in the field of *Graphical Authentication Systems (GASs),* through:

- usability evaluation, to explore the memorability of employing multiple passwords in Recognition-Based Graphical Authentication System (RBGS);
- security evaluation, to explore the guessability of RBGS passwords to written descriptions;
- presenting a novel authentication system that could improve the memorability of multiple graphical passwords and offer adequate security in an appropriate setting

1.2 Motivation

The most widely used authentication codes, such as text passwords and personal identity numbers (PINs) to control the access to resources (e.g. websites, bank accounts, mobile devices), are plagued with various usability and security problems. One major drawback with

text passwords and PINs is that users find it difficult to remember increasing numbers of authentication codes (Klein, 1990; Sasse et al., 2001). Therefore, the users are often faced with a choice between forgetting their passwords, which can be frustrating and inconvenient, or employing various coping strategies such as writing down the passwords, reusing them or sharing them with known associates, which compromises the security of the system (Adams & Sasse, 1999; Herley et al., 2009; Sasse et al., 2001). Various technical solutions such as imposing password polices, encryption and communication algorithms to protect the passwords have not resolved the primary problem, which is related to the human factors in authentication, of which the most important is the memorability of the multiple passwords.

In the recent years, GASs which use images as the password have been proposed as an alternative to text passwords, due to their potential to improve memorability. The motivating idea is that humans can supposedly remember images better than recalling alphanumeric text (Paivio, 1986), so this may be a way of devising more memorable passwords. However, as we will discuss in Chapter 2, most studies with graphical password systems, especially RBGSs (focus of this thesis) have focused on the unrealistic usability example of a single password. In the last fifteen years (to our knowledge), only four studies- Moncur & Leplatre (2007), Chiasson et al. (2009), Everitt et al. (2009) and Hlywa al. (2009), have explored the memorability of multiple graphical passwords. In general, user studies exploring the usability of multiple RBGS passwords are sparse, which is currently a limitation in the field. Moreover, as we will discuss in Section 2.6, multiple password studies suffer from a high drop-out rate and hence fail to provide concrete evidence, whether RBGSs in their current form are able to solve the issue of remembering multiple passwords. This is reflected in the thesis statement presented in Section 1.4.

The literature reported in Adams & Sasse (1999) highlights that the sharing and recording of text passwords has become an indispensable coping technique to remember multiple credentials. However, graphical passwords are assumed to be particularly resistant to being written down or verbally communicated. For example, Real (2004) have claimed that "*faces when used as RBGS password can't be written down or copied and can't be given to another person"*. It is an unchallenged assumption that users will find it difficult to record or share their graphical passwords. Currently, there is no known methodology to measure the extent to which users can record/share their graphical passwords, and the strategy that users will adopt to record/share the passwords in real life. However, the real question that needs to be

examined is, whether it is possible to guess graphical passwords, using any sort of revelation produced by the legitimate user. The guessability aspect is also reflected in the thesis statement in discussed Section 1.4.

Since the context of this thesis is *Usable Security*, the research focusses on both the memorability of multiple RBGS passwords and the guessability of such passwords using descriptions of the images forming the password provided by the respective account holders.

1.3 Authentication

Information security systems must permit only legitimate users to gain access to the system and use its resources. This is done by a two-step process: *identification* indicates the account that the potential user wishes to access, while *authentication* establishes whether the user has the right to access that account. In computer security mechanisms, users are often required to authenticate themselves by using a secret known as a *password or authenticator*. The authentication mechanisms can be classified into the following categories, based upon the model proposed in Renaud (2005).

- *Something you know (recall)*: examples include, passwords and PINs (Personal Identification Numbers);
- *Something you recognize*: examples include, images or a specific location on an image;
- *Something you are (static biometrics)*: examples include, fingerprints, facial/iris scans;
- *Something you do (behavioral biometrics)*: examples include, keystroke dynamics, handwritten signatures;
- *Something you have (tokens):* examples include, smartcards (a card with embedded microprocessor chip);
- *Where you are (location based authentication)*: examples include, approved locations – identifying city or county of origin.

An addition to this model is *someone you know*, which was reported in Brainard et al. (2006).

1.3.1 Traditional Text-Based Passwords

For the purpose of this thesis, a *'traditional text-based password'* is a password which consists of any combination of characters from the ASCII set. These passwords are also referred to as '*alphanumeric passwords' or 'text passwords'*. Text passwords remain the

most widely used authentication mechanism, despite the large number of available options, for many reasons reported in Herley et al. (2009). They are also inexpensive as well as easy to implement, and most users are familiar with them. Users can select text passwords that do not contain any personal information to authenticate themselves without violating their privacy, unlike biometric systems (Jain et al., 2000; Coventry, 2005). Text passwords are also portable, i.e. users need to simply recall them, as opposed to tokens which must be carried. The research reported in the literature had investigated various issues related to the *text passwords,* which are discussed below briefly.

- Klein (1990) reported the seminal work on the password behavior of the users by collecting the UNIX password files of 15000 users. The experiment showed that the users selected common English words as their password, which made it easier to guess the password using dictionary attacks.

- The responses collected in the questionnaire study reported in Adams & Sasse (1999) provided evidence that users cope with increasing number of passwords by reusing the same passwords over the multiple accounts, writing down the passwords and sharing the passwords. The aforementioned coping strategies potentially compromise the security of the system, since it becomes easy to guess and capture the respective passwords.

- A two-part online study was reported in Komanduri et al. (2011) to explore the effect of various password composition policies on the user behavior. The results obtained from the study demonstrated that 31% of the participants wrote down the passwords and 11.1% forgot the passwords. The results also highlighted that 34.6% of the 5000 participants who took part in the study admitted to password reuse, and 17.7% admitted to modified reuse, i.e. manipulating a previous password by addition of numbers or special characters. The results reported in (Komanduri et al., 2011) also provided evidence of the same types of password reuse as in Adams & Sasse (1999). Inglesant & Sasse (1999) also gathered evidence using password diaries and interviews of users writing down their passwords. The study was conducted with employees in a University.

- Dhamija & Perrig (2000) conducted an interview with 30 participants and found evidence of password re-use. The authors also found that users had 10-50 accounts of various forms that required authentication and users had one to seven unique passwords. A survey of 218 students was reported in Brown et al. (2004) regarding the password habits of users. The results reported that the mean number of password systems used by the students was 8.18 (SD of 2.18 and range of 3-20), and the mean number of unique passwords was 4.45 (SD of 1.63 and a range of 1-11). Gaw & Felten (2006) also reported a similar study and provided evidence of a high number of passwords per user and password re-use.

- Notoatmodjo & Thomborson (2009) reported a study with 26 university students that examined how users mentally group their passwords for various accounts that require authentication. The participants were required to describe their passwords using the length, perceived security level and difficulty of recall. 253 accounts were classified as low importance: 32% were unique passwords issued by the system and 68% were re-used passwords. The results demonstrated that 68 passwords were classified as highly important; of which 63% were system assigned unique passwords and 37% were re-used. The authors also found that an increase in the number of accounts requiring user authentication led to password re-use, which would aid memorability of the passwords.

The literature on text passwords provides evidence for a number of password coping mechanisms employed by the users to aid memorability of multiple such passwords. There is a variation in the quantitative values reported in these studies due to different sample size, as well as variation in the user study frameworks.

1.3.2 Graphical Passwords

The weaknesses of text passwords, as discussed in the previous section, have led to the exploration of alternative authentication mechanisms. One viable alternative, which has been researched extensively in the last fifteen years, is *Graphical Authentication Systems (GASs)*. GASs have been categorised in many ways (Biddle et al., 2009). However, in this thesis we focus on the categorisation, which is based upon the type of memory task involved in remembering the password, as discussed below.

- *Recall-based:* users must draw an image either on a grid or canvas during the password creation stage, and they have to re-draw that same image during authentication. A canonical example is Draw-a-Secret, which is reported in Jermyn et al. (1999). However, recall is considered to be the least accurate type of memories because the accuracy would decay after a considerable amount of time, if the password is not used frequently (Baddeley, 1997).

- *Cued recall-based:* specific points on an image that is either selected by the users or issued by the system, form the password. An archetypal example of such a system is Passpoints (Wiedenbeck et al., 2005a). In this kind of systems, the authentication system provides a cue (image) to help the users remember their password (points on the image). This feature is intended to reduce the memory load on the user and is considered an easier task, compared to unaided recall (Parkin, 1993).

- *Recognition-based*: users can either choose their password images from a collection presented by the system, or the passwords are issued by the system to the users. The users can also provide their own images. During authentication, users must recognize the correct password image among a collection of decoys. A canonical example is Passfaces, which has also been commercially deployed (Real, 2004). The motivating idea is humans have a vast memory for images. Mandler & Ritchey (1977) suggests that, human beings have an exceptional ability to recognize images that they have previously seen, even if the image has been viewed for a very short period of time.

1.3.3 Graphical Authentication Category Studied

The lack of usability and security studies reported in the literature fails to demonstrate the potential of recall-based GASs, which will be further discussed in Section 2.2. We also find that these systems show an inferior performance even when compared to the other GAS categories (Section 2.5). Many studies have evaluated the cued recall-based systems and demonstrated their potential as an alternative to traditional text passwords. However, prior work reported in Renaud & Angeli (2004) has pointed out that it might be difficult to find images that may have many memorable locations (click-points), which might eventually pose a problem for any future deployment. In the context of the RBGSs, image types, except faces

(Everitt et al., 2009; Hlywa et al., 2011) and objects (Moncur & Leplatre, 2007), have not been evaluated upon the usability context of the users having multiple graphical passwords to remember. Moreover, RBGS passwords are assumed to be particularly resistant to being written down, or verbally communicated (Dhamija & Perrig 2000; Real, 2004). Hence we advanced our research in the area of RBGSs, to examine their usability in the context of multiple password use and evaluate their guessability using written descriptions.

This research focuses on recognition-based graphical authentication systems (RBGSs), and details regarding the configuration of the system will be further discussed in Section 2.7. During the registration process a user selects, *n*, number of *target images* from the collection (image archive) presented by the system, to form a single password. Each authentication session is usually an *n*- step process. At each step, the user is presented with a grid of images containing, at least one of the target images and a number of other images called *distractor/decoy* images. The grid consisting of a target image and decoy images is called a *challenge set* as shown in Figure 1.1. A legitimate user must recognise and select the target image in each challenge set to authenticate successfully.

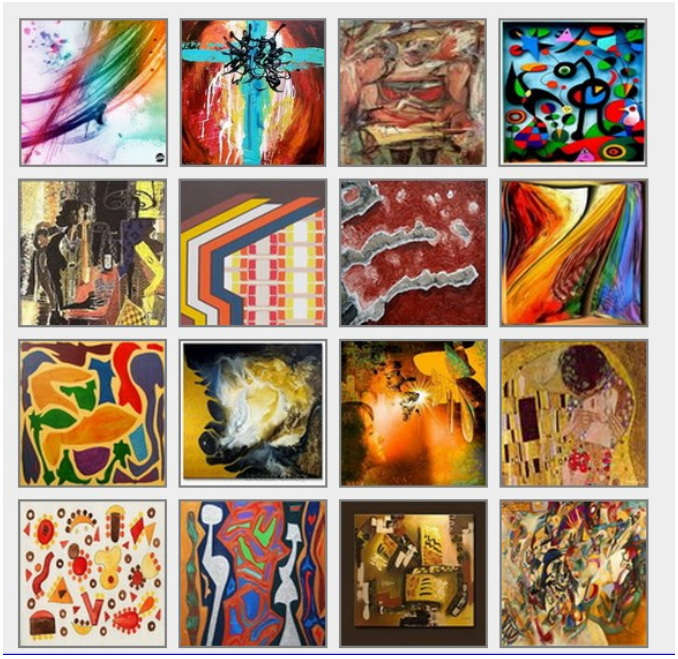


Figure 1.1: A challenge set comprising of 16 art images: 15 decoys + 1 target

The literature in the context of the usability and security of all the three categories of GASs presented in this sub section will be further discussed in Chapter 2.

1.3.4 Lack of Benchmarks

The traditional text passwords are the most common form of authentication. Hence it can be argued that these should be used as a benchmark to evaluate the usability and security of the alternative authentication systems. However, such comparisons will be biased by the years of experience, familiarity, good understanding of the registration and login processes and the range of strategies developed to cope with increasing number of such passwords, which a typical user now possesses. Moreover, the usable security community also lacks comprehensive and definitive results in the context of the effectiveness and efficiency of text passwords, which makes it even more difficult to use them as benchmarks (Biddle et al., 2009). Additionally, it might be also challenging to abide by the ethical regulations and provide definitive results in the context of the effectiveness and efficiency of text passwords in an ecologically valid setting: i.e. one which simulates a real-life scenario in a number of different contexts, such as using the passwords in real-time to access bank accounts, social networks (for e.g. Facebook, Twitter), online shopping services (for e.g. Amazon, ebay), online money transactions (PayPal, Western Union). In this thesis, we examine RBGSs as an alternative to the traditional text passwords and do not explicitly state the former as a replacement. Hence, we do not use text passwords as a baseline to compare the results reported in the existing literature to the results obtained from the empirical studies reported in this thesis.

Complicating matters further, existing research studies in the area of graphical passwords currently lack consistency in the sense that each of the single as well as multiple password studies have been conducted using a number of different experimental frameworks and authentication system design, which measure multiple dependant variables that have been interpreted differently in each study. This has been also pointed out in Biddle et al. (2009), which presents a review of all the graphical password studies in the last decade. However, we believe that the lack of rigor in evaluating GASs, as well as the lack of suitable benchmarks to define an acceptable baseline performance can be dealt with in this thesis by comparing the results obtained in each of our multiple graphical password studies and guessability studies, to the literature that has reported similar studies.

1.4 Thesis Statement and Research Objectives

The thesis statement is:

Multiple image passwords are memorable, and cannot be guessed using a description of the target images forming the password, given the current state-of-the-art in recognition-based graphical authentication systems (RBGSs).

The research reported in this thesis is refined into six stages as given below. Moreover, the thesis statement is further divided into four objectives, each of which is addressed separately in the thesis.

- *Stage 1*- Reviewing all the existing work in the field of GASs, especially studies that have examined the usability of multiple RBGS passwords and vulnerability of such passwords to descriptions. This stage will help us to gather the statistics reported in the existing studies, which will be used to compare the results obtained from each empirical study reported in this thesis.

- *Stage 2* – In this stage, we aim to address the first objective related to the thesis statement, in the context of the memorability of multiple RBGS passwords.
    - *Objective 1*: Designing and conducting a usability study (US1) using a suitable experimental protocol with a sample from the student population to compare the effectiveness as well as efficiency of the different image types, when multiple passwords of each type are used in a RBGS, and compare the results obtained in US1 with the other multiple graphical password studies. This objective will address the research question (RQ1): *Whether multiple RBGS passwords in the current state-of-the-art are memorable, in a given experimental setting?*

      To explicitly answer RQ1, we will compare the effectiveness results obtained in US1 to all the multiple graphical password studies reported in the literature (Section 2.6).

We conducted each user study reported in this thesis with a distinct sample recruited from the student population because of the following reasons:

- It is easier to recruit and get access to a large number of students, compared to the general population, for the kind of longitudinal studies reported in this thesis;
- It is also easy to control students, i.e. provide instructions to them for the purpose of the hybrid (both lab-based and web-based) studies reported in this thesis;
- Moreover, most multiple graphical password studies and a single study that has evaluated the topic of descriptions in the context of RBGS, recruited a sample from the student population.

We acknowledge that general population is much diverse, but the sample population used in our studies are sufficient to evaluate the various aspects (memorability of multiple RBGS passwords and their guessability to descriptions) reported in this thesis. Moreover, the results obtained in each study could justify more extensive field studies using the best performers (i.e. image types) in the future.

- *Stage 3* – We aim to address the second objective related to the thesis statement, in the context of the guessability of RBGS passwords using the written descriptions.
    - *Objective 2:* Designing and conducting a guessability study (GS1) using a suitable experimental protocol with a sample from the student population, to examine the vulnerability of the passwords created in US1 using the corresponding textual descriptions, written by the subjects (respective account holders) who took part in US1. This objective will address the research question (RQ2): *Whether RBGS passwords can be guessed using their corresponding textual descriptions, provided by the respective account holders, in a given experimental setting?*

      To explicitly answer RQ2, we will compare the results obtained in GS1 to the only empirical study (Dunphy et al., 2008) that has also examined the topic of RBGS password descriptions.

- *Stage 4* – We aim to address the third objective related to the thesis statement, in the context of remembering multiple RBGS passwords (story passwords), each created using a mnemonic strategy.
    - *Objective* 3: Designing and conducting a usability study (US2) using a suitable experimental protocol with a sample from the student population to compare the effectiveness and efficiency of the multiple RBGS story passwords (using

the same image types as in US1), i.e. when a mnemonic strategy is employed to choose the target images forming each of the passwords. This objective will address the research question (RQ3): *Whether the memorability of multiple RBGS passwords improves by employing a mnemonic strategy, to choose the passwords during the password registration stage, in a given experimental setting?*

To explicitly answer RQ3, we will compare the effectiveness results obtained in US2 to US1, and all the multiple graphical password studies reported in the literature (Section 2.6).

- *Stage 5*- We aim to address the final objective related to the thesis statement, in the context of guessing the RBGS story passwords using written descriptions.
    - *Objective 4:* Designing and conducting a guessability study (GS2) using a suitable experimental protocol with a sample from the student population, to examine the vulnerability of RBGS passwords (created in US2) using the corresponding descriptions, given by the subjects (respective account holders) who took part in US2. This objective will address the research question (RQ4): *Whether RBGS passwords created by employing a mnemonic strategy are guessable, using their corresponding descriptions provided by the respective account holders, in a given experimental setting?*

      To explicitly answer RQ4, we will compare the results obtained in GS2 to GS1 and the only empirical study (Dunphy et al., 2008) that has examined the topic of RBGS password descriptions.

- *Stage 6* – Based on the results obtained in each of the empirical studies (US1, US2, GS1 and GS2), we propose a novel authentication system that could further improve the memorability of multiple image passwords. We will design and conduct a usability study (US3) and a guessability study (GS3) with a sample from the student population, using a suitable experimental protocol, to examine the performance of the proposed authentication system.

The thesis statement will be evaluated as given below. Figure 1.2 further illustrates the various objectives in the thesis statement.

- Objectives 1 and 3 will examine the claim regarding the memorability of multiple RBGS passwords.
- Objectives 2 and 4 will examine the claim regarding the vulnerability of RBGS passwords to written descriptions.

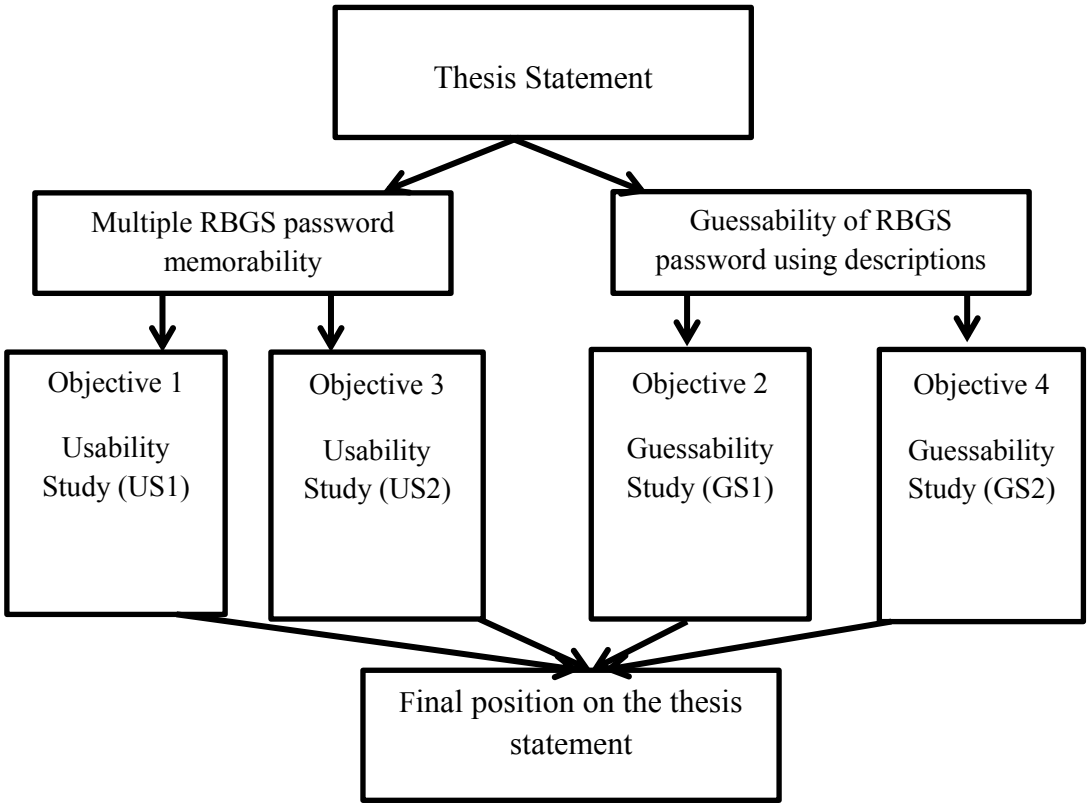


Figure 1.2: Evaluating the thesis statement

1.5 Research Approach

The methodology used in this thesis to answer the research questions and address the respective research objectives is primarily, empirical and quantitative. A technique called *human-subject experiments*, i.e. evaluating a system with human subjects by applying principles of experiment design (Miller, 1984; Field & Hole, 2003), which is very prominent in HCI community has been used to conduct the empirical evaluations.

The user studies reported in US1, GS1 and GS2 are web-based. There is no agreed-upon definition for web-based studies. According to Biddle et al. (2009), this term can be used for a scenario, where the experimenter has minimal face-to-face contact with the subjects taking part in the study. The reason for using web-based study is: (1) these studies can be conducted with large number of subjects; (2) subjects could be prompted several times to complete the tasks involved in the study; (3) the behaviour of the subjects may be more realistic than in a controlled lab settings; (4) it is easier to conduct the longitudinal studies over web.

In the web-based studies, it is necessary for the experimenter to keep track of the tasks and ensure that the experimental protocol is followed. In each of the web-based studies reported in this thesis, all the subjects were given clear instructions about the tasks, information about the system they will use, adequate time to get used to the system and email prompts reminding them about the tasks to be completed. The length of the study, especially in the case of US1 and US2 were chosen appropriately to capture the phenomena under study. In order to capture the user opinions, usability questionnaires were used. Questionnaires were the secondary source of gathering information, used in conjunction with system logs.

US2 is a hybrid study combining a lab-based training session followed by the tasks that are completed in the subject's regular environment. Hence, the study gained the advantage of both an initial controlled environment and increased the ecological validity of the task. US3 and GS3 are lab-based studies, which were conducted in a controlled environment to evaluate the success of the design decisions in isolation and usability problems in the new authentication system, before resources can be invested in large scale field studies. The idea was to ensure that subjects focus on the task assigned to them and statistical testing of the different measures could be done to assess the effectiveness of the new design decisions.

Statistical analysis is also used to assess, whether the differences in the data obtained from the user studies reflect actual differences between the experimental conditions (Mcguigan 1993; Field & Hole, 2003). The statistical tests, i.e. parametric or non-parametric, are chosen based upon the experimental design (independent measures or repeated measures), normality of the data obtained from the user studies and the type of the data (nominal, ordinal or categorical).

1.6 Structure of the Thesis

The remainder of the thesis is organised as follows:

- Chapter 2 presents a brief overview of some cognitive and psychological theories relevant to the thesis. The different categories of GASs are also reviewed. The chapter also addresses *Stage 1* by presenting a review of the literature, in relation to the usability of multiple graphical passwords and their vulnerability to descriptions;

- To address *Objective 1 (Stage 2)*, Chapter 3 presents a multiple password study (US1) to compare the usability of four different images types: Mikon, doodle, art and object, when used as RBGS passwords. The four different image types were chosen based on the review presented in Chapters 2 (Sections 2.2 to 2.6). The review suggested that none of the aforementioned image types except objects, were used in multiple password usability studies ;

- To address *Objective 2 (Stage 3)*, Chapter 4 presents a guessability study (GS1) to examine the vulnerability of RBGS passwords created in US1 using the corresponding textual descriptions, which are provided by the subjects (respective account holders) who took part in US1;

- Chapter 5 addresses both the *Objectives 3* and *4 (Stages 4 and 5)*, by reporting a usability study (US2) and a guessability study (GS2), to examine the performance of RBGS passwords created using a mnemonic strategy;

- Chapter 6 addresses *Stage 6* by presenting a novel authentication system. The chapter reports a usability study (US3) and a guessability study (GS3), to analyze the performance of the proposed authentication system. The performance of the proposed authentication system is also compared to the results reported in the existing literature discussed in Section 2.4 and 2.6.

- Finally, in Chapter 7 the thesis statement is revisited and all the research questions corresponding to each objective are answered. The final position on the thesis

statement is discussed, followed by further research directions that fall beyond the scope of this thesis.

1.7 Thesis Contributions

The research presented in this thesis contributes original ideas and knowledge in the field of RBGSs. The main contributions of the research are enumerated below.

Identifying the research problem

Chapter 2 identifies an important limitation in the field of GASs, i.e. most usability studies have focused on the unrealistic use of a single password. In this context, Section 2.6 identifies that in the last 15 years (to our knowledge) only four studies have explored the usability of multiple graphical passwords, and two of these studies had a high drop-out rate. The survey in Chapter 2 reveals that none of the existing studies have explored the vulnerability of RBGS passwords to descriptions, or any sort of revelation produced by an account holder, except (Dunphy et al., 2008). Hence, the thesis statement was established together with the research objectives systematically to explore the usability of multiple RBGS passwords, and their vulnerability to written descriptions.

Memorability of multiple RBGS passwords

The usability of multiple RBGS passwords has been examined in Chapter 3 with four distinct image types: Mikon, doodle, art and objects, over an online study (US1) conducted for a period of eight weeks. The results demonstrate that object images are most usable in the sense of being more memorable and less time-consuming to employ, Mikon images are close behind but doodle and art images are inferior. Another usability study (US2) is presented in Chapter 5, which examines the usability of multiple RBGS passwords when such passwords are created using a mnemonic strategy, using the same four image types as in US1. The results obtained in US2 follow the same trend as that of US1. However, the results obtained in both the studies provide concrete evidence that multiple RBGS passwords are difficult to remember, and time consuming to employ, given the current state-of-the art.

Guessability using descriptions

The vulnerability to third-party guessing of RBGS passwords, created in US1 and US2, using descriptions provided by the respective account holders is examined in Chapters 4 and 5 respectively. Both the studies show that most descriptions provided by the account holders were annotated/ non-annotated sketches of the target images forming the password. In the case of textual descriptions, these were denotative (i.e. described the elements in the image), which again helped in guessing the respective passwords. The results obtained from both the studies (GS1 and GS2) demonstrated that all the Mikon, doodle and object passwords were guessed, whereas 50% of the art passwords were guessed. It was difficult to guess art passwords using the textual descriptions and these passwords were the least amenable to sketching, compared to the three other image types. Hence these results provide evidence that art images are more resistant to being guessed using written descriptions, compared to the other image types.

Novel authentication system

A novel authentication mechanism, Passhint (PHAS), is proposed in Chapter 6. A prototype was created and two empirical lab-based studies (usability – US3 and guessability- GS3) were conducted. The results obtained from the multiple password usability study show that PHAS have memorability advantages, over other existing GASs. The results of the guessability study (GS3) with PHAS reveal that art passwords are the least guessable, followed by Mikon, doodle and objects in that order. The results strongly suggest that the use of art passwords in PHAS, would offer usable as well as secure authentication. This thesis offers the results of the initial usability and guessability studies as a proof of principle for the Passhint system.

1.8 Origins of the Material

A significant portion of the research presented in this thesis has been peer-reviewed and published in various academic venues. The author of this thesis is the primary author for all the publications, which are also based on the work presented in this thesis. Much of the text presented in the thesis is taken from these publications.

The peer- reviewed full- paper publications are:

- Chowdhury, S., Poet, R. & Mackenzie, L., 2013. A Comprehensive Study of the Usability of Multiple Graphical Passwords. *In the Proceedings of the Human-Computer Interaction – INTERACT 2013*., Springer Berlin Heidelberg.

- Chowdhury, S., Poet, R. & Mackenzie, L., 2014. Do Graphical Authentication Systems Solve the Password Memorability Problem? *In the Proceedings of the Human Aspects of Information Security, Privacy, and Trust - Second International Conference, HAS 2014*., Springer Berlin Heidelberg.

- Chowdhury, S., Poet, R. & Mackenzie, L., 2014. Passhint: Memorable and Secure Authentication. *In the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*., ACM Press.

- Chowdhury, S., Poet, R. & Mackenzie, L., 2014. A Study of Mnemonic Image Passwords. *In the Proceedings of the 12th International Conference on Privacy, Security and Trust (PST'14)*., IEEE.

- Chowdhury, S., Poet, R. & Mackenzie, L., 2014. Exploring the Guessability of Image Passwords Using Textual Descriptions. *In the Proceedings of the 7th International Conference on Security of Information and Networks (SIN'14)*., ACM Press.

The research reported in this thesis was selected to be presented as a poster at the *SET for BRITAIN* (2013) exhibition in the Engineering section held at the House of Commons.

The work reported in the following paper examined the guessability of image passwords using verbal descriptions. But this thesis is focused on written descriptions (textual or sketches). Hence the following work is not coherent with the thesis statement and therefore, has been excluded from this thesis. Please note that none of the guessability studies reported in this thesis are compared with the following publication and any comparisons made is beyond the scope of this thesis.

- Chowdhury, S., Poet, R. & Mackenzie, L., 2013. Exploring the Guessability of Image Passwords Using Verbal Descriptions. *In the Proceedings of the 12th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*., IEEE.

Chapter 2
Overview of Graphical Authentication Systems

In this chapter, a background on psychological research related to recognition-based graphical authentication is presented in the first section. The next three sections present a brief overview of the three different categories of graphical authentication systems (GASs) in the existing literature. A review of all the existing studies that have explored the memorability of multiple graphical passwords and the vulnerability of RBGS passwords to descriptions recorded by the respective account holders is also presented, followed by the scope of the thesis, and a conclusion to the chapter. The contents of Section 2.6 have been published in the proceedings of the 2nd International Conference, Human Aspects of Information Security, Privacy and Trust 2014, held as part of HCI International 2014.

2.1 Cognitive Theories

This section provides a background on psychological research related to the information processing in human memory, followed by the picture superiority effect, and guessability of images.

2.1.1 Information Processing in Human memory

The authentication systems require the users to remember their secrets, i.e. passwords. Hence it is essential to understand how a piece of information is processed in human memory, which is discussed below, and further illustrated in Figure 2.1. The discussion presented below is based upon the model proposed by Atinkson & Shiffrin (1968).

- *Sensory inputs:* The information from the outside world is gathered by the different sensory organs, and then it is stored in the sensory storage, which generally lasts for a very short period of time.
- *Short-term memory (STM):* If an individual is paying attention, then the information is transferred to the STM. STM holds the information as memory codes, i.e. mental representation of the selected parts of the information.
- *Long term memory (LTM):* The information is transferred from the STM to LTM, but only if it can be further processed and encoded. An elaborative encoding will take place, if the information can be associated with something meaningful.

- *Retrieval*: The encoding is the most essential part because a superior encoding will help to remember and retrieve the processed information easily over an extended period of time. The encoded information would be retained by the long term memory, if the information has been rehearsed and practiced for a considerable period of time.

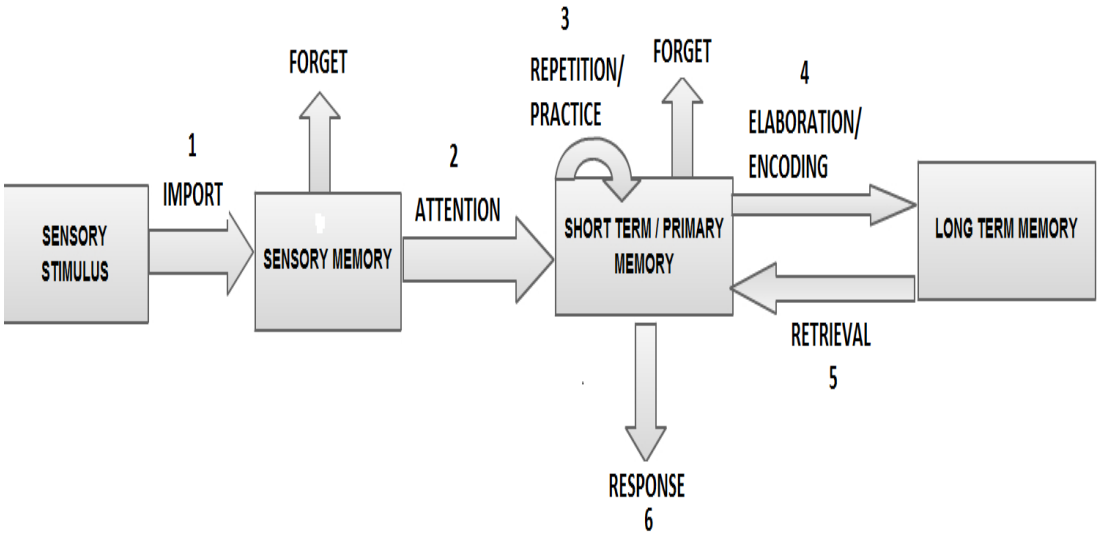


Figure 2.1: Information processing in human memory (Atinkson & Shiffrin, 1968)

2.1.2 Superior Memorability of Images than Words

RBGSs require the users to memorize their passwords and recognize them during authentication. The superior memorability of images compared to text passwords is predicted by the picture superiority effect (Shepard, 1967; Madigan, 1983). According to this theory, human beings have a vast memory to store visual information and images are more likely to be remembered than words. This effect has been further explained by the dual code theory (Paivio, 1986). This theory suggests that the images will have higher memorability than words, because images can be represented in memory as: (1) visual code, which stores the pictorial information; (2) verbal code, which stores the linguistic information. In other words, images are represented in the memory with their visual features as well as the perceived meaning. These two code representations stored in the memory are used to recognize images in subsequent use (Figure 2.2). Hence, in order to remember an image, it is also essential that the users can interpret it in a meaningful way. On the other hand, the textual information is a form of knowledge representation. They are represented symbolically in the human memory, as opposed to the dual encoding in images.

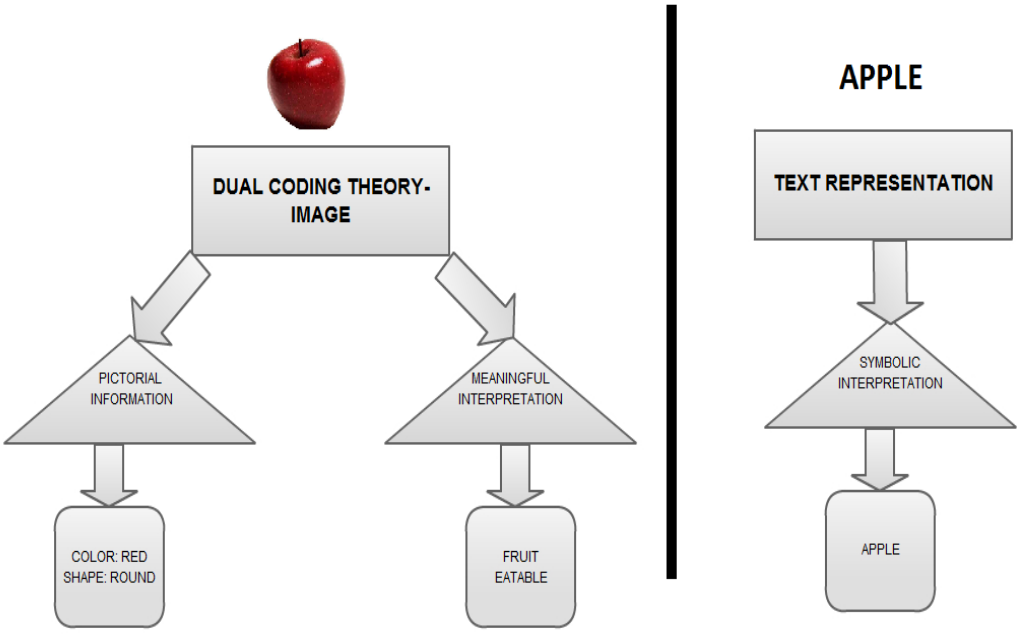


Figure 2.2: Dual coding in images and symbolic interpretation of a text

2.1.3 Guessability of Images

Sturken and Cartwright (2012) suggested that an image can have multiple meanings for different viewers. This will depend upon the viewer's cultural, social, historical and political background, as well as the context in which the image is viewed. The authors contend that an image is also associated with a *sign*. A *sign* consists of a *signifier*, usually the image itself, and a *signified*, the interpretation evoked by the image under consideration. The relationship between *signified* and *signifier* forms the meaning of an image to the viewer. Thus the meaning of an image would vary with the change in context. For example, a cigarette might signify friendship or romance in a classic Hollywood film. But, in an anti-smoking advertisement, it would signify disease and death.

The authors further suggest that images have two levels of meaning:

- *denotative*, associated with the literal and descriptive meaning of the image;
- *connotative,* associated with the cultural and historical interpretation.

For example, the denotative meaning of an image of a rose is a flower. However, depending upon the given context it may have connotative meanings involving romance, love or loyalty.

Mathur (1978) enumerates the following factors that may make an image difficult to guess.

- *Personality difference:* This stems from the fact that each human being has a distinct personality, which makes him or her unique. There are differences in terms of experience, communication abilities, expressions and many other aspects. Since people usually interpret things based on their experience and knowledge, any given person may have a distinct personal understanding of an image.

- *Difference in perception:* This is due to the way people usually interpret those parts of an image which are of some interest to them. This will depend on their perception and memories, which again vary from one individual to another.

- *Use of colors and difference in culture:* The author also cited some examples to show that colors can be culturally dependent. For instance, in western society red denotes danger, whereas in China red is perceived as a lucky mascot. Hence the viewer's background would be instrumental in influencing the interpretation of the image.

Hence an image can be interpreted in different ways by each viewer. An image can be easily guessed, if the description provided by a viewer is denotative, i.e. describes the elements in it. But, if the description is connotative, i.e. the viewer relates it to something personal (an idea or event that only has relevance to them), a sign (a secret meaning) or state (how it makes them feel), then it might be very difficult to guess the image, without being aware of the relation between the context of the description and the image.

2.2 Recall-Based GASs

The following subsections offer an overview of the usability of recall-based GASs, followed by a brief discussion of some common security vulnerabilities in these systems. Table 2.1 summarises the usability results reported in the existing literature of recall-based GASs.

2.2.1 Usability of Recall-Based GASs

Jermyn et al. (1999) proposed *Draw-A-Secret (DAS)*, which is also the archetypal example for this category. In this scheme, users are required to draw their password on a 2D grid using a stylus or a mouse as shown in Figure 2.3. To authenticate, users must draw the same password using the same order of pen strokes and pen-up events. Hence a drawing may

comprise of either one continuous pen stroke, or a series of pen strokes separated by pen-ups, which restart the next stroke.

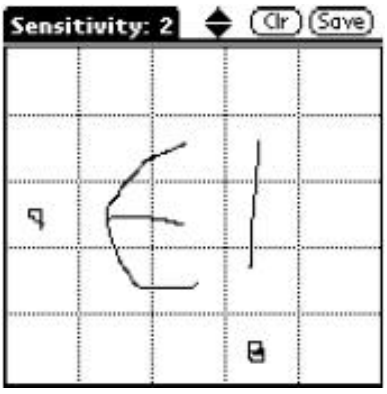


Figure 2.3: An example of a DAS password (Jermyn et al., 1999)

It is difficult to get an accurate analysis of the usability or security of DAS, as it has only been tested through paper prototypes. The results reported in Nali & Thorpe (2004) showed that participants tended to draw symmetric images with (1-3) pen strokes and place the drawings approximately in the centre of the grid.

A novel variant of DAS scheme, i.e. *BDAS* was presented in Dunphy & Yan (2007). In BDAS, users are first required to choose a background image to be overlaid on the grid, and then draw their password. The authors have reported two user studies with paper prototypes, to compare the effectiveness of DAS and BDAS. The results from both the studies demonstrated that the use of background images made the users draw, complicated and longer passwords, which were memorable to the same extent as the DAS passwords. The amount of symmetry within the password images (drawings) in BDAS was also reduced, compared to that of DAS.

PassShapes was proposed in Weiss & Luca (2008). In this system, users are required to draw geometric shapes. The geometric shape is translated into alphanumeric strings based on 8 stroke directions, recognised at 45 degree intervals as shown in Figure 2.4. During authentication, users must draw the same shape using the same order of the strokes. The size and position of the drawing is not important. The lab studies reported by the authors showed that the login success with Pass Shapes varied between 63% and 100%, over a period of ten days. The average login time was found to be 6.5 sec.

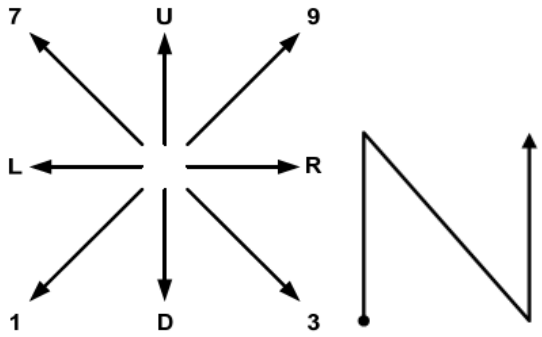


Figure 2.4: Eight different strokes used in PassShapes; An example PassShape with alphanumeric representation U3U (Weiss & Luca, 2008)

Passdoodle was reported in Goldberg at al. (2002) and Varenhorst (2004). This system is similar to DAS, in the sense that users are required to create free hand drawings as their passwords, but in the absence of a grid (unlike DAS). None of the Passdoodle studies have reported the relevant statistics related to the usability metrics such as login success, registration and authentication time.

Pass-Go was reported in Tao & Adams (2008). In this system, users are required to draw their password on a grid using the grid intersection points, instead of grid cells, as in DAS. The idea was based upon the Chinese board game of Go, which involved strategically placing tokens on the intersection points of a grid. This is the only recall-based system which was evaluated in a field study conducted with 167 participants. The login success rate was 78%, but no login times were reported. The results also showed that a large number of passwords were symmetrical and would be susceptible to attack, though the passwords drawn by the users were more complex than DAS.

A recall-based scheme using a haptic input device to measure the pen pressure, while users draw their passwords, was reported in Orozco et al. (2006). The motivating idea was it would be difficult for an observer to distinguish between variations in pen pressure; hence the system will be resistant to shoulder surfing attacks. The results showed that users applied very little pressure and hardly lifted the pen, while drawing their passwords. It was concluded that the use of haptics would not decrease the susceptibility of recall-based systems against should surfing attacks.

An android grid-based screen-unlock scheme, which is similar to Pass-Go has been commercially deployed, but was shown to be susceptible to smudge attacks (Aviv et al.,

2010), i.e. attackers would be able to determine a user's password through the finger smudges left on the smart phone's screen. Blackberry had also introduced a similar system (PatternLock), where users are required to draw their password on a 3 × 3 grid, instead of typing a 4 digit PIN.

In Table 2.1, we summarise the results reported in the literature in relation to the recall based GASs. Where the information was not reported/available in the literature, *Unknown* is used to denote it. For each user study, the number of sessions and the duration of the study in weeks are presented. For e.g. in the case of DAS, two paper studies were conducted: one comprised of a single session (1×) and the second had two sessions spread over single week (2×1wk). The registration and login time was not reported (Unknown) and login success rate varied between 57% and 80%.

System	Type of study	Study duration	Registration time (seconds)	Login time (seconds)	Login success rate	Multiple password studies
Draw-A-Secret	Paper	1 x 2 × 1wk	Unknown	Unknown	57%-80%	No
BDAS	Paper	2 × 1wk 2 × 1wk	Unknown	Unknown	50%-80%	No
PassShapes	Paper	3 × 1.5wk	Unknown	Unknown	63%-100%	No
	Lab	3 × 1.5wk				
Passdoodle	Paper	2 × 1wk	Unknown	Unknown	38% -46%	No
	Lab	1 x				
Pass-Go	Field	13wk	Unknown	Unknown	78%	No
Haptic	NR	1 x	Unknown	Unknown	Unknown	No

Table 2.1: Summary of recall-based GASs

2.2.2 Security Overview

Recall-based systems discussed in section 2.2.1 are prone to a number of security vulnerabilities as discussed below.

- In the context of the guessing attacks, Oorschot & Thorpe (2008) showed that dictionary attacks on DAS and Pass-Go would require much less effort, than implied by their theoretical password space. The authors categorized the passwords (drawings) into various classes based on their symmetry and pen strokes. Using this classification, they showed that many DAS passwords reported in Nali & Thorpe (2004), fall within the predictable categories. This information could be easily used by the attackers to identify candidate passwords and efficiently launch a dictionary attack.
- These systems are generally susceptible to shoulder surfing attacks, since the entire drawing is visible on the screen, when the user draws it. An observer can observe or record the entire drawing (which is the password) accurately.
- It is also possible to describe these passwords by explaining the path through the grid squares or simply drawing a sketch of the password. Hence social engineering attacks remain a concern. However, none of the existing studies have examined this aspect.

2.2.3 Summary

The key motivation behind recall-based schemes is associated to the superior memorability of the images. However, the lack of suitable usability as well as security studies makes it difficult to demonstrate the effectiveness of these systems.

2.3 Cued Recall-Based GASs

The following subsections will present a brief overview of the usability of cued recall-based GASs, followed by a brief discussion of some common security vulnerabilities in these systems. Table 2.2 summarises the usability results reported in the existing literature of cued recall-based GASs.

2.3.1 Usability of Cued Recall GASs

The literature in cued recall-based systems is dominated by PassPoints (Wiedenbeck et al., 2005a; 2005b). A password in this system is a sequence of five click-points selected by the user on an image assigned by the system. During login, the user must choose each click-point in the correct order and within a tolerance area specified by the system, as shown in Figure

2.5. The results of the lab-based studies reported in Wiedenbeck et al. (2005a) showed that the password creation time was 64 sec, and the users required an additional average training time of 171 sec to memorise their respective passwords. The mean login success rate varied between 55% and 90%, whereas the mean login time varied between 9sec and 19 sec.

Figure 2.5: An example of PassPoints (Wiedenbeck et al., 2005a)

Cued Click Points (CCP) was reported in Chiasson & Oorschot (2007), where a user is required to choose a single point (click-point), on each of the five distinct images presented by the system in a sequential order, unlike Passpoints (five points on one image). The images in sequence are dependent on the coordinates of the click point in the prior image. Hence a wrong selection during the login process, would not display any of the images that were used to create the respective click points. The result of the authentication session is shown only after the final point is clicked. The study reported in Chiasson & Oorschot (2007) showed that the mean registration time was 25 sec and the mean authentication time was 7 sec. The mean success rate was 96%. The analysis of the click points chosen by the users revealed that, users tended to choose points falling within known hotspots, i.e. popular points or areas of an image with higher probability of being chosen as click-points.

Persuasive Cued Click Points (PCCP) is a variant of CCP, designed to persuade the users to select more secure passwords (Chiasson et al., 2011). During the password registration, the image is dimmed, except for a small square view-port area randomly positioned on the image, as shown in Figure 2.6. Users can either select a click point from within the viewport, or press the shuffle button to randomly reposition the viewport, until they are satisfied with the location. The login process is same as that of CCP. The idea was to reduce the hotspots, by flattening the distribution of the click-points across multiple users. The login success rate of PCCP (83%-94%) was similar to CCP. However, the mean password registration time

increased to 50 sec in PCCP, mainly due to the participants who shuffled the viewport repeatedly. It took almost 8 sec to login in PCCP. The review of PCCP presented in Biddle et al. (2009) reported that login time in PCCP varies between 11 sec – 89 sec.

Figure 2.6: PCCP with a view-port area (Chiasson et al., 2011)

Suo (2006) proposed a modified version of PassPoints to resist against the shoulder surfing attacks. The registration process is same as the Passpoints. During login, the image is blurred except a small focus area. Users must enter Y or N using a keyboard, or use the right and left mouse buttons, to indicate, if the click points selected during registration are within the focussed area. This process is repeated for at most 10 rounds, until all the click points are identified. However, the author did not report any user study to demonstrate the effectiveness and efficiency of the proposed approach.

Alsilaiman & El Saddik (2006) proposed a system, where users would navigate a 3D environment, and perform a sequence of actions such as clicking on certain areas, interacting with certain parts of the virtual world, typing or drawing on the virtual surface, which will be interpreted as their password. The idea is that 3D environment will act as the cue to prompt the users to repeat their actions. A prototype was implemented, where users can walk through a virtual art gallery, and enter text passwords at the virtual computers or select pictures to form their passwords. However, no user studies were conducted to demonstrate the effectiveness and efficiency of the proposed system.

A variant of cued recall-based system, also known as Windows Picture Passwords has been commercially deployed on Windows operating system (Windows, 2011). In this mechanism, users are first required to choose an image from their personal collection (any collection), and then highlight the parts of the image that are either important or interesting to them. The highlighted sections (gestures) on the image form the password. Ideally, the user has to choose three gestures (analogous to three click-points in CCP). The users can choose the gestures by: clicking on the image using a mouse in the case of a desktop; tapping on the image (handheld devices); connecting/highlighting points in the image by drawing lines or circles. This is one of the widespread deployments of the graphical authentication.

In Table 2.2, each of the cued recall-based GASs discussed above is evaluated, based on the usability results presented in the literature. The structure of the table and terminologies used are the same as in Table 2.1. Please note that the statistics presented for CCP is obtained from the single password studies. The statistics related to multiple passwords study reported in Chiasson et al. (2009) will be presented in Table 2.5 (Section 2.6).

System	**Type of study**	**Study duration**	**Registration time (seconds)**	**Login time (seconds)**	**Login success rate**	**Multiple password studies**
PassPoints	Lab	2 × 1wk 1 × 2 × 2wk	64 sec	9 -19 sec	38% - 94%	No
	Field	7-9wk				
CCP	Lab	1 × 2 × 1wk	25 sec	7 sec	96%	Yes (Chiasson et al., 2009)
Suo	No study	Unknown	Unknown	Unknown	Unknown	Unknown
PCCP	Lab	1 × 2 × 1wk	50 sec	8-89sec	83%-94%	No
3D	No study	Unknown	Unknown	Unknown	Unknown	Unknown
Windows Picture	No study	Unknown	Unknown	Unknown	Unknown	Unknown

Table 2.2: Summary of cued recall-based GASs

2.3.2 Security Overview

Passpoints and its variants have been analysed rigorously in the context of security.

- The two major weaknesses that enable efficient dictionary attacks in PassPoints are: hotspots (Chiasson & Oorschot, 2007) and patterns (Chiasson et al., 2009). If an attacker can predict the hotspots in an image through image analysis or predictable behavior of the users, then a dictionary of passwords comprising of such hotspots can be created, to launch a dictionary attack. Chiasson et al. (2009) showed that geometric patterns were absent in case of CCP, as the password was constructed across 5 images, as opposed to 5 click-points in a single image (PassPoints). According to Chiasson et al. (2011), PCCP eliminated the major concerns regarding hot spots and patterns.
- In the context of shoulder surfing, the user's password is observable on the screen, while the user enters it. Hence, the system may not be resistant to shoulder surfing attacks. For example, it might be easier for an observer to detect the change in images in the case of CCP, than the movement of the mouse pointer in PassPoints. It is also easy to capture the user's screen or authentication session using cameras and other recording equipment. However, none of the existing studies have evaluated this aspect in the context of the CCP.
- Social engineering attacks remain a concern, as it may be possible to describe the click-points verbally or simply using screen captures and other recording equipment. However, none of the existing studies have examined this aspect.

2.3.3 Summary

The potential problems in a cued recall system, discussed in Renaud & Angeli (2004) are listed below:

- According to Parkin (1993), the effectiveness of cued recall would depend on the strength of the association between the cue (memorable objects in the image) and response (correctly pointing the location). However, it is difficult to find an image which has many memorable locations, does not have too many bright objects and consists of smooth well defined objects. In case the image does not have many memorable locations, the same location are likely to be used by many users as their

passwords. This may make the system insecure to use. PCCP has claimed to solve this problem of hotspots, however the system is time consuming to employ.

- The users may also find it difficult to click on the exact location (click-point) which will lead to an unsuccessful authentication attempts (Renaud & Angeli, 2004). This may make the system difficult to use. However, this aspect has not been evaluated in the existing studies.

To summarise, early cued recall-based systems such as PassPoints were found to be usable in single password studies, but suffered from security issues due to hotspots and patterns in the user selection. Later schemes such as PCCP alleviate the security issues, but their performance in terms of usability (especially, efficiency) deteriorated. Chiasson et al. (2009) reported the only multiple password study in the field of cued recall-based systems, which will be further discussed in Section 2.6.2.

2.4 Recognition Based Graphical Authentication systems (RBGSs)

The RBGSs reported in the existing literature have used various types of images, such as faces, random art, objects, doodles, icons and personal photographs. The following subsections will present an overview of the usability of RBGSs, followed by a brief discussion of the some common security vulnerabilities in these systems. Table 2.3 summarises the usability results reported in the existing literature of RBGSs and is presented at the end of this section.

2.4.1 Usability of Different Image Types in RBGSs

(A) Face passwords

The RBGS studied most extensively to date is *Passfaces* (Real, 2004). In most configurations, during registration users must select 4 target faces from a collection presented by the system. The authentication is a four step process. At each step, a challenge set consisting of 8 decoy faces and 1 target face is displayed in a 3×3 grid. The user has to recognise and select the target face at each step.

The results from a field study reported in Valentine (1999) with 77 users demonstrated that the success rate of the authentication varied between 72% and 100%, over a period of five months. A field study was reported in Brostoff & Sasse (2000) to compare the usability of face and text passwords. The result obtained from the field study showed that the authentication errors were much less for face passwords (mean success rate: 96.1%). Davis et al. (2004) reported a 16 week field study with students, who had to use faces as their password to access their class materials. The login success rate varied between 85 % and 97%. The analysis of the faces revealed that the target faces chosen by the users is biased by the race, gender and attractiveness of the face. This would increase the guessability of the passwords. Hence it was concluded that user involvement in the password creation should be minimised, when faces are used. None of the above studies have reported the password creation and login time. However, the Passfaces corporate website Real (2004) suggested that the password creation would take 3-5 minutes for a panel of 9 faces and 5 rounds.

All the aforementioned studies have explored the usability of a single face password. To our knowledge, Everitt et al. (2009) and Hlywa et al. (2011) are the only studies in the field of RBGSs that have explored the usability of multiple face passwords. A detailed review of these studies will be presented in Section 2.6. Dunphy et al. (2008) reported a study to examine the vulnerability of faces to verbal descriptions. A detailed review of this study will be presented in the Section 2.6.5.

(B) Dejavu (Random art passwords)

A RBGS named *Dejavu,* which used abstract art images as the password was reported in Dhamija & Perrig (2000). The motivating idea behind the use of random art images was that it would be difficult to communicate/record such images, through written and verbal descriptions. Hence this may be a way of devising memorable passwords, which are difficult to communicate or share.

In this system, users must select five target images from a collection presented by the system and recognise all the target images among a collection of 20 decoy images, displayed in a single challenge set, to authenticate successfully. A user study conducted with 20 participants demonstrated that users took an average of 45 sec to create their passwords, and an average of 32 sec to login immediately, after the password creation. The participants took an average

of 36 sec to log into the system after a period of 1 week. The login success rate was 100% just after the password creation and 90% one week later, when a single password was used.

It was also reported that the participants in the user study found it difficult to describe their password images. Some of the participants who chose the same images as their password gave different descriptions. The authors concluded that abstract art images would make it difficult to launch a social engineering attack, if the attacker is relying on the user to verbally describe their password images. But, this claim was not supported by any experimental evidence.

(C) Object passwords

A RBGS using *story* passwords was reported in Davis et al. (2004). In this system, users were required to employ a mnemonic strategy to select their password comprising of four target images. To authenticate users must select the four target images among a collection of 5 decoy images in the correct order. This system was also tested with faces as part of the same study. It was found that the mean login success rate of the story passwords was 85%. It was also found that out of the 236 incorrect password entries, 75% of the password images were selected correctly, but in the wrong order. The survey with the participants revealed that, 50% of the participants did not formulate a story as a memory aid. The authors claim that it may be easier for the users to describe their respective passwords in the story scheme, as they belong to everyday objects. However this claim is yet to be tested.

Angeli et al. (2005) reported two studies, comparing three different configurations of a *Visual Identification Protocol (VIP),* which used object images as the password. VIP 1 required users to recognise four target images issued by the system in a challenge set comprising of 6 decoy images, which were displayed in the same position for all authentication sessions. VIP 2 was same as VIP1, except that the four target images were displayed in random positions for each authentication attempt. In case of VIP3, 8 target images were assigned to the user. During authentication, any four target images were randomly displayed in a challenge set together with 12 decoy images. The results obtained from a lab-based and a web-based experiment showed that the login failure rate in case of VIP1 and VIP2 was less than 5%, whereas VIP 3 had a failure between 10% and 12%. The login time was lowest in case of VIP1 (5sec - 8 sec), followed by VIP2 (10sec - 13sec) and highest in case of VIP 3 (15sec - 22 sec).

PassImages was reported in Charrau et al. (2005), which again used object images as the password. In this system, users were required to select six target images from a collection provided by the RBGS. During authentication, six target images were displayed across four challenge sets. Each challenge set was presented as a 5×5 image grid, comprising of one or more target images and rest decoy images. The results of a web-based study conducted with twenty-nine users over a period of 3 months showed that the login success rate was 95%, mean registration time was 180 sec and the mean login time varied between 20 sec and 32 sec.

All the aforementioned studies have explored the usability of a single object password. Moncur & Leplatre (2007) and Hlywa et al. (2011) are the only known studies in the field of RBGSs that have explored the usability of multiple object passwords. A detailed review of both the studies will be presented in Section 2.6.

(D) Doodle passwords

A visuo-biometric authentication mechanism known as *Handwing* was proposed in Renaud (2005). The authentication mechanism used handwritten PINs, handwritten postal codes and hand drawn doodles as the password in a three stage authentication process. During authentication, users must recognise the credentials provided by them among a collection of decoys, as shown in Figure 2.7. The main idea behind the proposed system was to provide usable and secure authentication mechanism for older users. A user study was conducted with users over the age of 50. The results demonstrated that most users were able to remember their password credentials. However, the time taken to register and login was not reported. The author claimed that the authenticators can be recognized by the cohabitant or family members of the user, but it will be difficult for an unknown to predict it.

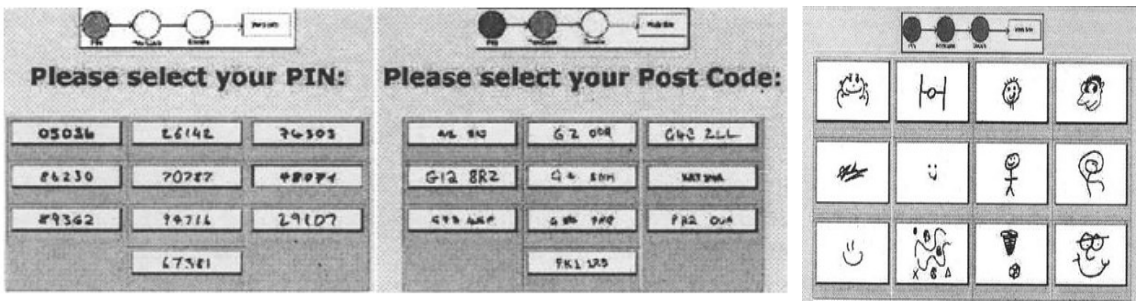


Figure 2.7: Three authentications steps comprising of PINs, Postal codes, doodles provided by the user, which are displayed among a collection of decoys.

A longitudinal study was reported in Renaud (2009a) to determine the efficiency of the doodles, photographs and object, using three levels of user involvement as follows:

a. The system assigned the first set of users a random set of object images;
b. Second set of users drew doodles;
c. Third set of users used photographs captured by them;

During authentication users must identify their target image in each challenge set comprising of 16 images (15 decoys and 1 target). The results of the memorability study conducted over a period of three months showed that the mean success rate of hand drawn doodles was highest (88.5%), followed by photographs (77%) and lowest in case of system assigned object images (67%). Another experiment was performed, to test the memorability of the doodle images over a period of 6 months, with school students of 11- 12 years of age. The results demonstrated that there was an overall login success rate of 85.5%. These results showed that the doodle images have long term memorability.

The above studies provide evidence that doodle images are suitable to be used as authenticators in RBGSs. However, in all such systems users would need to draw the doodles, which would make the registration process time consuming.

(E) Mikon passwords

Renaud (2009b) reported an authentication system which used Mikon[1] images as the password. These are icon like images that were drawn by the users using an online tool developed by the Mikons.com Company[1]. The authentication mechanism relied upon the recognition of previously drawn Mikon as shown in Figure 2.8. A web-based study was conducted with 26 school students of 11-12 years of age, for a period of 3 months, to determine the performance of the Mikon authentication system. In terms of memorability, it was found that the mean login success rate was 87%. In terms of scalability (i.e. the registration process is not time consuming; hand drawn images need not be scanned in manually and uploaded to the system), it performed better than that of doodle authentication systems, which required human involvement to scan and upload the authenticator images. The results do not give any information about the efficiency (enrolment and authentication

[1] www.mikons.com

time) of the Mikon authentication system. The findings of the study show the potential of the Mikon images to be used as authenticators.

Figure 2.8: A challenge set in Mikon system comprising of 1 target and 15 decoy images

A user study to compare the usability of doodle and Mikon images selected by the user from a collection of images provided by the system was reported in Chowdhury & Poet (2011). The results reported in the study conducted with 20 participants demonstrated that the mean login success rate over a period of 5 days for both the doodle and Mikon passwords was 100%. The mean registration time was 31 sec in case of Mikons and 39 sec for doodle passwords. The mean authentication time was around 30 sec for Mikon passwords and varied between 47 sec – 49 sec in case of doodle passwords.

(F) Shoulder surfing resistant RBGSs

An attacker can capture a user's credentials by direct observation or by recording the individual's authentication session. This is known as a shoulder surfing attack. In this section, we present a brief review of some RBGSs that were designed to resist such attacks.

Weinshall (2006) reported a *Cognitive Authentication Scheme*, where users are assigned 30-60 pictures randomly which is chosen by the system and they must memorize them. During authentication, (20-30) pictures are displayed in a single challenge set consisting of (40-50) decoys. The authentication task involved computing a path through the challenge set starting from top-left corner based on whether an image belongs to the user. On reaching the challenge set's bottom or right edge, users must identify the corresponding label for that row/column. A multiple choice question is presented, which includes the label of the path's

correct endpoint. Users must perform several such rounds. At the end of each round, the system would calculate the cumulative probability that the correct answer was entered by chance. Once the probability is over a certain threshold value, the authentication is considered as successful. The results obtained from a user study demonstrated that the login success rate was 90-100% and the login time was 180 seconds, for the configuration (10 rounds) reported in the published research.

Convex hull scheme was reported in Wiendenbeck et al. (2006). The icons of various software applications were used as the visual cue. During registration, users were required to choose a number of target icons to form their respective password. The authentication required a series of challenge-response round. In each round, users must first visually locate the target icons among a collection of decoys randomly distributed in the challenge set, visualise the convex hull triangle formed by the target icons and then click anywhere within the triangle. At least three target icons were displayed in each challenge set, since forming a convex hull requires three icons. In order to authenticate successfully, users must respond to the challenge presented in each round correctly. The usability study was conducted with fifteen users, who used 5 target icons to form a password and had to successfully click in the convex hull area for each of five rounds. The mean success rate was 97.95% just after selecting the password. The mean login time was 71.66 sec.

A system known as Use Your Illusion, where users had to recognise a degraded version of a previously selected image was reported in Hayashi et al. (2008). The aim was to make the scheme resilient to social engineering and observation attacks, while maintaining its usability. During registration the users were required to capture three photographs (target images) using their phone and upload it to the system. The system then presents a distorted form of each of the photographs, which make up the password. During authentication users must identify, each target image among a collection of 8 decoys, at each step to authenticate successfully. The results of the user study conducted with 54 participants showed that the mean login success rate was between 89% and 100% over a period of four weeks. The login time varied between 11.5 sec and 25.8 sec. A system based on the technique where you see is what you enter, as a defence against shoulder surfing attacks was proposed in Khlot et al. (2012). In this technique the users had to identify the pattern of pre-selected x number of target images in an M × M grid and then map the position of the identified pattern on to another N × N grid. A usability study was conducted with 24 participants and the results demonstrated that all the

participants were able to login successfully within three attempts after the registration stage. The mean login time varied between 12.1 sec and 35.5 sec.

The basic issue with the shoulder resistant systems discussed in this section are:

- Users would need to remember a large number of target images for a single password, which might make the system less usable, if multiple passwords are employed;
- The registration as well as the login time in these systems is very high, which would make them unpopular in the real world.

Based upon the issues listed above, we believe that shoulder surfing resistant schemes are unlikely to be memorable and less time consuming to employ, if multiple RBGS passwords are used. Hence these systems won't be considered further in this thesis.

(G) Personal Photos

Tullis & Tedesco (2005) reported the use of personal photographs as passwords in RBGSs. In this system users were required to provide 8-20 personal photographs, which they believe wouldn't be recognised by someone else. During authentication, a single challenge set comprising of 15 photos, i.e. 2-5 target photographs and rest decoy photographs was presented. In order to authenticate successfully, a user must recognise all the target photos in the challenge set. The results of the user study conducted over a period of 2 months showed that the login success varied between 94 % and 100% and the mean login time varied between 11.2 sec and 21.3 sec.

Takada et al. (2006) proposed a RBGS named *Awase-e*, which used the user's personal photographs as the password. Each user had to upload certain number of photographs to the system. Then they had to select four of the photographs (target) which would be used as their password. During authentication a user was presented with four challenge sets, each containing 9 images and a 'no pass image option'. A user must select the target photograph, if it is included in the challenge set or choose the option if it is not in the challenge set. Memorability evaluation experiments were conducted in week 2, 4, 8, 16 weeks after the training stage. The mean success rate over the sixteen weeks was found to be 100% and the mean registration time was 24.6sec. The password creation time was not reported in the study.

In the photographic authentication system reported in Perrig et al. (2003), users were first required to provide a set of personal photographs. During authentication, the users must identify the target among a collection of 3 decoy photographs, at each step of authentication comprising of 10 rounds. The results of the memorability study showed that all the users were able to authenticate successfully and the mean login time was approximately 40 sec. The password creation time was not reported in the study.

2.4.2 Security Overview

The most prominent security analysis of RBGSs was reported in Davis et al. (2004). The results highlighted that face as well as story passwords had exploitable patterns, when they were selected by the user themselves. In this context, a probabilistic model was developed and a dictionary was generated, using random subsets containing 80% of the passwords chosen by the user in both face and story scheme. The results of the dictionary attack demonstrated that: (i) the weakest 25% of the face passwords could be guessed in 13 attempts and the weakest 10% of the passwords corresponding to male users in 2 guesses; (ii) The weakest 25% of the story passwords could be guessed in 112 attempts, whereas the weakest 10% in 35 guesses. Shoulder surfing is a concern in most RBGSs since the entire password is revealed on the screen. Hence an attacker can record or observe the images selected by the users during authentication because the number of images in a challenge set is relatively few. Some of the RBGSs discussed in Section 2.4.1 proposed various shoulder surfing resistant approaches. However, these systems are time consuming to use and possess additional usability issues, for e.g. users in Weinshall (2006) had to remember at least 30 images as their password.

In the context of social engineering attacks, some of the widely held assumptions related to Passfaces and other GASs are listed below.

- Dhamija & Perrig (2000) reported that participants in the usability study found it difficult to describe their art passwords concretely, and often related them to objects or actions. Hence authors believe that it would be difficult for an adversary to guess art passwords based on the descriptions;
- According to Real (2004), one claimed advantage of the Passfaces over text passwords is that: "*Passfaces can't be written down or copied and can't be given to*

another person". "*Passfaces can be used in grayscale on all platforms in order to make it even harder for a user to describe their Passfaces to someone else*";

- Poet & Renaud (2009b) claimed that doodles are hard to describe without reproducing them, i.e. drawing them. Hence it will be difficult for someone to verbally communicate the doodles.

However, none of the above claims have been tested in the existing studies. Dunphy et al. (2008) is the only known work that had explored some aspects of Passfaces in the context of the descriptions: (i) various approaches to Passfaces description; (ii) evaluate approaches to reduce vulnerability of Passfaces to descriptions, through strategic selection of decoy images in the challenge set; (iii) exploring gender differences in the context of the creation of descriptions. This work will be further reviewed in Section 2.6.5.

2.4.3 Summary

To summarise, all the user studies discussed in section 2.4.1 have focused on the unrealistic use of a single password. In the last fifteen years (to our knowledge), only three multiple RBGS password studies have been reported. All these studies are critically reviewed in the Section 2.6. The research reported in Dunphy et al. (2008) the only work in the field of RBGS that is closely related to the focus of this thesis, i.e. guessability of RBGS passwords to written descriptions. Hence the study will be further discussed with relevant details in the next section. Some significant findings in the context of the RBGSs are listed below:

- RBGS passwords have good memorability characteristics, even over extended periods of non-use (Tullis & Tedesco, 2011);
- The memorability, i.e. login success rate is negatively impacted by ordered selection of the target images (Davis et al., 2004);
- User involvement during password creation has a positive impact on the memorability (Renaud, 2009a).

In Table 2.3 a summary of the results reported in the existing RBGS studies (single password) is presented. However, the table does not present any results reported in the multiple RBGS password studies. These studies will be discussed in Section 2.6 and the corresponding results are presented in Table 2.5.

Image Type	Type of study	Duration of study	Login success rate	Registration time (seconds)	Login time (seconds	Multiple Password	Password recordability and descriptions
Faces	Field	≥ 16wk 10wk	72% – 100%	NR in single password studies	180- 300 sec	(Everitt et al., 2009) and (Hlywa et al., 2011)	(Dunphy et al., 2008)
	Lab	1-20wk					
Random art	Lab	2 × 1wk	90% - 100%	45 sec	32-36 sec	No	Claim made – no user study
Story (objects)	Field	≥ 16wk	85%	Unknown	Unknown	Yes (Moncur & Leplatre, 2007) (Hlywa et al., 2011)	No
VIP 1, 2 and 3 (Objects)	Lab	1 ×	85 % - 95%	Unknown	5 – 22 sec		
	Web	≥ 16wk					
PassImages (Objects)	Web	12 wk	95%	180 sec	20 sec – 32 sec		
Doodle	Lab	12wk 2 × 1wk	85.5% - 100%	39 sec (user selects the targets)	47 – 49 sec	No	Claim made – no user study
Mikon	Web	12 wk	87% - 100%	31 sec (user selects the targets)	30 sec	No	No
	Lab	2 × 1wk					
Personal photographs	Lab	8wk 4× 16wk 1 ×	90% - 100%	Unknown	11.2 – 40 sec	No	No
Shoulder surfing scheme	Lab	2 × 1wk 4 × 1wk 13 × 1wk	89% - 100%	Unknown	71.66 sec 90-180 sec 11.5-25.8 sec	No	No

Table 2.3: Summary of Recognition-based GASs

2.5 Comparing the GAS Categories

In this section, we compare the three GAS categories (presented in Sections 2.2 to 2.4), based upon the results reported in the existing studies, which have been already presented in Tables 2.1 to 2.3. The comparisons are based upon the aspects listed below. Table 2.4 further summarises the comparisons reported in this section.

- *Type of study:* The review shows that most recall-based GASs are evaluated using paper prototypes. PassShape and Passdoodle systems are examined in a lab-based environment, and Pass-Go is examined in a field trial. In the context of the cued recall-based GASs, most variants are either evaluated in lab or field trials. Similarly, most studies with RBGSs have been also examined in web, lab or field trials. Hence most variants of cued recall based systems and RBGSs have been thoroughly evaluated, in the context of remembering a single password.

- *Login success rate:* The login success rate varies between 38-100% for recall based GASs, 38-94% for cued recall systems and 72-100% in the case of RBGSs, in the context of remembering a single password. In relation to the minimum login success rate, RBGSs (72%) perform better compared to the other two GAS categories. However, due to the differences in the experimental frameworks and dependent variables reported in the literature, it is difficult to explicitly state the best performer.

- *Registration time:* The registration time has not been reported for any recall based system in the existing studies. In the case of cued recall systems, the registration time varies between 25sec and 64sec, whereas the variation in the case of RBGSs is between 31sec and 180 sec. Based on these figures cued systems perform better than RBGSs. However, the registration time would also depend upon a number of factors such as the number of images/click-points used as the password and time taken to load the images in the interface. Hence the registration time would differ from one system to another, depending upon its configuration.

- Login time: The login time has not been reported for any recall based system in the literature, except PassShapes (6 sec). In case of cued recall systems the login time

varies between 8sec - 89sec, whereas the variation in the case of RBGSs is 5sec – 300sec. Hence based on these figures cued recall systems perform better than RBGSs. However, the login time in the case of RBGSs would also depend upon a number of factors such as the number of challenge sets and time taken by the application to load each challenge set.

- *Multiple password study:* There are very few studies of multiple password use in the field of GASs. To our knowledge, four studies have examined the memorability of multiple graphical passwords: three studies have evaluated a RBGS and one study has evaluated a cued recall based system. None of the studies in the literature have reported the memorability of multiple graphical passwords using a recall based system. A review of these studies will be further presented in Section 2.6.

- *Password description study:* The topic of descriptions has not been examined extensively in relation to the graphical passwords. Moreover, it is a widely held assumption that graphical passwords are resistant to being written down and difficult to guess by an adversary using written descriptions (Dhamija & Perrig, 2000; Real, 2004). To our knowledge, only one study has examined the guessability of graphical passwords in a RBGS using faces as the visual cue, which will be further reviewed in Section 2.6.5.

System	**Type of study**	**Login Success rate (%)**	**Registration time (seconds)**	**Login time (seconds)**	**No of multiple password studies**	**No of password description study**
Recall based	Paper Lab Field	38 – 100	Not reported	6 (PassShapes)	None	0
Cued recall based	Lab Field	38 - 94	8 - 89	9 - 89	1	0
RBGS	Lab Field Web	72 - 100	31 – 180	5 - 300	3	1

Table 2.4: Comparing the three GAS categories

2.6 Literature Related to the Thesis Objectives

The survey of the existing GASs in Sections 2.2, 2.3 and 2.4 shows the existing interest of the research community in graphical passwords as an alternative to text passwords. An important limitation of the field, which has been identified from the previous sections, is: most user studies in the field of GASs have focused on the usability of a single password. However, if GASs were to be widely adopted, users would, in general, need to use multiple graphical passwords, just as they currently use many text passwords. To our knowledge, in the last fifteen years, only four studies in the field of GASs have explored the memorability of multiple graphical passwords. The literature presented in Section 2.4.2 also provides evidence that RBGS passwords are assumed to be particularly resistant to being written down, or verbally communicated. But, none of the existing studies except Dunphy et al. (2008) has examined this assumption. Moreover, there is no standard methodology to:

- quantify the extent to which users can share their RBGS passwords;
- identify the coping mechanism that will be used to record such passwords;
- measure the guessability of RBGS passwords using various coping mechanisms.

This section will present a review of all the four user studies in the field of GASs that have explored the memorability of multiple graphical passwords, and one empirical study that has examined the vulnerability of face passwords to verbal descriptions. The review will discuss the design of the system used for the user experiments, the experimental protocol followed, the results obtained from the experiments and our inferences (which are based upon the published research), for each study. In this context, Table 2.5 presents a summary of the results reported in the existing multiple graphical password studies, at the end of this section. The details and relevant statistics in Table 2.5 are reported from the original publications.

2.6.1 Multiple Password Study with Object Images

The first user study that had examined the memorability of multiple graphical passwords in a RBGS was reported in Moncur & Leplatre (2007). The lab-based study compared the memorability of multiple graphical passwords to multiple PINs (four digit numerical passwords). Images of everyday objects such as food, music, sports, and flowers were used as the visual cue in case of graphical passwords.

(A) Configuration of the RBGS

Each participant was assigned five numerical or graphical passwords. In case of graphical password, each password comprised of four colourful and meaningful images. During authentication, a challenge set containing 10 images was presented to the participants, as shown in Figure 2.9. The participants had to select four target images in the correct order, among the collection of six decoy images. In case of PINs, the numbers (0-9) were displayed on the screen, and the participants had to click on the numbers in the correct order that formed the digits of their PIN.

Figure 2.9: Authentication screen comprising of 10 images (4 target + 6 decoys) (Moncur & Leplatre, 2007)

(B) User Study

The usability study examined the memorability of five system-issued passwords with 172 university students, who were assigned randomly to one of the five groups, as given below.

- *Group 0:* participants used four digit PINS.
- *Group 1:* participants used object passwords.
- *Group 2:* participants used object passwords displayed with a signature color background to enhance memorability.
- *Group 3:* participants employed a mnemonic strategy to remember their passwords.
- *Group 4:* participants employed a mnemonic strategy to remember their passwords, which were displayed with a signature color background.

Three memorability tests (RT1, RT2, and RT3) were conducted, with a gap of two weeks between each one of them. There were no practice sessions in between each of the memorability tests. The overall drop-out rate in the user study was 64.91%. The high drop-out rate made it difficult to analyse the results, because the number of participants who completed each memorability test varied significantly.

(C) Results

The mean login success percentages discussed below have been obtained from the graph reported in Moncur & Leplatre (2007), which is shown in Figure 2.10. The authors do not provide raw statistics; hence the results discussed below are close approximations.

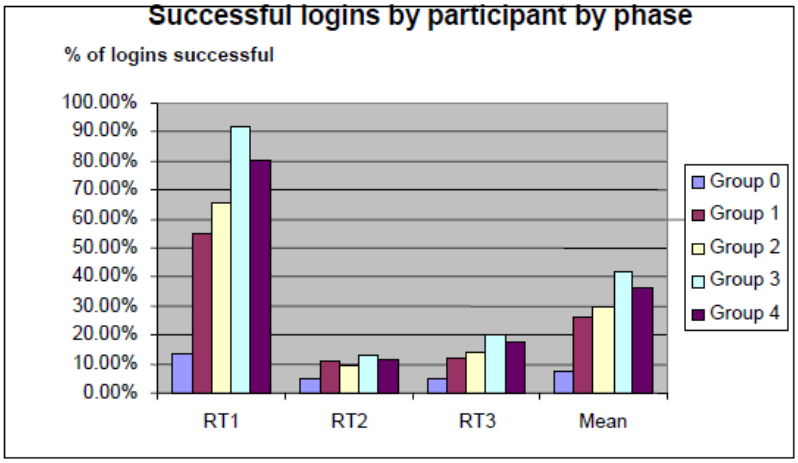


Figure 2.10: Password retention rates presented in (Moncur & Leplatre, 2007)

- *Group 0:* The percentage of successful login in RT1 was close to 14%, dropped to almost 5% after 2 weeks, and remained almost the same as RT2 after RT3.
- *Group 1:* The percentage of successful login in RT1 was close to 54%, dropped to almost 11% after 2 weeks, and remained almost the same as that of RT2 (12%) after RT3.
- *Group 2:* The percentage of successful login in RT1 was close to 66%, dropped to almost 10% after two weeks, and was slightly more than 14% after RT3.
- *Group 3:* The percentage of successful login in RT1 was almost 92%, dropped to 14% after two weeks, and was approximately 20% after RT3.
- *Group 4:* The percentage of successful login was almost 82% in RT1, dropped to 12% in RT2, and close to 18% after RT3.

The mean value for each group presented in Figure 2.10 is high because the login success percentages just after the training session (i.e. RT1) was taken into account. The mean login time was not reported. The mean registration time was not reported, as the passwords were issued by the system. The results reported in the study also revealed that most participants in groups (1-4) chose the correct target images in the wrong order, which decreased the login success percentages.

(D) Inferences

- The results demonstrated that the best mean login success percentage among all the groups was very low, i.e. 12% (Group 3) after RT2 and 20% (Group 3) after RT3.
- The high dropout rate of the participants (i.e. 64.91%) also made it difficult to obtain concrete results.
- The reported study did not provide any concrete information regarding the mnemonic strategies used by the participants, to enhance the memorability of multiple system-assigned object passwords.

2.6.2 Multiple Password Study with Click-based Passwords

Chiasson et al. (2009) reported a lab-based study with 65 university students to compare the memorability of multiple text passwords (MTP) and multiple click-based passwords (MCP).

(A) Study groups and configuration of cued recall-based system

The participants who took part in the study were randomly divided into two groups:

- Members of the first group were required to remember six text passwords created by them, during the registration stage;
- Members of the second group created six click-based passwords. Each password comprised of five click-points on an image, which were chosen by the participants themselves (Figure 2.11). Each participant was provided with six distinct images to create each of their six passwords.

Figure 2.11: MCP, 5 click points on one image

.

(B) User study

The lab-based study was divided into two sessions as given below.

- Session 1: All the participants were required to register with six passwords depending upon their group. After completing registration for each password, they were asked to login, once they had performed a distraction task. The login success for each of the participants in the session was reported as Recall1.
- Session 2: The second session was conducted two weeks after the first session, and only 26 participants took part in it. There were no practice sessions between the two sessions. The login success for this session was reported as Recall2. The authors did not report the number of participants in each group, who took part in session 2.

(C) Results

- Recall1: The mean login success percentage was 95% for MCP and 68% for MTP during the training session, when the participants logged in successfully in the first attempt. The mean login success for multiple attempts was 88% in case of MTP and 99% in the case of MCP. However, these are the mean login success percentages during the training session, just after the passwords were created. Hence, the results

do not reveal much in the context of the long term memorability of multiple click-based passwords.

- Recall2: The mean login success percentage was 38% for MCP and 30% for MTP, when the participants logged in successfully in the first attempt. The mean login success for multiple attempts was 70% for MTP and 57% for MCP, and this was found to be statistically insignificant.
- The registration time for each password is calculated as the summation of the time taken to create and confirm it (i.e. select the password and re-enter the selected password). The mean registration time was almost 43.9 sec in case of MCP and 43.5 sec for MTP. The mean login time for MCP was 15.1 sec during the training stage and 47.0 sec in the second session (i.e. two weeks after the training stage).

(D) Inferences

- The results of session 1 (Recall1) revealed that the short term memory for MCP is significantly better than MTP. However, the results obtained in session 2 (Recall2) revealed that the participants found it difficult to remember six click-based passwords.
- Since the participation in the second session dropped to 40% (i.e. only 26 participants took part), the results may not completely reflect the phenomenon of memory interference, when multiple click-based passwords are used.
- The performance of MCP (mean login success - 76%) is better than the multiple object passwords (best mean login success is approximately 42%) reported in Moncur & Leplatre. (2007). This might be attributed to the fact that multiple object passwords in Moncur & Leplatre. (2007) were issued by the system, which made it difficult for the participants to remember them, irrespective of employing a mnemonic strategy. Moreover, most authentication errors in Moncur & Leplatre. (2007) were due to order confusion, i.e. target images were chosen in the wrong order.

2.6.3 Multiple Password Study with Facial Images

Everitt et al. (2009) reported a web-based study with 100 university students, over a period of five weeks, to examine the memorability of multiple face passwords.

(A) Configuration of the RBGS

Each participant was assigned *x* number of passwords by the system. Each password comprised of five faces. During authentication, participants had to select the correct face from a sequence of 3x3 grids of decoy faces, at each step of a five step login process, as shown in Figure 2.12.

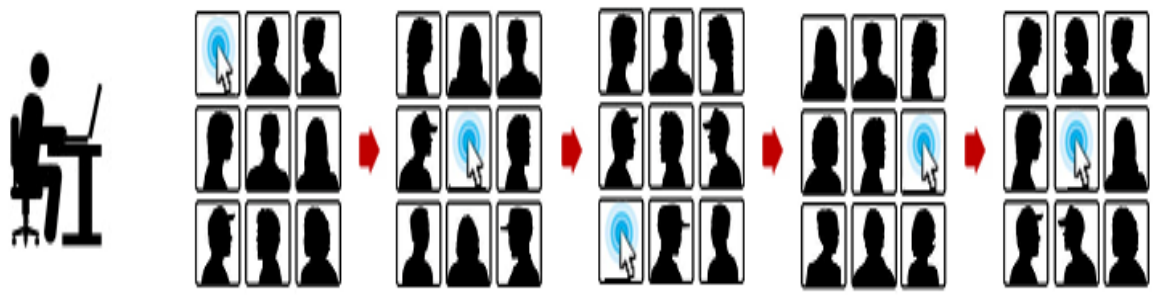


Figure 2.12: Login process reported in (Everitt et al., 2009)

(B) User study

A user study was conducted using a between-subjects design, where each participant was randomly assigned to one of the five conditions as discussed below, and further shown in Figure 2.13.

- *C1:* Participants used one face password (5 faces) once a week, for a period of five weeks.
- *C2:* Participants used one face password three times in a week, for a period of 5 weeks.
- *C3:* Participants used two face passwords. One password was used three times in a week for a period of 5 weeks, and the second password was used once a week for a period of 5 weeks.
- *C4:* Participants used four face passwords (20 faces). Each password was used once a week, for a period of 5 weeks. Hence all the four passwords were used at least once during the week.
- *C5:* Participants used four face passwords (20 faces). In this condition, a distinct password was used four times in a week.

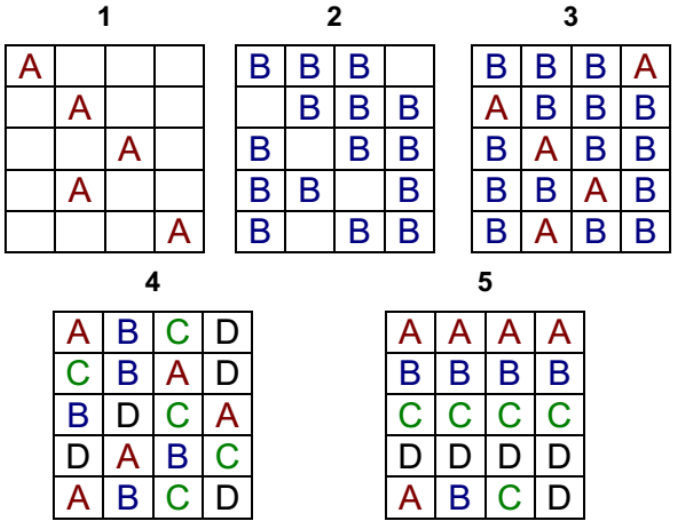


Figure 2.13: Overview of the five conditions reported in Everitt et al. (2009). Each row represents a week of the study and each alphabet represents a distinct password.

(C) Results

- The results revealed that the participants using a single face password once a week (C1) required more attempts to authenticate successfully, compared to the participants who used one face password thrice a week (C2). Hence, the frequency of the password usage would significantly affect the ease to authenticate successfully.
- The results also indicated that the participants accessing four passwords per week (C4) were ten times more likely to have an authentication failure, compared to the participants using a single password per week (C1). Thus, the memory interference occurring from the use of multiple face passwords would significantly affect the user's memorability.
- The results also demonstrated that the participants who were trained using multiple face passwords, each week during a month (C4) were four times more likely to have an authentication failure, than the participants who were trained using one graphical password per week (C5). Hence the password training pattern would significantly affect the ease of access. The failure rate in the case of C4 was 15.23%.
- In context to long term recall, it was found that the participants using one graphical password throughout a month could remember their password correctly after four months. But, the participants using four face passwords had problems remembering their passwords, due to memory interference (failure rate in C4: 14.29%). Thus, the long term recall is significantly affected, as the number of face passwords increase.

(D) Inferences

- The performance of multiple face passwords is better compared to the results reported in Moncur & Leplatre (2007) and Chiasson et al. (2009). However in Everitt et al. (2009) participants who were assigned to C4 used each of their four passwords, at least once in a week. This may have helped to recall the passwords in subsequent use.
- The long term recall of the face passwords also seemed promising, compared to the other two multiple graphical password studies discussed in the previous sections. However, the number of participants in each condition who took part in the long term recall study is not reported in Everitt et al. (2009).
- Overall, Everitt et al. (2009) demonstrated that memory interference and the frequency of password use would significantly affect the memorability of multiple face passwords.

2.6.4 Multiple Password Study comparing Three Image Types

Hlywa et al. (2011) have reported two lab-based studies, one with sixty participants and the other with twenty participants, to compare the memorability and efficiency of multiple RBGS passwords, using three distinct image types: faces; objects; pictures of houses.

(A) Configuration of the RBGS

The RBGS was implemented as a plugin for some popular open source websites. In the first study (S1), each authentication session comprised of six panels of 26 images (25 decoys and 1 target), involving a password space of 28 bits, which is same as that of text passwords with six random single-case letters (Figure 2.14). In the second study (S2), each authentication session comprised of five panels of 16 images (15 decoys and 1 target), involving a password space of 20 bits, which is more than a four digit PIN. Please refer to Appendix E for the theoretical password space computations.

(B) User Study

- *S1* used a between-subjects design and each of the 60 participants was randomly assigned to one of the conditions (image types). Each participant was issued three passwords of a single image type.

- *S2* used a within-subject design and each of the 20 participants was assigned two object and two face passwords.

In both of the studies (S1 and S2), a memorability test was conducted a week after the passwords were issued.

Figure 2.14: A single challenge set consisting of 26 images in Study 1 (Hlywa et al., 2011)

(C) Results

- *S1:* The login success rate was highest for object passwords (78.33%), followed by faces (63.33%), and the least for house images (38.33%). The mean authentication time was best for objects (31.03 sec), followed by faces (41.45 sec) and the slowest for house images (83.06 sec). The results demonstrated that objects performed the best both in terms of memorability and password authentication time, when multiple RBGS passwords are used.
- *S2:* The login success rate was highest for object passwords (95%), followed by faces (87%). The mean authentication time was best for objects (22.55 sec), followed by faces (35.96 sec). The results demonstrated that decreasing the password space (i.e. number of challenge sets) improved the login time, as well as the memorability of the respective passwords.

The authors concluded that though humans do possess special ability to process and memorize faces, as claimed by the psychological research studies, this claim may not hold true in the context of the authentication mechanisms.

(D) Inferences

- The performance of the object passwords in Hlywa et al. (2011) in terms of memorability was better than all the three studies discussed in the previous sections. The performance of the face passwords was similar to the results reported in Everitt et al. (2009)
- The login time of the object and face passwords reported in Hlywa et al. (2011) was either similar or more than all the other studies discussed in the previous sections. This can be attributed to the increase in number of challenge sets in Hlywa et al. (2011) compared to the other studies, which decreased the efficiency of the system, though the security offered in terms of theoretical password space was high.
- The lab-based study did not report any drop our rate and provides conclusive evidence regarding the superior memorability of multiple object passwords compared to the face passwords.

2.6.5 Guessability of faces Using Verbal Descriptions

Dunphy et al. (2008) examined the guessability of faces to verbal descriptions, when used as RBGS passwords. This is the only work regarding the recordability of RBGS passwords showing that research in this area has been limited.

(A) Configuration of the system

The authentication system comprised of five 3×3grids. Each grid contained 1 target face, 8 decoy faces and an audio description corresponding to the target face, as shown in Figure 2.15. In order to authenticate successfully, each participant was required to first hear the audio description corresponding to the target face, and then click on the face they believe was being described in each challenge set. Each participant was required to identify the target face in each of the five challenge sets using the audio descriptions to authenticate successfully.

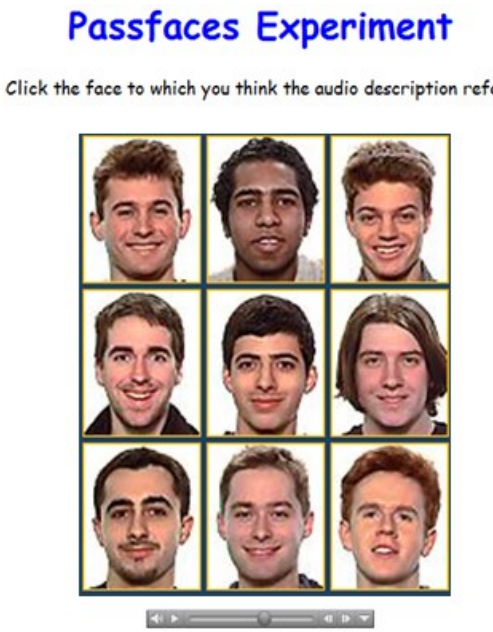


Figure 2.15: Challenge set consisting of 1 target face, 8 decoy faces and audio description corresponding to the target face (Dunphy et al., 2008).

(B) User Study

(Dunphy et al., 2008) explored the ability of 56 participants to associate verbal descriptions with target faces, across three conditions (i.e. three different approaches used to choose decoy faces for each target face) as given below.

- *Random groups:* The decoy faces for each target face were chosen randomly.
- *Visual groups:* The decoy faces were chosen based on their visual similarity with the target face. The contributors were asked to select visually similar decoys for a face.
- *Verbal groups:* The decoy faces were chosen based on their similarity with the verbal descriptions of each face.

The descriptions of the 45 faces (27 female and 18 male) were collected from 18 contributors (9 male and 9 females). Each contributor was asked to record verbal descriptions of 15 faces, randomly assigned to them. The contributors were asked to imagine a scenario that they were describing the faces to a friend.

(C) Results

Out of 158 authentication attempts made by 56 participants, 13 (8%) were successful. The results demonstrated that success rate was minimal, when the participants were presented

with visually and verbally grouped decoys, and maximum in the random selection of decoys. The authors concluded that the vulnerability of faces to verbal descriptions can be reduced by judiciously choosing the decoys.

(D) Inferences

- The verbal descriptions used in the guessability study were collected from nine male and female contributors, who described a set of faces given to them. They did not create or use the faces as their password. The contributors were also not told that the faces would represent passwords in real life.
- The contributors were given a set of faces to describe, which is analogous to a system where passwords are issued by the system. Hence, these results may hold true in a scenario where images are issued by the system, but may not be extended when images are chosen by the users.
- Lastly, the authors pointed out that the low success rate may be due to the experimental set up used in the study, rather than being a completely accurate reflection of the phenomenon of password guessability using descriptions.

Study	Mean Login Success %				Registration time (sec)	Login time (sec)		Comments
(Moncur & Leplatre, 2007) 5 system assigned object passwords	***Groups***	***RT1***	***RT2***	***RT3***	***Mean***	Not reported as passwords were issued by the system	Not reported		High dropout rate of 64.91% after RT1.
	(0) PINs	14	5	5	8				
	(1) OP (object password)	54	11	12	26				
	(2) OP + background color (bg)	66	10	14	29				
	(3) OP + mnemonic	92	14	20	42				
	(4) OP + bg + mnemonic	82	12	18	37				
(Chiasson et al., 2009) 6 user selected click based	***Conditions***	***Recall1***	***Recall2***	***Mean***		43.9 s (creating and confirming MCP)	15.1 in recall 1 to 47.0 in recall 2		High dropout (60%) in recall 2 stage.
	MTP- Multiple Text password	68	70	69					
	MCP- Multiple Click password	95	57	76					
(Everitt et al., 2009) system assigned face passwords	***Conditions***	***After five weeks***		***After four months***		Not reported as passwords were issued by the system	***Login time***		Number of participants in each condition who logged in after 4 months has not been reported.
	(1) 1 password used once a week	98.55%		0%			18.14 -20.76		
	(2) 1 password used thrice a week	99.65%		NA			13.78		
	(3) one password (A) once a week + one password (B) twice a week	A - 92 B - 98.06		NA			A- 24.58 B- 16.31		
	(4) 4 passwords, all used once in each week for four weeks	84.77		85.71			24.27 – 31.71		
	(5) 4 passwords, a distinct password used every week – 4	97.5 (Week 5 data only)		0			26.88 – 28.22		
(Hlywa et al., 2011) 2 or 3 system assigned passwords	***Conditions/ Image types***	***Study 1 (S1)***		***Study 2 (S2)***		Not reported as passwords were issued by the system	***S1***	***S2***	Evidence: multiple object passwords are better than multiple face passwords
	Objects	78.33		95			31.03	22.55	
	Faces	63.33		87			41.45	35.96	
	Houses	38.33		Not used (NU)			83.06	NU	

Table 2.5 Summary of the results reported in the existing multiple graphical password studies

2.7 Scope of the Thesis

2.7.1 Configuration of the Existing RBGSs

Each RBGS discussed in the Section 2.4.1 is further examined for details on the different configurations. The different aspects contributing to the configuration of RBGSs is further illustrated in Table 2.6 and 2.7. The different aspects are as follows:

- *Image type*, i.e. the type of image used as RBGS password
- *User involvement*, i.e. whether the target images are
  - Assigned by the system to the user, i.e. *system assigned*
  - Selected by the user from a collection provided by the system, i.e. *user selected*
  - Supplied by the user themselves
    - Drawn by the user themselves, i.e. *drawn*
    - Photographs uploaded to the system by the user, i.e. *user provided*
- *Configuration of the challenge set*
    - Number of decoy images in a single challenge set
    - Number of target images in a single challenge set
    - Number of challenge sets for each authentication session
    - Use of constant decoy images for a given target image
    - Ordered selection of target images
    - Theoretical password space- number of all possible passwords in the password space (calculations are shown in Appendix E).
- Approach to select decoy images for each target image in the challenge set
    - *Visually similar* decoy using a similarity measure
    - *Randomly* selecting the decoy
    - Decoys selected from a *specific semantic category*

All the details in each of the relevant fields in Tables 2.6 and 2.7 are obtained from the literature. Where the information was not available 'Unknown' is used.

System	Image Type	User Involvement	Target images	Decoy images	Challenge sets	Constant decoys	Ordered login	Decoy Selection	Theoretical space
PassFaces (Real, 2004)	Faces	System assigned	4	8	4	8	No	Random selection	12.67 bits
Faces (Davis et al., 2004)	Faces	User selected	4	8	4	8	No	Random selection	12.67 bits
(Everitt et al., 2009)	Faces	System assigned	5	8	5	8	No	Random selection	15.72 bits
(Hlywa et al., 2011)	Faces Object	System assigned	S1: 6 S2: 5	S1: 25 S2: 16	S1:6 S2: 5	Unknown	No	Random selection	S1: 27.86 bits S2: 20 bits
Story (Davis et al., 2004)	Object	User selected	4	5	1	5	Yes	Random selection	11.56 bits
(Moncur & Leplatre, 2007)	Object	System assigned	4	6	1	Unknown	Yes	Each decoy from one distinct category other than target category	12.29 bits
VIP 1 and 2 (Angeli et al., 2005)	Object	System assigned	4	6	1	0	Yes	Category other than target image Random selection	12.29 bits
VIP 3 (Angeli et al., 2005)	Object	System assigned	8	12	1, only 4 targets	0	No	Category other than target image Random selection	10.82 bits
PassImages (Charrau et al., 2005)	Object	User selected	6	Max 25	4, but all 5 targets displayed	Unknown	Yes	NA	26.90 bits
Dejavu (Dhamija & Perrig, 2000)	Art	User selected	5	20	1	Unknown	No	Random selection	15.6 bits

Table 2.6: Configuration of RBGS used in existing studies (Part 1)

System	Image Type	User Involvement	Target images	Decoy images	Challenge sets	Constant decoys	Ordered login	Decoy Selection	Theoretical space
Doodle (Renaud, 2009a)	Doodle	Drawn by user with pen on paper	4	15	4	15	No	Random selection	16 bits
Mikon (Renaud, 2009)	icons	Drawn by using a tool	4	15	4	15	No	similarity algorithm	16 bits
Cognitive (Weinshall, 2006)	NA	System assigned	30-60	40-50	several	NA	No	Unknown	10-73 bits depending on the configuration
Convex hull (Wiedenbeck et al., 2006)	Icons of software applications	User selected	5	40-119	1 (3-5 rounds)	No	No	Unknown	20.22 to 32 bits
Use Your Illusion (Hayashi et al., 2011)	Target image distorted	User provided	3	8	3	Yes	No	Random selection	10 bits
WYSWYE (Khlot et al., 2012)	Distinct visual object	User selected	4	21-24	1	Unknown	No Place in correct position	Unknown	13.62 bits
(Tullis & Tedesco, 2005)	Personal photographs	User provided	8-20	15	1 with 2-5 targets	Unknown	No	Random selection	Will depend on number of targets
AWASE (Takada et al., 2006)	Personal photographs	User provided	4	8 or 9	4	Unknown	No	Random selection	12.67 bits
Photographic authentication (Pering et al., 2003)	Personal photographs	User provided	*X (not known)*	3	10	0	No	Unknown	20 bits

Table 2.7: Configuration of RBGS used in existing studies (Part 2)

2.7.2 Configuration of the RBGSs to be used in the Thesis

The configuration of the RBGS to be used in this thesis is based upon the different aspects extracted from the review of existing RBGSs (Tables 2.6 and 2.7). This is further discussed below.

(A) User involvement

In this context, it was found that seven studied RBGSs used system-issued passwords, six had them selected by the user from a collection displayed by the system, four had them provided by the user and in two systems users were required to draw sketches of the target images. Tullis & Tedesco (2005) and Renaud (2009a) showed that user involvement during the password registration stage, i.e. selection of target images has a positive impact on the memorability. Tullis & Tedesco (2005) used personal photographs provided by the user and Renaud (2009a) used hand-drawn doodles; both password types were required to be uploaded to the system, which would make the registration process time-consuming. The existing studies reported in Dhamija & Perrig (2000) and Charrau et al. (2005) have shown that the memorability of images selected by the user themselves is also quite high (> 90%).

Hence in this thesis, all the target images are selected by the user themselves from a collection displayed by the system.

(B) Image type

Tables 2.6 and 2.7 shows that four RBGSs used faces as the password, six used objects, two used icons, and three used personal photographs. Doodle, art and image of a distinct visual object were also used as RBGS passwords. Table 2.3 shows that multiple password studies have been conducted with faces and objects only.

Davis et al. (2004) suggested that the choice of face passwords is affected by the race and gender of the user, as well as the attractiveness of the faces. To eliminate such biases, face passwords should be issued by the system. Moreover, Perrig et al. (2003), Tullis & Tedesco (2005) and Takada et al. (2006) do not report filtering the photos provided by the user. Hence personal photos are likely to be easily related to the user, which would be too insecure to use in a real-life setting. Clear guidelines in terms of ethics, privacy as well as security have to be provided to the user, so that such photos are suitable for authentication purposes.

Hence in this thesis all the user studies were conducted with four different image types: Mikon, doodle, art and object.

(C) Configuration of the challenge set and Theoretical password space

In this context, RBGSs can be divided into two groups. One approach consists of a single challenge set with multiple target images, which was used by eight of the 19 systems presented. This approach can be further refined by the target image selection restricted to a specific order or the order being irrelevant. Four of the systems had ordered selection and the rest have unordered selection. Davis et al. (2004) and Moncur & Leplatre (2007) found that most users recognised the target images correctly, but in wrong order, which led to many login failures. Hence, they recommended that ordered selection of target images should be avoided. In the second approach, which is represented by the 10 remaining systems, a single target image is displayed in multiple challenge sets, except for Charrau et al. (2005).

For RBGSs, the theoretical password space would depend upon the number of rows, columns challenge sets and type of selection (order/unordered) in a given system (Appendix E). In most systems, except Charrau et al. (2005), Everitt et al. (2009) and Hlywa et al. (2011) , the theoretical password space is either equivalent (13 bits) or slightly more (16 bits) than that of a 4 digit PIN (13 bits). The results reported in Hlywa et al. (2011) showed that decreasing the password space would increase the usability, i.e. increased memorability (less number of target images to remember) and decreased login time (less number of challenge sets).

The default configuration used in this thesis comprised of four challenge set, each comprising of 1 target image and 15 decoy images, without any ordered selection. This configuration has a theoretical password space of 16 bits, i.e. slightly more than a four digit PIN (13 bits).

(D) Approach to select decoy images for each target image in the challenge set

In most RBGSs (10 out of 19) the decoy images were chosen randomly for each target image from a collection which is stored in the system. In VIP (Angeli et al., 2005) the decoy images were chosen from any category except the category that the target image in the challenge set belonged to. Moncur & Leplatre (2007) used an approach where decoy images for each target image were chosen from a distinct category, except that of target image. Renaud (2009) used a similarity algorithm to choose the decoy images, which were visually similar to the target

image. In the context of the visual similarity, Nelson (1979) had reported that if two images are visually similar then the picture superiority effect could be counteracted.

Hence in this thesis, the decoy images for each target image are randomly selected for a collection stored in the system. This would also help to preserve consistency across image types in a scenario where many visually similar decoys for a target image are not available, or it is difficult to categorise the target images (for e.g. art images).

2.7.3 Authentication Environment and Threat Model

In order to define the threat model, it was necessary to identify the possible threats. In this context, Angeli et al. (2005) have proposed three dimensions to assess the security of a RBGS, as discussed below.

- *Guessability:* the probability an attacker can guess the user's password.
- *Observability:* ability of an attacker being able to observe the authentication process, i.e. able to actually see the target image as the user enters it.
- *Recordability*: ease with which a user can record the target images, thereby making it easier for an attacker to steal and eventually impersonate the user.

These dimensions were further extended in Renaud (2007) to include *analysability*, i.e. exploiting implementation details/ bugs in a software and *resistibility*, i.e. auxiliary attempts to secure the system, for e.g. a lockout policy after a certain number of failed login attempts.

To put the guessability attacks reported in this thesis into context, a threat model is discussed in this section, taking into account each of the dimensions listed above. The model represents the attacker under consideration in this work, the information they have access to and their abilities. An *attacker* will attack the authentication stage in a RBGS. The attacker is classified as *casual,* i.e. an individual who is not a skilled hacker, for the purpose of this thesis. It is assumed that an attacker has already identified a target user, knows their username and has access to the user recorded prompts of the target images forming a password, i.e. description of each target image forming a password. The aim of the attacker is to authenticate by impersonating the target user. In order to impersonate a target user, an attacker must identify all the target images in each challenge set during the authentication session, using the user recorded prompts (descriptions). The attacker does know that a user has four target images

and each target image has to be selected in each challenge set. Hence the attacker is aware that each challenge set has one target image and 15 decoys. But, the attacker does not have any idea about the decoy image selection approach. It is also assumed that an attacker does not have any knowledge about the likes, dislikes, personal life of the target user. Hence educated guessing attacks are excluded.

The environment being considered in all the studies reported in this thesis is a local authentication environment (including a web environment), which can be physically observed by an attacker. Moreover, the user is required to select the target images using a mouse on the computer screen, during the password creation as well as the authentication stages. Since the RBGS is being used in a local authentication environment, attacks which involve intercepting authentication communication between the client and server (e.g. man in the middle attacks) as well as other observation attacks (e.g. shoulder surfing) are excluded.

Additionally, an attacker does not attempt offline guessing attacks such as dictionary attacks. In order to launch an offline dictionary attack in case of RBGSs, the form of communication between the client and the server, when a user selects a target image needs to be considered. This can take the form of: hash of the image; an identifier for the image; a hash of the identifier for the image; a temporary identifier for the image (Biddle et al., 2009). Hence this attack would rely on copying the communication, and establishing the connection between the information that is being sent over the network and the target image. As the thesis considers local authentication environment and all communication attacks are excluded, offline guessing attacks are also outside the scope of this thesis.

It is assumed that the attacker cannot establish the semantic category for each image in the challenge set and won't exploit any bias or pattern in the target images selected by the user. It is also assumed that an attacker does not exploit any bugs in the implementation of the mechanism. Hence attacks based on analysability are also outside the scope, and not considered further in this thesis. The RBGS used in this thesis will implement a strike out policy, i.e. the attacker is permitted a limited number of unsuccessful login attempts, before being locked out of the system. Hence online dictionary attacks are also considered out of scope.

The threat model applies to both the guessability studies GS1 and GS2, which are presented in Chapters 4 and 5 respectively. The model presented here will be further discussed in

Chapter 4 (Section 4.1.2) in the context of the guessability study. The threat model also applies to GS3 presented in Chapter 6, except that attackers do not have access to the descriptions, i.e. prompts recorded by the user.

2.7.4 Summary of the Scope

Configuration of RBGS

- Four image types, viz. Mikon, doodle, art and object are used as RBGS passwords.
- The target images are selected by the user from a collection presented by the system.
- The authentication is a four step process. At each step, a challenge set comprising of 1 target image and 15 decoy images will be presented to the user. The decoy images are chosen randomly from the collection presented to the user during registration. Further details about the decoy image selection approach is presented in Chapter 3 (Section 3.3.2)

Guessability Attacks

- The authentication environment is local and the user inputs are observable.
- The purpose of an attacker is to impersonate a legitimate user. The attacker does so by using prompts/ descriptions of the target images recorded by the user. Further details about the attack and its context will be discussed in Chapter 4 (Section 4.1.2).
- The attacker does not have any knowledge about the decoy selection approach and how the images are presented in each challenge set.
- The attacker won't perform offline brute force or guessing attacks.
- The attacker does not have any information about the likes/ dislikes of the user; hence educated guess attacks are out of scope.
- The attacker does not exploit biases in user selection.
- The attacker won't perform any observation attacks (shoulder surfing attack) or communication attacks (man in the middle attack).
- The attacker won't perform any attacks by exploiting the bugs in the implementation.
- The attacker is permitted a limited number of login failures, before being locked out of the system.

2.8 Conclusion

The survey of the existing research in the field of GASs (Sections 2.2 – 2.4) showed that, most studies have focused on the usability of a single password. To our knowledge, in the last fifteen years, only four studies (3 with RBGS and 1 with CCP) have examined the use of multiple graphical passwords. The review of these four studies (presented in Section 2.6) shows that the memorability statistics (login success rate) reported in each of these studies, except (Everitt et al., 2009) and (Hlywa et al., 2011) is much less than the ones reported in the single password studies (Chapter 2, Table 2.2 and 2.3). The review also highlighted that two studies, (Moncur & Leplatre, 2007) and (Chiasson et al., 2009) had a very high dropout rate during the password retention stage, which made it difficult to obtain conclusive results. Hence there is a need to conduct user studies with a considerable participation rate and a suitable experimental protocol, to better understand the memorability of multiple passwords in RBGSs. In this context, two usability studies (US1 and US2) are reported in Chapters 3 and 5 respectively, which have examined the memorability of multiple RBGS passwords.

The review presented in Section 2.4 shows that it is an untested assumption that graphical passwords will be resistant to being written down or verbally described. Moreover, the extent to which graphical passwords can be guessed using written or verbal descriptions have not been assessed yet, and the studies investigating the topic of description in the context of RBGS passwords are sparse. Only a single study (Dunphy et al., 2008) in the field has explored the vulnerability of face passwords to verbal descriptions. This may be attributed to the lack of information regarding, to what extent users will record their RBGS passwords and the mechanisms they would use to record such passwords. However, it is worth investigating the extent to which different image types used as passwords in RBGSs are guessable using their corresponding descriptions. In this context, two guessability studies (GS1 and GS2) are reported in Chapters 4 and 5 respectively.

An important limitation identified in this chapter is that published details, methodologies and reported results in each study vary greatly. Hence the usability and security studies lack consistency, for e.g. metrics reported in each study is different and in some cases the same metrics are calculated/ interpreted in different ways. Therefore, it is difficult to compare the results reported in the existing literature. In this context, some future research directions are proposed in Chapter 7 (Section 7.3).

The survey presented in Chapter 2 does not include a review of the empirical studies discussed in the later Chapters (i.e. 3 to 6) of this thesis. Tables 2.1-2.7 do not include any results reported in the later chapters of this thesis. However, Table 7.1 (Chapter 7) provides a summary of the results obtained from the usability studies (US1, US2 and US3) reported in this thesis, together with the results presented in Table 2.5 (i.e. multiple password studies).

Chapter 3
Usability of Multiple RBGS passwords

The rapid increase in the technologies requiring user authentication has increased the number of passwords that users have to remember. In this chapter, a user study (US1) comparing the usability of multiple RBGS passwords with four different image types: Mikon, doodle, art and objects is presented. This chapter addresses Objective 1, which has been discussed in Section 1.4 of Chapter 1. The contents of the chapter have been published in the proceedings of the 14th IFIP TC.13 International Conference on Human-Computer Interaction - INTERACT 2013.

3.1 Introduction

Users will need to remember and use multiple RBGS passwords in the same way that they currently use multiple text passwords, if RBGS passwords were to become widely adopted. The literature presented in Sections 2.4 and 2.6 has identified that most prior work with RBGSs have focused on the usability of a single password, except, (Moncur & Leplatre, 2007; Everitt et al., 2009; Hlywa et al., 2011). Hlywa et al. (2011) used the images of objects, faces and houses as RBGS passwords. Everitt et al. (2009) used faces as the visual cue, whereas Moncur & Leplatre (2007) used objects (images of flowers, food, sculptures, nature etc.). The usability of multiple RBGS passwords with other image types such as Mikon, doodle and art, when the target images are selected by the user themselves has not been explored yet. This chapter presents a usability study (US1), which was conducted with 115 subjects who used multiple RBGS passwords over a period of eight weeks. The study compared the usability of four different image types: Mikon, doodle, art and everyday object, when used as passwords in RBGS. The motivation of the study was to investigate, '*whether multiple RBGS passwords, each composed of the same image type, are memorable, in a given experimental setting*?'

3.1.1 Contributions

The main contributions of this chapter are as follows:

- The usability of four different image types are assessed, in terms of their effectiveness and efficiency, when multiple passwords of each type are used;

- The results obtained in US1 are compared to the existing studies that have reported the usability of multiple graphical passwords.

3.2 Image Types Used in the Thesis

All the usability and guessability studies reported in this thesis were conducted using the four image types discussed below.

- Mikon: These are icon-like images which are drawn using a tool called the *Mikon engine*[2] (Renaud, 2009). Figure 3.1 presents a sample of four Mikon images. The results of a single password study reported in Renaud (2009), which has been already discussed in Section 2.4, demonstrated the memorability of Mikon passwords. However, multiple password studies with this image type have not been reported in the existing literature.

  In this thesis, subjects in all the usability studies (Chapters 3, 5 and 6) did not draw the Mikon images. The subjects were required to select the images to form their respective password, from a collection presented by the RBGS. The Mikon collection reported in Poet & Renaud (2009a) was used in all the studies reported in this thesis.

Figure 3.1: Sample Mikon images

- Doodle: These images have been evaluated in a number of studies (Renaud, 2005; Poet & Renaud, 2009b; Renaud, 2009a). Figure 3.2 presents a sample of four doodle images. The results reported in the existing single password studies, which have been already discussed in Section 2.4, demonstrated the memorability of doodle passwords. However, none of the existing works have explored the performance of doodle images in multiple password studies. In this thesis, the

[2] http://www.mikons.com/create/machine/

subjects in all the usability studies (Chapters 3, 5 and 6) did not draw the doodle images. They were asked to select the doodle images to form their respective passwords, from a collection presented by the RBGS. The doodle collections that has been used in (Poet & Renaud, 2009a; Poet & Renaud 2009b), was also used in all the studies reported in this thesis.

Figure 3.2: Sample doodle images

- Art: These images were collected from two free media repositories: FreeFoto[3] and Wikimedia commons[4], where images are licensed under Creative Commons, i.e. students can use the downloaded media for educational purposes. The images comprised of paintings of different styles, such as cubism, abstract and modernism (Figure 3.3). Dhamija & Perrig (2000) used random art images to investigate the usability of a single password. The usability of art images have not been explored, in the context of the use of multiple passwords.

Figure 3.3: Sample art images

- Object: These comprised of images of food and drinks, sculpture, buildings, sports and leisure activities, which were again collected from FreeFoto[3] and Wikimedia commons[4] websites (Figure 3.4). Many studies conducted in the field of RBGSs have used this image type, as shown in Table 2.3 (Section 2.4). Hence, we chose to use object images, and compare their performance with the other three image types in a number of multiple password studies.

[3] http://www.freefoto.com

[4] http://commons.wikimedia.org/wiki/Main_Page

Figure 3.4: Sample object images

Davis et al. (2004) suggested that the choice of face passwords is affected by the race and gender of the user, as well as the attractiveness of the faces. To eliminate such biases, face passwords should be issued by the system. According to the results reported in Renaud (2009a), system-issued passwords are difficult to remember. Additionally, the results of the study reported in Everitt et al. (2009) gives evidence that users do have problems remembering multiple face passwords (as opposed to remembering a single face password), when these are issued by the system. Hlywa et al. (2011) also showed that multiple object passwords are more memorable than multiple face passwords, when these are issued by the system. Since the main aim of US1 is to examine the effectiveness and efficiency of multiple image passwords selected by the user themselves, the use of faces was omitted. Moreover, to alleviate ethical and privacy issues, personal photographs were not used.

3.3 Design of the RBGS

Four online-study websites (RBGS prototypes) were developed for the usability study because the aim was to compare the usability of four distinct image types (Mikon, doodle, art and object). Hence each of the RBGS prototypes used a distinct image type as the password. Each RBGS password comprised of four target images of the same image type.

Each RBGS had a distinct web address. The home page of the website comprised of four hyperlinks (each hyperlink corresponding to one password, see Figure 3.5, Screen 1): *My Jokes; My Movies; My News; My Status*. In each RBGS, all the hyperlinks except *My Status* used a collection of 150 distinct images of the same image type. My status had a collection of 150 images (50 images each from *My Jokes, My Movies and My News* respectively). Table 3.1 gives an overview of the image collection used in each of the hyperlinks. Example images for each category are shown in Appendix A.

Upon successful authentication, the subjects were required to post content in the respective link upon successful authentication. The content can be any text that the subjects would like

to post, for example, jokes in *My Jokes* section, movie reviews in *My Movies*, any news or something similar in *My News* and status updates (similar to the social networking application Facebook) in *My Status*. This task ensured that the subjects had a context to use in differentiating their four passwords. It was decided to use distinct image collections of the same image type for most links, to ensure that the subjects do not select the same target images to create all their passwords.

	Category 1 **MyJokes (MJ)** **150 images**	**Category 2** **My Movies (MM)** **150 images**	**Category 3** **My News (MN)** **(MN)-150**	**Category 4** **My Status (MS)** **(MS)-150**
Mikon	Colourful images (No annotations)	Black and white.(No annotations)	Colourful as well as Black and white images with annotations	50 images each, from MJ, MM and MN
Doodle	Black and white (No annotations)	Black and white - not same as MJ doodle (No annotations)	Black and white with annotations	50 images each, from MJ, MM and MN
Art	Abstract paintings	Paintings different from MJ (cubism/ modernism etc.)	Paintings different from MJ and MM (cubism/ modernism etc.)	50 images each, from MJ, MM and MN
Object	Food and drinks	Sculpture and buildings	Sports and leisure	50 images each, from MJ, MM and MN

Table 3.1: Image collection used for each link

3.3.1 Registration Process

Figure 3.5 shows the registration screens of the RBGS prototype developed for US1. Each subject could register in each of the hyperlinks by entering a username (Screen 2, Figure 3.5), and then selecting four target images to form a single password (Screen 3, Figure 3.5), from the collection presented by the system.

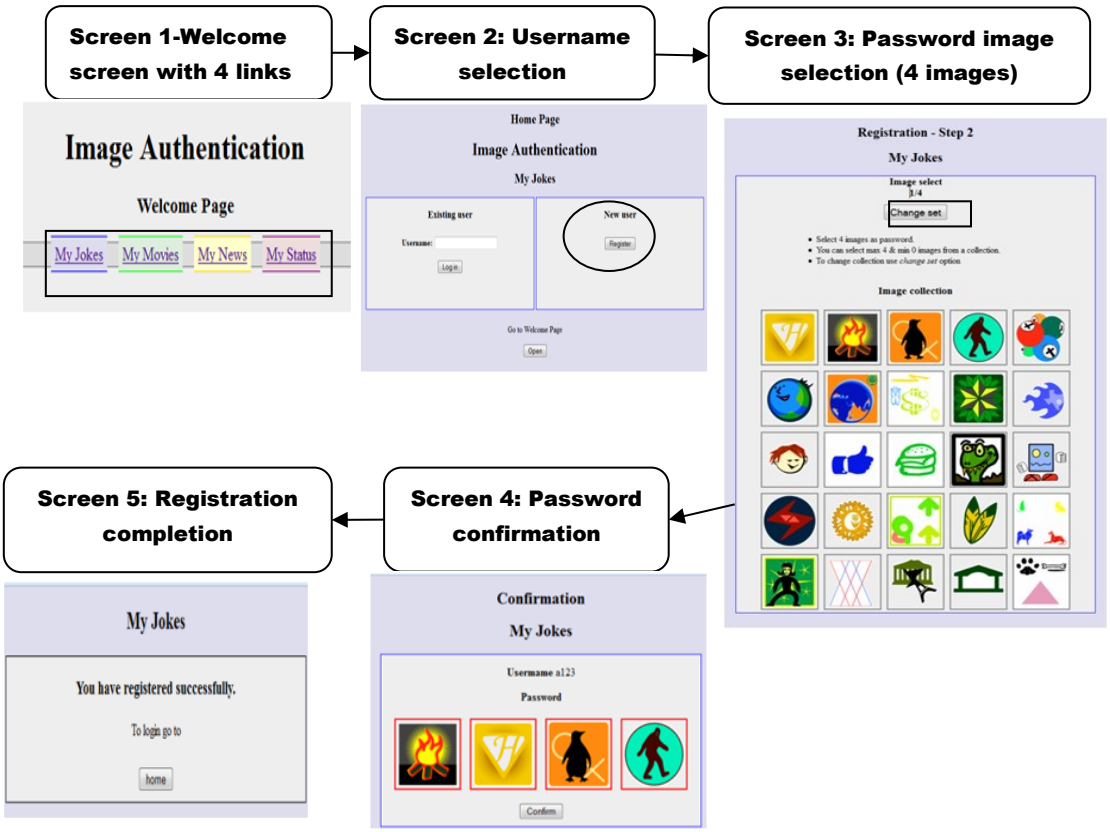


Figure 3.5: Registration screens in the RBGS prototype

Each hyperlink had a collection of 150 images, presented on the screen as six sets of 25 images in the form of 5×5 grids (Screen 3 in Figure 3.5). A decision was made to display the images in a 100 × 100 pixels resolution to minimize vertical browsing, and accommodate an entire challenge set (authentication process) even in the smaller screens. This decision ensured that the subjects can see all the images at the same time during the authentication process. The subjects could browse from one set to the other using the 'change set' button on the web page (Screen 3 in Figure 3.5). The subjects could choose all the four target images from a single set, or each one from a different set (browse multiple sets to look through the whole collection). The RBGS website was designed such that each of the subjects would use a different collection of the same image type, while registering for each of their passwords. For example, each subject in the Mikon condition created four passwords, i.e. one password selected from the image collection of the first link (*My Jokes*), one from the second link (*My Movies*) and so on. The image archive in each hyperlink of the Mikon website contained Mikon images only. The image collection used for each hyperlink was the same for all the subjects. However the images displayed in the sets were randomised for each registration

session, i.e. a subject will use the same image collection, but may not see the same images in each set as the other subjects..

3.3.2 Authentication Process

Upon accessing the RBGS website, the subjects had to select each target image (the ones they selected during registration) from a sequence of 4×4 grids at each step of a four-step procedure. The authentication screens for the RBGS are presented in Figure 3.6. Each target image was displayed with fifteen decoy images forming one challenge set. In order to successfully complete an authentication session, the subjects were required to select the target image in each of the four challenge sets.

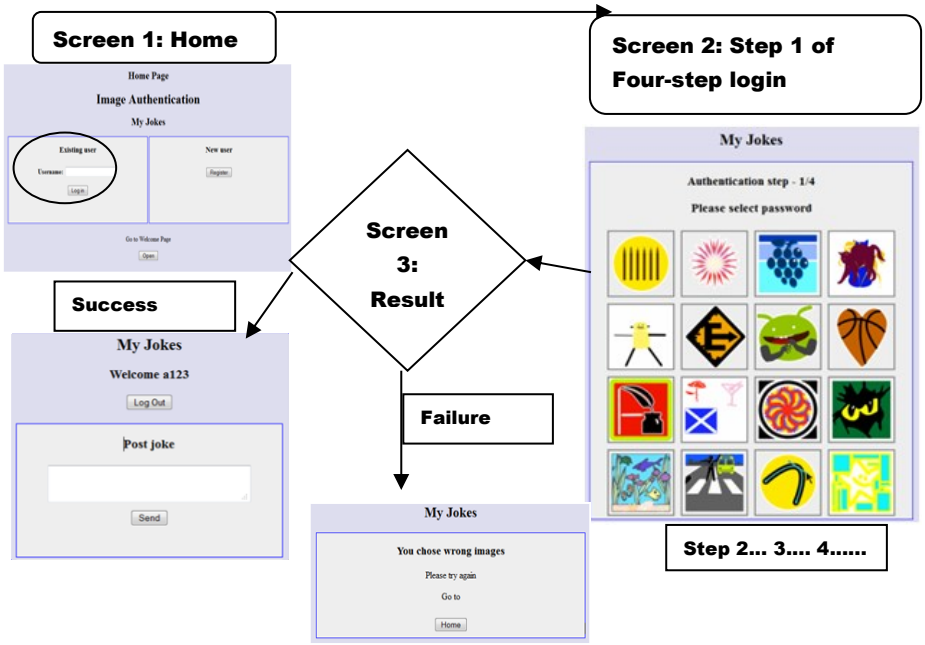


Figure 3.6: Authentication screens in the RBGS prototype

The challenge set configuration developed for the user study is given below:

- The decoy images corresponding to each target image forming a password for all the hyperlinks were chosen randomly, from the collection of 150 images (specific to the hyperlink).
- The decoy images for each of the four target images were distinct and never repeated. The fifteen decoy images for each of the four authentication steps did not include any of the target images;

- The fifteen decoy images for all the authentication steps were fixed after registration (once a user had selected the four target images), to ensure that an intruder would not be able to guess the target images, merely refreshing the web page. Hence the challenge sets for a user did not change, when the web page was reloaded by using the refresh button of the browser;
- If the subjects at any step during the authentication procedure selected a wrong target image, then they would never get any of their registered target images in the subsequent steps. In such a scenario, sixteen decoy images (without the target image) different from the original challenge set was displayed.
- The result of an authentication session was displayed only after the last step (i.e. step 4) of the procedure. In case of three continuous/non-continuous failed login attempts, the subjects were automatically reminded of their password. The reminder was given only for the first 20 login attempts (i.e. week-1) of each password. Once the subjects had authenticated successfully, the task was to post any information which could be seen by other users using the system, who could like it.

3.4 Usability Study (US1)

The study reported in this chapter investigated the usability, i.e. effectiveness in terms of the memorability and the efficiency in terms of mean registration time and the time taken for successful authentication attempts, of multiple passwords in RBGS.

3.4.1 Recruitment of the Subjects

In order to recruit the participants (subjects) for this study, emails were sent to the first and second year student email distribution lists in a university. The mail comprised of:

- The main objectives of the study;
- Summary of the registration and authentication tasks involved in the study ;
- Type of study – Lab-based/Online/Hybrid;
- Approximate duration of each task;
- If there is a requirement to meet the experimenter and details regarding task schedules as well as face to face meetings with the experimenter (if any);

- Mode of communication that will be used during the period of the experiment;

A total of 208 subjects volunteered to take part in the study, by responding to our email. These subjects were asked to answer a pre-study web-based questionnaire. Out of 208, 150 subjects answered the pre-study questionnaire. 58 subjects who did not respond to the questionnaire were not contacted further for the purpose of this study. Out of 150, 10 subjects were randomly approached to take part in a pilot study. The aim of the pilot study was to ensure that:

- Subjects could use the instructions easily, to successfully register and authenticate in our RBGS prototypes;
- They could understand their tasks from the documentation provided to them and emails that were framed for the purpose of the study;
- The post-study questionnaire is interpreted correctly, and they could understand the terminologies used to frame the questions;

Once the pilot study was completed, an email was further sent to the remaining 140 subjects, to confirm their participation. 115 subjects responded to the mail and agreed to participate by signing the participation consent form, once they were given all the relevant documentation (registration and login process, detailed description of the task for each week). The remaining 25 subjects either did not respond to our email or decided not to take part in the study. Of the 115 subjects who took part, 10 subjects had a very low participation rate (i.e. did not follow the experimental procedure), and five subjects had to withdraw due to other circumstances. Hence the usability study was completed by 100 subjects. The dropout rate of this study, i.e. subjects who did not complete the study after confirming their participation, was 13.1 %, which is much less than the dropout rate reported in Moncur & Leplatre (2007). For the purpose of the research reported in this Chapter, demographic information of all the subjects who took part (100 completed + 15 drop-outs) has been reported. However, the results reported in Section 3.5 do not include the subjects who did not complete the study (15 drop-outs). Table 3.2 summarises the responses received during each stage of the recruitment process.

Stage	Subjects responded	Comments
Email invitation	208	
Pre-study questionnaire	150	58 no response
Pilot study	10	Randomly approached 140 Subjects left
Confirmation to participate	115	25 no response
Study completed	100	15 dropped out

Table 3.2: Recruitment information for the user study US1

3.4.2 Demographic Information of the Subjects

115 undergraduate students (30 female and 85 male, age range: 20-24 years) took part in US1. They were enrolled in to different degree programmes as follows:

- Mechanical Engineering – 22;
- Electrical Engineering- 19;
- Aerospace Engineering- 25;
- Computer Science- 24;
- Electronics Communication Engineering- 25.

None of the subjects were experts in usable security or studying this topic as a part of their curriculum. Ethics approval (ethics no CSE00864) was granted by the college ethics committee to conduct US1.

3.4.3 Study Framework

A three stage user study was designed which comprised of:

(A) A *pre-study survey*, evaluating the demographics of the subjects and their current password strategies;

(B) An *8-weeks online study*, where subjects authenticated using multiple RBGS passwords;

(C) A *post-study* questionnaire.

The experimental framework has also been summarised in Figure 3.9.

(A) Pre-study survey

A pre-study questionnaire was conducted in the form of an online survey, to decide the number of image passwords that will be used in US1. A total of 150 subjects took part in the survey, which included the 115 subjects who took part in US1 (Table 3.2). The online survey was designed using the Grounded theory (Strauss & Corbin, 1990). The steps followed to design the survey are illustrated in Figure 3.7 and listed below:

- The key point, i.e. the aim of the survey was identified;
- The various categories and factors related to the key point was identified ;
- The parameters or concepts to be examined under each category were identified to frame the questions (refer to Appendix B.1 for the questionnaire);
- Finally, the results obtained from the survey were analyzed.

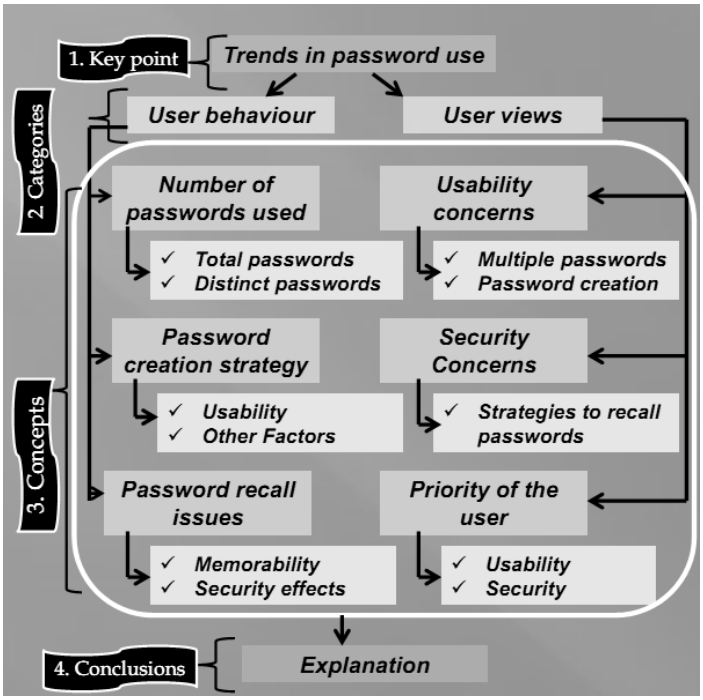


Fig.3.7: Framework used to design and analyze the pre-study survey

The results of the survey revealed that subjects used as many as 4-6 distinct passwords in their day to day life. This enabled us to make the choice of using four passwords (lower bound) in US1. In relation to the password creation strategies employed by the subjects, we received 500 responses in total, as shown in Table 3.3. Each survey respondent (subject)

chose one or more strategies from a list given to them (Column 2 in Table 3.3). For example, we received 93 responses favoring strategy S1. Since the total number of respondents was 150, 62% (93/150) of the respondents (subjects) chose this strategy.

Strategy	List of strategies to create text passwords	Responses
S1	Same passwords for different accounts	93/150
S2	Similar password for different accounts	128/150
S3	Different combination of the letters in a master password	95/150
S4	Password somehow personally related to you	107 /150
S5	Use a random password and write it down	77/150

Table 3.3: Responses in relation to password creation strategies

The responses in the context of the password creation strategies employed by the subjects show that: 62% used same passwords for different accounts (S1); 85.33% reported using similar passwords for all their online accounts (S2); 63.33 % used different combinations of letters in a master password (S3); 71.33 % used passwords that could be linked to their personal likings (S4).

The results obtained from the other questions highlighted that 80% of the subjects forget their passwords, either due to the strategy used to aid memorability, or constraints imposed by the system, while creating the text passwords.

(B) Eight weeks online study

US1 used an independent measure style of experimental design with four conditions (equal number of subjects in each condition) as given below.

- *Mikon:* Register with four Mikon passwords and authenticate using them.
- *Doodle:* Register with four doodle passwords and authenticate using them.
- *Art:* Register with four art passwords and authenticate using them.
- *Object:* Register with four object passwords and authenticate using them.

Each subject was randomly assigned to only one of the four conditions. The subjects in the Mikon condition used Mikon images as their password. They registered with four Mikon passwords and authenticated using them. Each password comprised of four Mikon images. Hence, each subject was required to remember 16 images in total. Similarly, the subjects in

the doodle, art and object conditions used the respective types as the passwords. The subjects were instructed to register one password per day, i.e. not to register all the passwords in the same day. Each subject's log details consisted of the day, date and time (timestamp) as well as the time taken to complete the registration. Hence, the log details of all the subjects helped us to review their registration sessions and eliminate any subject, who did not follow the instructions that were given to them. These subjects are reported as drop-outs in Table 3.2.

Once the subjects signed the consent form to take part in the experiment, they were given a task information sheet. The task sheet contained all the information on the steps to register with the system, i.e. select four target images to create a password, and steps to authenticate after successful registration. However, the subjects were neither given any instructions regarding the strategy they should use to select the target images forming the respective passwords, nor the strategy they should employ to remember them.

Email prompts were used to inform the subjects about the experimental procedures and tasks, once they gave their consent to receive emails. Emails were sent to the subjects on Day 1, 3 and 5 of each week to notify the progress that they have already made and they are expected to make, to complete the weekly tasks. These emails were meant to help the subjects to keep track of the tasks, and make sure that the experimental protocol was followed. In US1, the frequency of login differed in each week to simulate a scenario, where subjects are using RBGS passwords more frequently in the first four weeks, and less frequently over the last four weeks, as shown in Figure 3.8. The subjects were instructed as given below:

- ***Week 1:*** They must register with 4 passwords and login *20 times* with each password (total 80 logins in the week). This was the training week to get used to the system, since all the subjects were using a RBGS for the very first time.
- ***Week 2:*** They must login *20 times* with each password (total 80 logins in the week).

Meeting after week 2: The subjects were asked to meet the experimenter after completing the tasks assigned to them in week 2, to provide a textual description for each of their four passwords. This phase was conducted in the presence of the experimenter to ensure that the subjects do not use the textual descriptions as an aid to remember their RBGS passwords in the subsequent weeks. The main aim of taking the password descriptions was to conduct a guessability study (GS1), which will be further discussed in Chapter 4. A detailed discussion of the descriptions provided by the subjects in US1 is also presented in Chapter 4.

- ***Week 3-4:*** Subjects must login with each password *ten times* (total 40 logins in the week).
- ***Week 5:*** They must login with each password *twice* (total 8 logins in the week).
- ***Week 6:*** They must login with each password *four times* (total 16 logins in the week).
- ***Week 7:*** They must login with each password *twice* (total 8 logins in the week).
- ***Week 8:*** They must login with each password *thrice* (total 12 logins in the week).

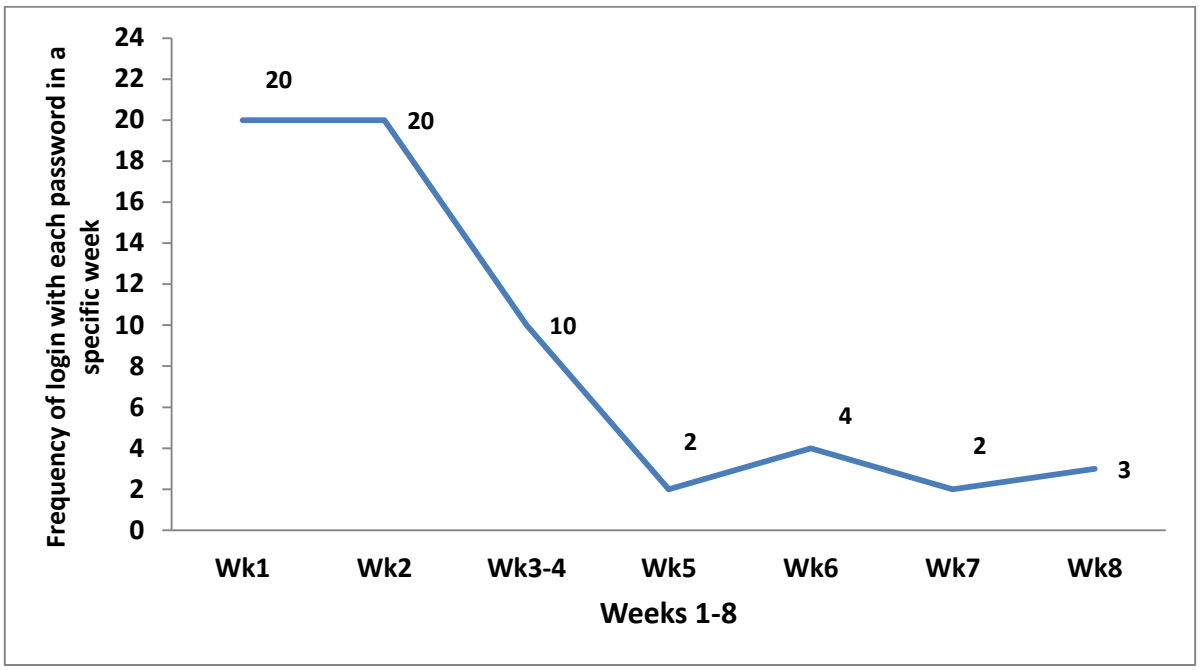


Figure 3.8 Login frequencies with each password in a week.

The subjects were instructed to distribute their authentication sessions over a period of time, instead of completing them simultaneously on the same day. Each subject's log details consisted of the day, date and time for each authentication session. Hence, the log details of the subjects helped us to review the progress made by them and eliminate any subject, who did not follow the instructions that were given to them. These subjects are reported as drop-outs in Table 3.2.

All the existing multiple RBGS password studies except Everitt et al. (2009) have used an experimental protocol, where memorability tests are conducted after a considerable gap (1-2 weeks after the registration stage), without any practice sessions. According to Baddeley (1997), the acquisition of a new piece of information (in the current scenario - remembering

four RBGS passwords) requires practice. In this context, Baddeley (1997) suggested two general principles of practice:

- The total time hypothesis states that a new piece of information could be learned effectively, if more time is spent practicing it;
- The distribution of practice principle states that it is better to spread the practice over a considerable amount of time, instead of doing it en masse.

According to the aforementioned principles, regular password users should experience less memorability problems than the subjects in the multiple password experiments reported in existing literature (for e.g. Chiasson et al., 2009; Moncur & Leplatre, 2009). Hence, the experiment protocol (especially the frequency of password usage) in US1 was designed considering the two learning principles (total time and distribution of practice). The frequency of usage, combined with the distribution of practise in US1 is assumed to simulate the password usage of a regular user in a real world scenario. In this context, the protocol used in US1 ensured that:

- Subjects used their passwords regularly, i.e. frequently (10-20 times) in the first four weeks, and less frequently (i.e. 2-4 times) in the last four weeks, as shown in Figure 3.8.
- Subjects were required to spread the authentication sessions for each of their four passwords throughout the week, instead of doing it in a single session (one day).

(C) Post-study questionnaire

A post-study online questionnaire was circulated among all the subjects, after they had completed the online (8 weeks) study. The questionnaire asked the subjects to rate certain aspects related to the usability of the system (refer to Appendix B.2).

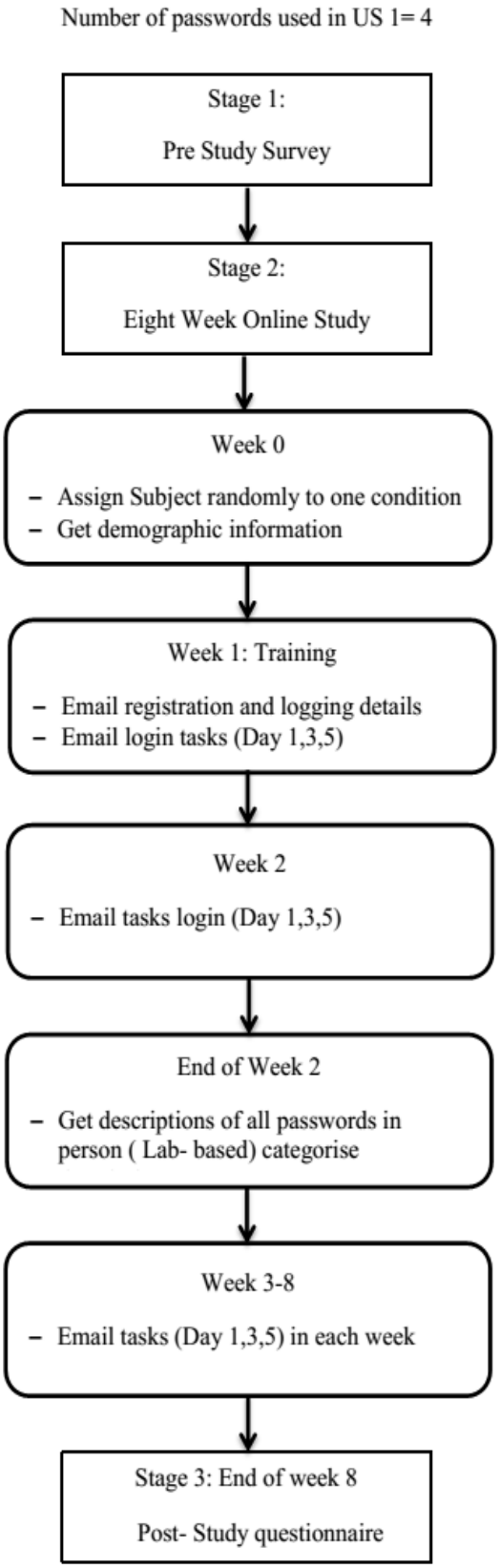


Figure 3.9: Summarising the experimental framework

3.5 Usability Study (US1) Results

The independent variables for the usability study were the four different image types (i.e. conditions): Mikon, doodle, art and object. The dependent variables and the corresponding results are discussed below.

3.5.1 Effectiveness (memorability)

The *mean successful login percentage (SP1)* for each of the conditions is calculated using (Eq. 3.1) as given below.

$$SP1 = \frac{\textbf{\textit{Total number of successful login in the condition}}}{\textbf{\textit{Total number of login in the condition}}} \times \mathbf{100} \text{ (Eq. 3.1)}$$

The mean login success percentages obtained from week 2 to week 8 were considered to analyse the results. The data obtained from week 1 is eliminated, as it was the training week, where subjects got familiar using the system. We did not examine the individual categories of each image type, presented in the Table 3.1 because the aim of this study was to evaluate the memorability of multiple RBGS passwords, composed of the same image type, regardless of the category. For a category-specific study, the experimental protocol would need to be augmented to evaluate the effectiveness of each category in the context of multiple password use, i.e. each subject would need to create four passwords for each category of a specific image type. The descriptive statistics for SP1 is presented in Table 3.4 and the box plots for all the conditions are shown in Figure 3.10

Condition	**SP1 (%)**	**Standard Deviation (SD)**	**Standard Error (SE)**	**Median**	**Range**
Mikon	74.17	4.00	0.80	74.39	65.87-79.87
Doodle	67.04	4.22	0.84	67.68	57.92-76.82
Art	54.90	5.27	1.05	54.26	45.73-65.85
Object	77.31	3.81	0.76	78.04	68.90-84.75

Table 3.4: Descriptive statistics for mean successful login percentage in US1

SP1 for each of the conditions was normally distributed, as assessed by the Shapiro-Wilk test. Given the use of the independent measure experimental protocol with four conditions and the

normal distribution of the data, *One-way independent measure ANOVA* is used to examine the statistical significance.

The results of the *ANOVA* showed significant differences between all the conditions [F (3, 96) = 129.659, p<0.01]. These results indicate that the type of images used as the RBGS passwords would significantly affected the memorability. The results of the *Tukey* post hoc tests revealed significant differences between all the conditions ($p < 0.001$ for all tests), except between Mikon-object ($p = 0.059$). According to the descriptive statistics presented in Table 3.4 and the significance test results, the order of decreasing memorability is: *Object* ≥ *Mikon* > *Doodle* > *Art*

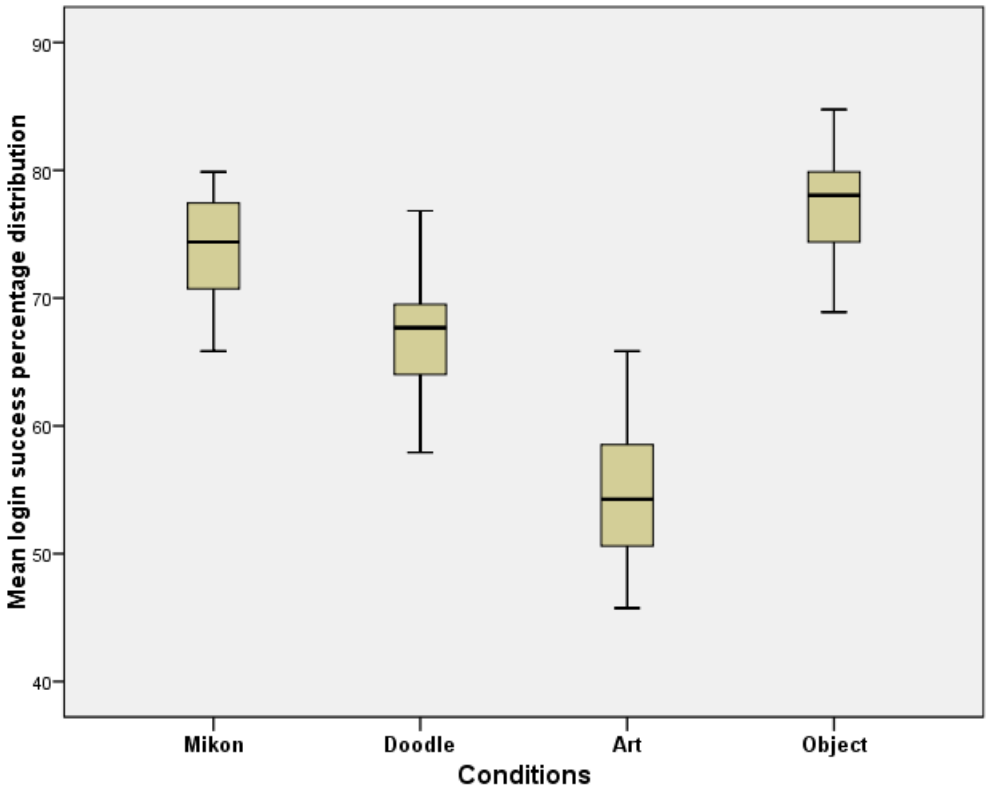


Figure 3.10: Box plot showing the distribution of login success in US1

3.5.2 Mean Weekly Login Success Percentages

The login success percentages in each week (2 to 8) for each of the image types is also analysed, and the results are presented in Figure 3.11. The results show that the mean login success percentages for each of the image type falls from week 2 to week 8, as the frequency of the password usage decreases. A two-way ANOVA is conducted with week and image

types as the two independent variables. The dependant variable is average weekly login success percentages. The week × image interaction is found to be significant, ($F(15,576) = 6.102$, $p < 0.001$). Hence, mean weekly login success percentages significantly varied for each of the image types.

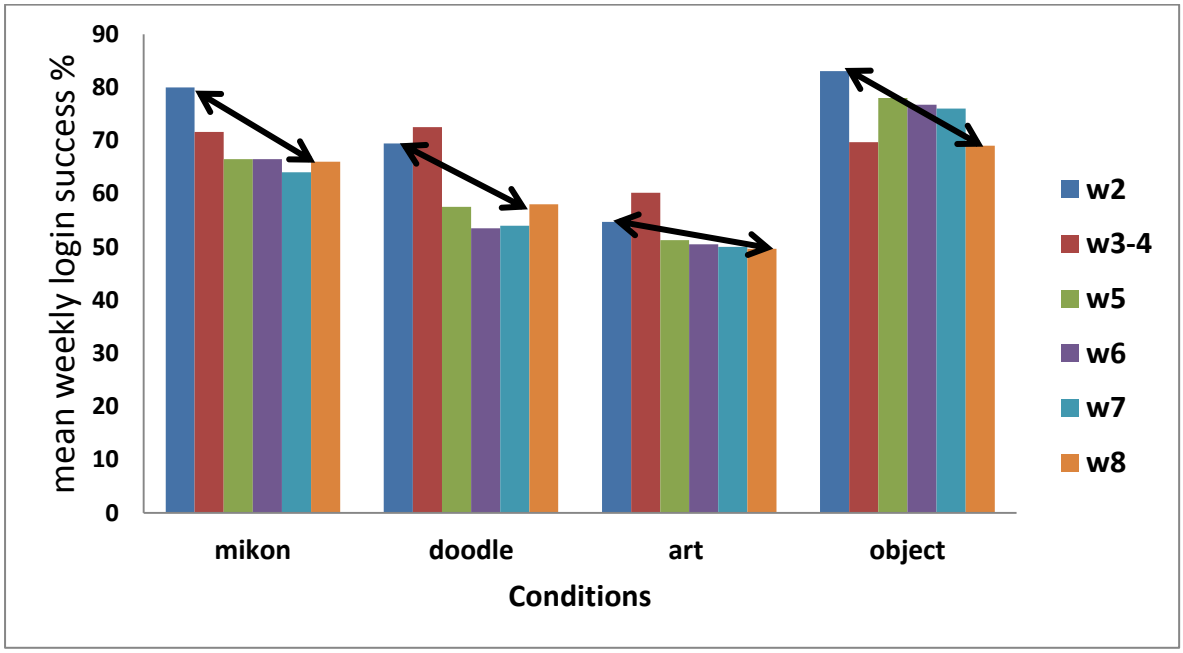


Figure 3.11: Mean weekly login success percentages for each condition

A comparison of the mean weekly login success percentages between week 2 (w2) and week 8 (w8) for each of the image types (bi-directional black arrows in Figure 3.11) shows that the percentages dropped by: 11.44 % in case of Mikon; 12.55 % in case of doodle; 7.74 % in case of art; 14% in case of object. Figure 3.11 also shows that the mean weekly login success percentage for Mikon, doodle and art passwords remained almost the same after week 4. However, in case of object passwords the success percentages are almost the same from week 5-7, but dropped in week 8. Thus the decrease in memorability for the best performers, i.e. Mikon and object are almost the same. Similar characteristic is shown by the doodle passwords. In the case of art passwords (lowest mean login success percentages), the difference is comparatively low, which clearly suggests that subjects had problems remembering these passwords throughout the study. These results show that irrespective of the same frequency of the password usage and the login sessions spread out over time, the memorability of multiple image passwords varies for each image type. This is likely due to the superior encoding of some image types in the human memory. The superior encoding of the image types can be attributed to the familiarity and meaningfulness of the image to the

user. This is also evident from Figure 3.11, where all the types have different login success percentages in each week, despite being used the same number of times by all the subjects.

3.5.3 Efficiency

This measure examined the mean registration time (RegT1), and mean time for successful authentication attempts only (AuT1).

(A) *Registration time:* The registration time for each of the passwords is the time taken to move from screen 1 to screen 5 during the registration process, as shown in Figure 3.5. The mean registration time (RegT1) for each condition is calculated using (Eq. 3.2) as given below.

$$\boldsymbol{RegT1} = \frac{\mathbf{1}}{\mathbf{4}}\sum_{\boldsymbol{i=1}}^{\mathbf{4}} \boldsymbol{Registration\ time\ for\ password\ (i),\ where\ i\ denotes\ password\ 1,2,3,4} \qquad \text{(Eq. 3.2)}$$

The descriptive statistics for the mean registration time in each condition is shown in Table 3.5 and the box plots are presented in Figure 3.12.

Condition	RegT1(sec)	SD	SE	Median	Range
Mikon	72.18	5.48	1.17	72.5	62-82.25
Doodle	75.42	4.27	0.88	75.5	65.75-84.25
Art	84.44	4.91	0.99	83.75	74.75-94.25
Object	70.61	3.84	0.76	71.50	62.5-76

Table 3.5: Descriptive statistics for mean registration time in US1

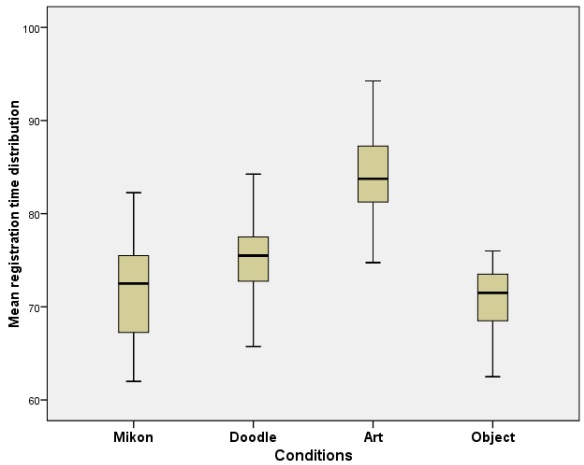


Figure 3.12: Box plot distribution for registration time in US1

The results presented in Figure 3.13 show that the registration time decreases as the subjects get used to the system, in all the conditions (registration time decreases from p1- first registered password to p4- last registered password).

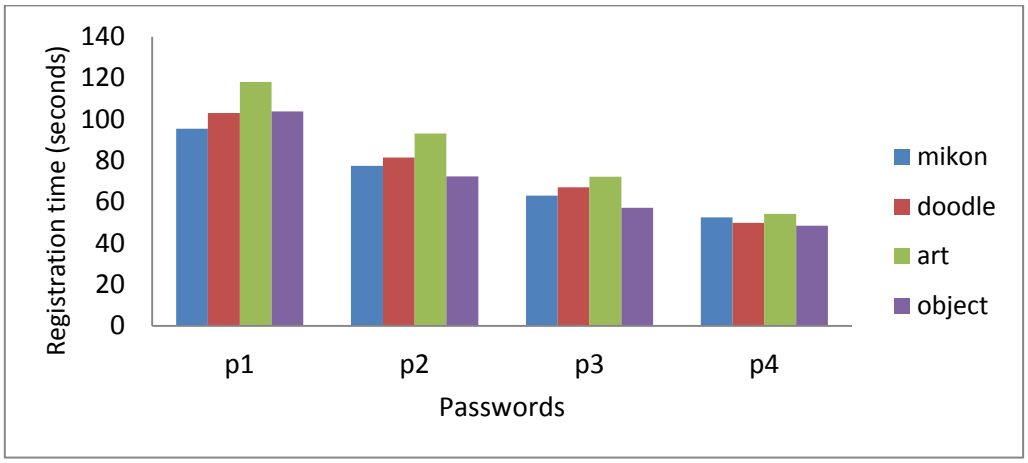


Figure 3.13: Mean registration time for each password in each condition

The mean registration time (of 4 passwords) in all the conditions was normally distributed as assessed by the *Shapiro-Wilk test*. A *One-way independent measure ANOVA* is used as the statistical test. The result of the *ANOVA* confirmed significant differences between all the conditions ($F(3, 96) = 41.277$, $p<0.001$). This shows that the type of images used as a password by the subjects affect the mean registration time significantly. The result of the T*ukey post hoc* tests revealed significant differences between all the groups ($p < 0.05$), except Mikon-doodle ($p = 0.091$) and Mikon-object ($p = 0.658$). According to the descriptive statistics presented in Table 3.5 and the significance test results, the order of increasing registration time is: *Object* ≤ *Mikon* ≤ *Doodle* < *Art*

(B) *Authentication time for successful attempts:* The authentication time for a password is the time taken to proceed from screen 2 to the success notification screen of the authentication process, as shown in Figure 3.6. The mean time for successful authentication attempts (AuT1) in each condition is calculated using (Eq. 3.3), where z represents the total number of successful authentication attempts.

$$\boldsymbol{AuT1} = \frac{\mathbf{1}}{\boldsymbol{z}} \sum_{\boldsymbol{n=1}}^{\boldsymbol{z}} \boldsymbol{Login\ time\ for\ successful\ login\ (n)} \qquad \text{(Eq. 3.3)}$$

The descriptive statistics for AuT1 in each condition is presented in Table 3.6 and the box plots are shown in Figure 3.14.

Condition	AuT1 (sec)	SD	SE	Median	Range
Mikon	19.52	3.60	0.72	19	13-26
Doodle	22.16	3.75	0.75	22	16-31
Art	24.56	4.8	0.96	24	17-35
Object	18.28	2.84	0.59	18	13-24

Table 3.6: Descriptive statistics for mean authentication time in US1

The mean authentication time (of 7 weeks) for each of the conditions was normally distributed as assessed by the Shapiro-Wilk test. The results of the one way ANOVA showed significant differences between all the conditions ($F(3, 96) = 13.199$, $p < 0.001$). The results of the *post hoc* tests revealed significant differences between all pairs of condition ($p < 0.05$), except Mikon-doodle ($p=0.091 > 0.05$) and Mikon-object ($p = 0.658 > 0.05$). According to the descriptive statistics presented in Table 3.6 and the significance test results, the order of increasing for authentication time is: *Object* ≤ *Mikon* ≤ *Doodle* < *Art*

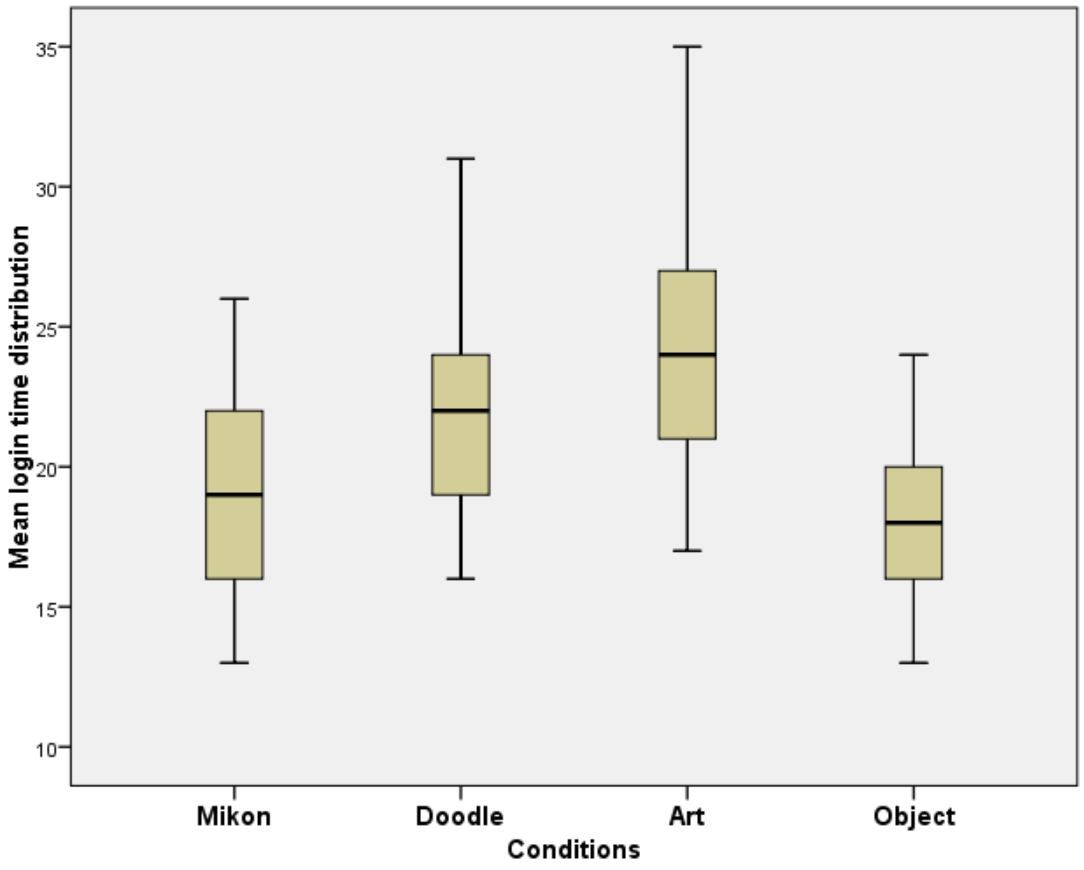


Figure 3.14: Box plot distribution for authentication time in US1

3.5.4 Post-Study Questionnaire Results

(A) Satisfaction using RBGS passwords

This dimension was assessed from the ratings (1- 5, 1 being highly dissatisfied to 5 being highly satisfied) given by all the subjects to the different aspects (*sat-1 to sat-4*), as follows:

- (*sat1*) Ease to register in RBGS;
- (*sat2*) Ease to authenticate in RBGS;
- (*sat3*) Meaningfulness of the target images forming the password;
- (*sat4*) Overall satisfaction wi th the type of image used as password.

The aforementioned aspects were based on some of the items in SUS (System Usability Scale) questionnaire (SUS, 2011).

The mean value of *sat (final)* for each of the conditions was calculated using Eq. 3.4.

$$sat\ (final) = \sum_{i=1}^{25} \frac{sat1(i)+sat2(i)+sat3(i)+sat4(i)}{4} \qquad \text{(Eq. 3.4)}$$

i represents the subjects (1-25) in each condition

The descriptive statistics of the measure *sat (final)* is presented in Table 3.7 and the box plots are shown in Figure 3.15. The mean statistics for each of the sat aspects (1 to 4) in each condition is presented in Table 3.8

Condition	sat (final)	SD	SE	Median	Range
Mikon	13.25	1.64	0.33	13	12-14.5
Doodle	12	1.61	0.32	12	10.5-12
Art	9.20	1.58	0.32	9	8-10.5
Object	13.91	1.88	0.38	14	13-15

Table 3.7: Descriptive statistics for mean satisfaction ratings in US1

Condition	sat1	sat2	sat3	sat4
Mikon	3.12	3.36	3.5	3.27
Doodle	2.95	3.05	2.95	3.05
Art	2.45	2.45	2.14	2.16
Object	3.24	3.56	3.65	3.46

Table 3.8: Descriptive statistics for each sat aspect in US1

Given the ordinal scale of the data (i.e. user ratings) and the independent measure experimental protocol with four conditions, a *Kruskal-Wallis* test is used to examine the statistical significance. The test result showed that the *sat(final)* scores for each of the conditions were significant [H (3) = 52.37, p < 0.001). In other words, the mean sat(final) ratings of the subjects are significantly affected by the type of images used as password. A Mann-Whitney test was conducted to follow up the findings by applying a Bonferroni correction, to report all the effects at a 0.008 level of significance. The Bonferroni correction (Field & Hole, 2003) is used to reduce the chances of obtaining false-positive results (type I errors), when multiple pair-wise tests are performed on a single set of data (especially, when a non-parametric test is used as a post-hoc test, to find statistical difference between multiple pairs of conditions). We performed a Bonferroni correction, by dividing the critical *P* value (α =0.05) by the number of comparisons being made (i.e. 6).

The results revealed that the mean satisfaction scores were significantly different in all conditions (p < 0.008 for all tests), except for the Mikon-object (p = 0.156). Hence, we conclude that the subjects were most satisfied with object and Mikon passwords (no significant difference), followed by doodles, and the least satisfied with art passwords.

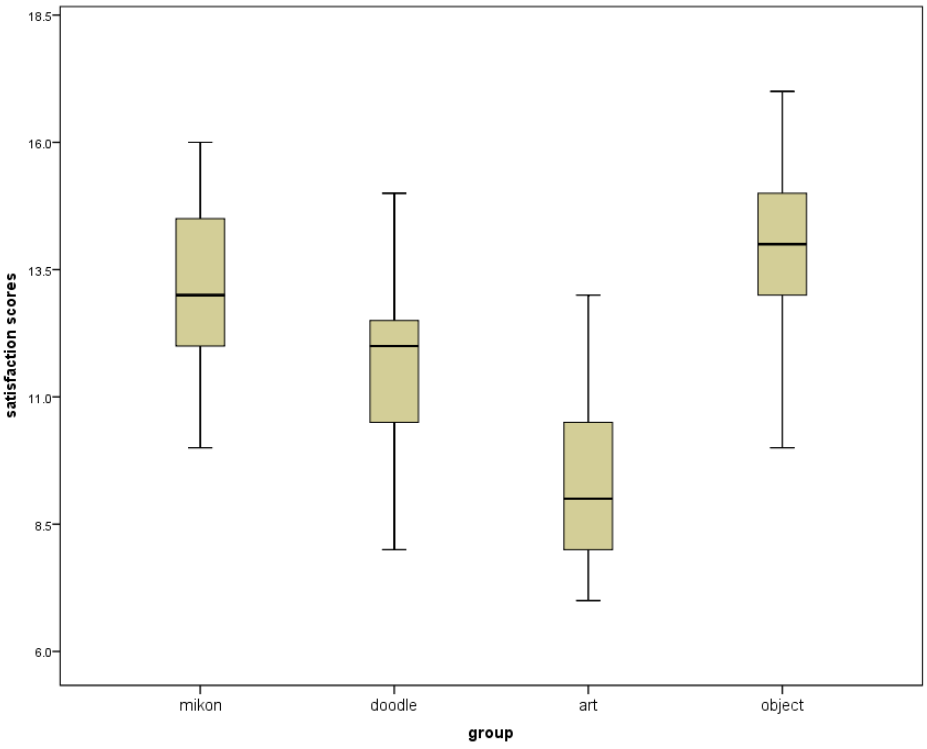


Figure 3.15: Box plot distribution for satisfaction ratings in US1

(B) Strategy used for password creation during the registration stage

The subjects were asked to provide information on the strategy/approach they had used (if any) to choose the target images forming the respective passwords, and remember them for subsequent use. They were asked to select one or more options from the list given below.

- *Story/pattern:* target images forming a password follow a pattern or a story.
- *Personal likings:* target images are personally related to you (something you like).
- *Visual appeal:* target images stand out compared to the other images, or you prefer them visually.
- *Caption/tags:* target images have a tag or annotation, which aid memorability.
- *Random:* any random strategy that came to your mind, or no strategy;

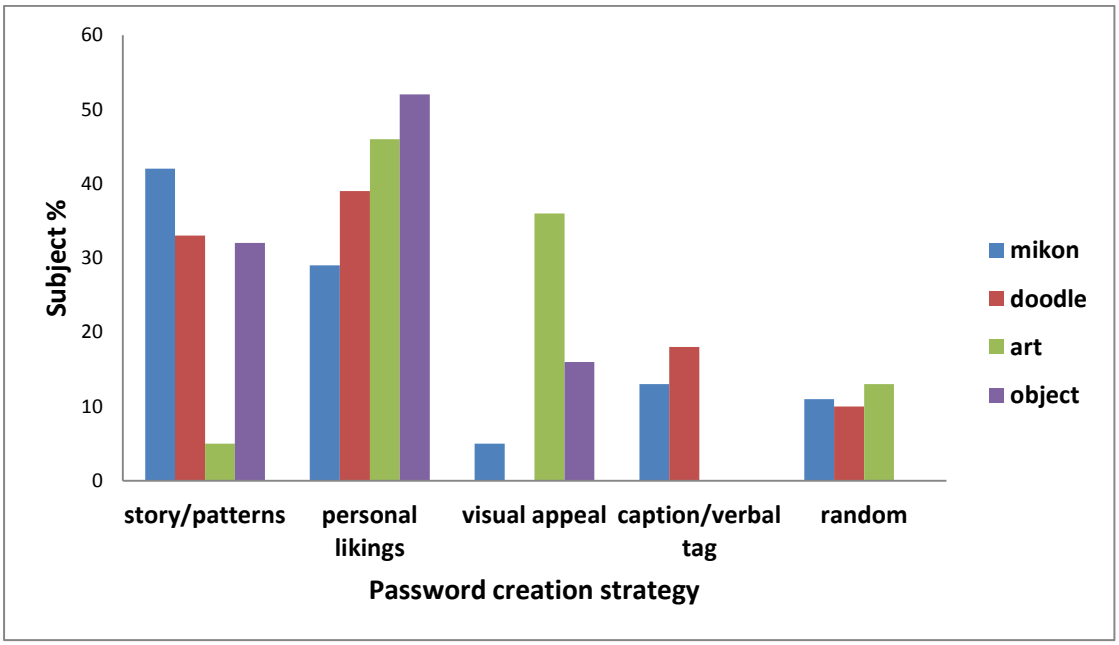


Figure 3.16: Password creation strategy reported by subjects in US1

The subjects were given examples for each item in the aforementioned selection list, so that they could choose their options appropriately. The results obtained from the subjects are shown in Figure 3.16. The results reveal that most of the Mikon and doodle subjects either employed a story/pattern strategy to remember their passwords, or they chose passwords according to their personal likings. Most of the subjects assigned to the art condition chose passwords, either based on their personal likings (favourite colour, objects, scene), or visual and aesthetic quality of the images (i.e. attractiveness). However, these strategies employed to aid memorability of multiple RBGS passwords may either make them guessable to an

intruder who knows the user quite well, or could be disclosed and shared easily, which needs to be examined in future studies.

(C) Recording passwords

The subjects were asked, if they have made an attempt to record their RBGS passwords, and how would they record such passwords (coping strategies) to aid memorability in the future. None of the subjects reported to have made an attempt to record their passwords. However, almost 84 % of the subjects who used art passwords, 68 % of the doodle, 36% of the Mikon and 32% of the object subjects reported that they might use screen captures, sketches or notes (written descriptions), to aid memorability of multiple RBGS passwords. It can be also assumed that the recording strategies would also depend upon the device being used for authentication (handheld devices or desktop machines), and the importance of the account from the user's perspective.

3.6 Comparison with Other Studies

In this section, the results obtained from US1 are compared to the existing studies that have explored the cognitive demands of using multiple RBGS passwords. The results reported in the existing multiple RBGS password studies has been summarised in Table 2.5 (Section 2.6), which is frequently referred to in this section.

US1 – Moncur & Leplatre (2007)

The results reported in US1 demonstrates superior performance, viz. 77.31 % mean login success percentage in case of object passwords, compared to the mean login success of any of the groups in Moncur & Leplatre (2007), as shown in Table 2.5. The registration and authentication times were not reported in Moncur & Leplatre (2007).

US1 – Chiasson et al. (2011)

In the memorability test reported in Chiasson et al. (2011), which was conducted two weeks after the training stage, the mean success rate of the click-based passwords was 57% and text-based passwords was 70%, which is a bit lower compared to the mean success rates reported in US1 (varied between 54.9% and 77.31%). However, the frequency of password use in Chiasson et al. (2011) was less than US1, and the number of passwords used in the former

study was also more, which might have contributed to the difference in the memorability performances. The registration time for click-based passwords was 43.9sec, which is lower than that of RBGS passwords in US1, ranging between 70.61sec (object) and 84.44sec (art). The mean authentication time for the click-based passwords varied between 15.1sec and 47.0sec after two weeks, which is more than that of RBGS passwords in US1, ranging between 18.28sec (object) and 24.56sec (art).

US1 – Everitt et al. (2009)

Everitt et al. (2009) demonstrated that subjects accessing four different facial passwords each week had a failure rate of 15.23% after a month, when each password was used once a week (Table 2.5). In US1, the lowest failure rate is in the case of object passwords (23.69%) and highest in case of art passwords (45.1%), which shows that the RBGS passwords in US1 had an inferior performance compared to Everitt et al. (2009). The mean login time reported in (Everitt et al., 2009) was 29.7 sec, which is more than the results reported in US1 (18.28 sec - 24.56 sec). The passwords in Everitt et al. (2009) were issued by the system, so the registration time was not reported.

US1 – Hlywa et al. (2011)

The results reported in Hlywa et al. (2011) demonstrated slightly better performance, viz. mean success rate of 78.33% in study 1 and much superior performance, i.e. mean success rate of 95% in study 2 for object passwords, compared to the login success results reported in US1 (Table 3.4). However, the number of object passwords used in both the studies (S1: 3 and S2: 2) reported in Hlywa et al. (2011*)* is less compared to that of US1 (4 passwords). In context to the mean authentication time, the performance of the object passwords in US1 (Table 3.6) is superior compared to the statistics reported in Hlywa et al. (2011) for both the studies, as shown in Table 2.5. The login performance with face passwords in Hlywa et al. (2011) was better in study 2 (Table 2.5) compared to the results reported in US1 (Table 3.4). The mean authentication time of face passwords in both the studies reported in Hlywa et al. (2011) were much higher compared to all the conditions in US1 (Table 3.4).

3.7 Discussion

3.7.1 Effectiveness of Multiple RBGS Passwords

The results presented in Table 3.4 shows that the memorability of RBGSs is significantly affected by the type of images used as passwords. In this context, the results show that the mean login success percentage is highest for objects, closely followed by Mikons, then doodles and lowest for art images.

According to the cognitive studies, dual coding theory (Paivio, 1986) and guided search process (Wolfe, 1994), an elaborative encoding of an image in the human memory makes it memorable. Thus, an image which is easily associated with a name (nameable), or can be interpreted in a meaningful way is more likely to be memorable due to its superior encoding in the human memory. In this context, Table 3.8 (column-4) shows that the mean score of the *sat3 parameter (meaningfulness of the image)* is: highest for objects (3.65/5); closely followed by Mikon (3.5/5); then doodles (2.95/5); lowest for the art images (2.14/5). Thus the results of sat3 parameter are in line with the mean successful login percentages reported in Table 3.4. Hence, the higher memorability of the object and Mikon images can be attributed to the fact that the subjects found these images to be meaningful. The doodle images are black and white line drawings and do not convey much meaning to aid memorability. Hence these images are unlikely to be encoded in an elaborate way in the human memory, unless they are drawn by the individual account holders. According to the subjects, the art images are very difficult to remember because it is difficult to associate them with something meaningful. Moreover, these images are visually complex, i.e. they contain a lot of information and colour, which may eventually lead to information overload in memory. This complements the work reported in Szekely & Bates (1999), which had suggested that the visual complexity of an image is linked to the ease of associating it with a name. The authors also suggested that it is difficult to assign names to visually complex images.

3.7.2 Efficiency of Multiple RBGS Passwords

The results also reveal that the mean registration time is: lowest for the objects; closely followed by Mikons; then doodles; highest for the art images (Table 3.5). In this context, the mean score of the sat1 parameter (i.e. ease to register) is found to be (Table 3.8, column-2):

highest for objects (3.24/5); closely followed by Mikons (3.12/5); then doodles (2.95/5); lowest for the art images (2.45/5). Thus the results obtained from the subjects (Table 3.8, column-2) complements the mean registration time statistics presented in Table 3.5. These results can be attributed to the fact that users find it difficult to choose meaningful images in case of doodle and art, which they could use as passwords. The authentication time follows the same trend as that of the registration time (Table 3.6): lowest for objects; closely followed by Mikons; then doodles; highest for art images. The mean score of the sat2 parameter (ease to authenticate) is found to be (Table 3.8, column-3): highest for the objects (3.56/5); closely followed by Mikons (3.36/5); then doodles (2.95/5); lowest for the art images (2.45/5). Thus the mean scores of the sat2 parameter (Table 3.8, column-3) complement the results of the mean login time statistics presented in Table 3.6.

The above discussion demonstrates that the effectiveness and efficiency results complement each other, i.e. objects are the best performers both in terms of effectiveness and efficiency, followed by Mikon, doodle and art passwords. Hence, it is concluded that images which are meaningful or can be associated with something easily are: effective in the sense of being memorable; efficient, i.e. less time consuming to employ. This conclusion is also supported by the mean satisfaction score obtained from the post-study questionnaire: highest for the objects (13.91/20); closely followed by Mikons (13.25/20); then doodles (12/20); lowest for the art images (9.20/20).

3.7.3 Limitations

In US1, the subjects were told that their passwords were for four specific accounts, which may have made them to select account specific passwords. US1 was conducted with students and we acknowledge that the general population is much more diverse. However, most of the studies both in the field of GASs and RBGSs have also been conducted with students. But, we believe that the user group in US1 is adequate to establish a baseline performance in the context of the memorability of multiple RBGS passwords.

We believe that despite the aforementioned limitations, examining the issue of multiple passwords in an online experimental setting can be considered an essential step in understanding the effect of increased memory load, and the password memorability problem. In the future, the same usability study could be conducted in a more ecologically valid setting, i.e. diverse user group, to get a better idea of the memorability issues. It is also

advisable to make the subjects familiar with the authentication systems beforehand, so that their behaviour is more natural and the novelty effects might be avoided.

3.8 Conclusion

The user study presented in this chapter compared the usability of multiple RBGS passwords using four different image types- Mikon, doodle, art and objects. The results of the study show that object and Mikon passwords performed the best in each of the usability criteria compared to doodle and art passwords.

The object passwords in US1 had the highest mean login success, i.e. 77.31 %, compared to the other three image types. The memorability statistics (login success rates) for each of the image types reported in US1 are lower than the success rates reported in the single password studies with RBGSs (Table 2.3, Chapter 2). Hence, subjects in US1 did find it difficult to remember multiple RBGS passwords. The results also reveal that RBGS passwords in US1 are time consuming to employ; the mean registration time is either equivalent to or higher than most RBGSs, except Charrau et al. (2005) (Table 2.3), while the authentication time is similar to the statistics reported in other studies.

The post study questionnaire results demonstrate that most subjects in US1 chose the target images forming their respective passwords, either by making a pattern/story or something which is related to them. These results underscore the need to examine, whether passwords created using patterns or mnemonic strategy aid memorability, when multiple RBGS passwords are used. The ease of employing such a strategy with different image types also needs to be assessed. This aspect is further discussed and examined in Chapter 5.

The post study questionnaire results also suggest that though meaningful images would aid memorability when multiple RBGS passwords are used, users may still engage into insecure coping mechanisms, like recording them through digital or non-digital media. In this context, the next chapter presents a user study, to examine the guessability of RBGS passwords using the corresponding password descriptions provided by the respective account holders.

Chapter 4
Vulnerability of RBGS Passwords to Textual Descriptions

This chapter presents a guessability study (GS1) with 70 participants using four different image types: Mikon; doodle; art; object, to examine the vulnerability of RBGS passwords to textual descriptions The study will examine, whether textual descriptions of the target images forming a password, provided by the respective account holders, could be effectively used by an attacker to authenticate. This chapter addresses the Objective 2 corresponding to the Stage 3 (Section 1.4). The contents presented in this chapter have been published in the proceedings of the 7th International Conference on Security of Information and Networks (SIN 2014).

4.1 Introduction

Most users find it difficult to remember text passwords, which are the most widely used authentication mechanism. As the number of passwords a user has to remember increases, the system proves onerous in terms of memorability (Adams & Sasse, 1999). This makes the user employ unsafe strategies like writing down the passwords and sharing them with others, which compromises the security of the system (Adams & Sasse, 1999; Herley et al., 2009). The various issues related to text passwords have been already discussed in Chapter 1 (Section 1.3.1).

Recognition-based graphical authentication systems (RBGSs) have received significant attention as a potential alternative to text passwords. The literature presented in Chapter 2 highlights that users find it difficult to remember multiple RBGS passwords. Hence, it appears to be a reasonable assumption that users would be likely to attempt to adopt the same coping strategies as with text password systems, if RBGS passwords were to become widely used. The use of such strategies would compromise the security of RBGSs. There is no substantial evidence to hold the assumption made in the past that RBGS passwords will be particularly resistant, to being written down or verbally communicated (Dhamija & Perrig, 2000; Real, 2004). In the context of the RBGSs, none of the existing studies have explored the nature of user's descriptions for the credentials forming the passwords, and the vulnerability of such passwords to descriptions. This chapter presents a user study (GS1) to

examine the guessability of RBGS passwords using their corresponding textual descriptions, provided by the respective account holders.

4.1.1 Graphical Password Description

According to Dunphy et al. (2008), a RBGS password description can be defined as "*any non-digital attempt to record or communicate a password, using either an external representation, or verbal/nonverbal means"*. This would include:

- sketches of the target images;
- word description of target images;
- verbal descriptions of the target images;
- instructions and physical gestures.

RBGS passwords may be recorded/shared in three different ways (Dunphy et al., 2008):

- producing written descriptions of the target images forming the password;
- verbally communicating the descriptions of the target images;
- recording the target images using screen captures, smart phones and other hand held devices.

Since RBGSs have not been widely adopted, it is unclear how users will record/share their passwords, in practice. Moreover, none of the existing studies in the field have examined this aspect. In the absence of any information on the coping mechanism that might be employed, this chapter examines the guessability of RBGS passwords using their textual descriptions.

4.1.2 Threat Model

The threat model discussed in Chapter 2 (Section 2.7.3) applies to the guessability study reported in this chapter. In this sub-section, the threat model is further discussed to make it more explicit, in the context of the guessability study. There are modern technologies like built-in camera in smart phones and other hand-held devices, which would enable the users to record and keep a copy of their RBGS passwords. But, recording RBGS passwords using such modern technologies would also depend on the technical expertise of the user, and its availability in a specific instance. The guessability study reported in this chapter assumes that recording the target images forming a RBGS password using such modern technologies

would be unattractive due to their permanence, and can be identified easily, if discovered by an adversary. Hence, a more transient and spontaneous way to record RBGS passwords will be based upon descriptions. A user may wish to record descriptions of the target images forming the password for a specific account, to aid memorability of the same credentials in subsequent use. An attacker may get hold or seek out the descriptions, if these are recorded externally or not kept safely (by the account holder, i.e. respective user), and try to guess the RBGS credentials to gain unauthorised access. It is also assumed that the attacker is aware of the username, corresponding to the specific password description. It is also assumed that the description of the target images forming a password is written sequentially, i.e. in the same order they are displayed during the authentication session.

4.1.3 Terminologies

For clarification, the definitions of aspects related to GS1 reported in this chapter which will be used frequently are as follows.

- *Account holder:* A legitimate user who has registered with an RBGS password and proves the authority using the same password;
- *Description/ textual description:* A description of the target images forming the RBGS password provided by an account holder. The description must take the form of words, written by the respective account holders only;
- *Subject:* An account holder who took part in the usability study (US1) reported in Chapter 3;
- *Attacker:* A participant in the guessability study (GS1) reported in the current chapter, who is trying to guess the RBGS passwords, using the corresponding word descriptions provided by the respective subjects. The attackers who took part in GS1 are different to the subjects in US1.

4.1.4 Contributions

The main contributions of this chapter are:

- A user study (GS1) is reported with 70 attackers to examine the guessability of four different image types when used as RBGS passwords, using their textual descriptions.

- In the absence of any known experimental framework, the research also contributes a methodology that enables realistic and practical studies to examine the vulnerability of RBGS passwords to textual descriptions.

4.2 User Study

GS1 was conducted to examine the extent to which RBGS passwords can be guessed using the respective descriptions of the target images. The descriptions were provided by the subjects who took part in US1, which has been reported in Chapter 3.

4.2.1 Recruitment of the Attackers

In order to recruit the participants (attackers) for this guessability study, emails were sent to the third and fourth year student email distribution lists in the university. The mail comprised of the following components:

- The main objective of the study;
- A brief summary of the tasks involved ;
- Type of study, i.e. Lab-based/Online/Hybrid;
- Approximate duration of each task;
- Mode of communication that will be used during the period of the study;

A total of 79 participants (attackers) volunteered to take part in the study, by responding to our email. Out of these 79, 5 attackers were randomly approached to take part in a pilot study. The aim of the pilot study was to ensure that:

- Attackers could use the instructions provided to them easily, to complete the guessability attacks by logging into the system;
- They are able to understand their tasks from the documentation provided to them and the emails that were framed for the purpose of the study (GS1);
- The experiment design need not be changed, during the actual guessability study (GS1) with the attackers;
- The dependent variables used for the purpose of this study can actually measure the phenomenon of guessability of RBGS passwords using written descriptions.

Once the pilot was completed, an additional email was sent to the remaining 74 attackers, to confirm their participation. 70 attackers responded to the mail and agreed to participate by signing the participation consent form, once they were given all the relevant documentation (login process, detailed description of the guessability tasks). The remaining 4 attackers, either did not respond to our email or decided not to take part. The guessability study was completed by 70 attackers. Table 4.1 summarises the responses received during each stage of the recruitment process.

Stage	**Subjects responded**	**Comments**
Email invitation	79	
Pilot study	5	74 left for actual GS1
Confirmation to participate	70	no response from 4
Study completed	70	No one dropped out

Table 4.1: Recruitment information for the user study GS1

4.2.2 Attacker Demographics

The guessability study was conducted with 70 attackers (42 males and 28 females, age range: 18 to 24 years). The attackers were studying different undergraduate degree programmes as follows:

- Mechanical Engineering – 15;
- Electrical Engineering- 19;
- Aerospace Engineering- 20;
- Computer Science- 16.

The attackers were neither expert in usable security nor studied the topic as a part of their curriculum. Ethics number (CSE01061) was assigned by the college ethics committee, once the ethics approval was granted. There were no drop-outs in GS1. This might be attributed to the fact that the attackers were able to complete their tasks online as per their convenience, without any time limits imposed on them.

4.2.3 Experiment Protocol and Framework

The guessability study was designed using a repeated measure protocol with four conditions as follows:

- guess Mikon passwords using the corresponding descriptions of the target images;
- guess doodle passwords using the corresponding descriptions of the target images;
- guess art passwords using the corresponding descriptions of the target images;
- guess object passwords using the corresponding descriptions of the target images.

Each attacker had to guess four passwords in each condition, using the corresponding descriptions and usernames that were given to them. Each condition was performed on a different day.

In the guessability study, seventy attackers were divided into five groups (G1-G5), i.e. fourteen attackers in each group. Each attacker in a group had to guess four passwords for each condition, using the corresponding descriptions, which were also the same for the other attackers in that group. Each group had descriptions of sixteen distinct passwords (four descriptions for each condition), different from the other groups. Hence, the guessability study was conducted with 20 distinct passwords (5 groups × 4 conditions) in each condition. The passwords that were allocated to each group have been illustrated in Table 4.2.

Group	**Attackers**	**Password number**				
		Mikon	**Doodle**	**Art**	**Object**	Study with twenty distinct passwords for each image type
1	1-14	1-4	1-4	1-4	1-4	
2	15-28	5- 8	5- 8	5- 8	5- 8	
3	29-42	9-12	9-12	9-12	9-12	
4	43-56	13-16	13-16	13-16	13-16	
5	57-70	17-20	17-20	17-20	17-20	

Table 4.2: Passwords allocated each group in GS1

GS1 was conducted online, so the experimenter communicated with the attackers, via email, once the attackers gave their consent to take part in the study, and adhere to the instructions (refer to Appendix C for details). Each attacker was sent four emails, i.e. one for each condition. The emails were sent in a sequential order, i.e. once the attackers finished the tasks

for the first condition, they were sent the details of the second condition and so on. Each email comprised of:

- URL of the login page for the condition currently under attack;
- All the instructions required to complete the guessability attacks involved in the task (authentication steps);
- The usernames to be used for logging in to the system;
- The respective password descriptions corresponding to each username;
- Access to all the required online resources including a thesaurus, web images.

Each attacker was allowed four attempts to guess a password, but there was no time limit imposed for the task. The prototypes used to conduct GS1 were the same as the ones used in US1.

4.2.4 Description Collection and Instructions

The RBGS passwords (for all the conditions in GS1) and the corresponding descriptions were collected from US1 reported in Chapter 3. In US1, each subject created four passwords using one of the image types. The subjects were asked to write down the descriptions of their passwords, once they had used the system for two weeks (i.e. completed 30-40 login sessions), as discussed in Chapter 3 (Section 3.4.3). This method of taking the descriptions was more realistic, as it might be easier to provide the descriptions, once a degree of familiarity has been achieved. Moreover, all the subjects in US1 were using a RBGS for the first time, and providing a recorded prompt of an RBGS password is not a task that they do in their daily life. If a degree of familiarity with the RBGS passwords as well as the system is not achieved, then the subjects would be describing images, rather than their passwords. This would put limitations to the basic aim of the guessability study. The instructions given to the subjects in US1 for producing their descriptions are discussed below.

I1: The textual descriptions must be written in English, but don't need to be grammatically correct (all the subjects were well versed in English, as they had studied all their modules in their respective schools and university in this language);

I2: The textual description for each target image must not be more than 25 words. A word limit was imposed to ensure that the descriptions are not too long and realistic;

I3: The subjects should describe, as if they were keeping a written prompt for themselves, to aid memorability of the passwords in the subsequent use. Since the subjects were recording a prompt for themselves, it was reasonable to show them the passwords, while they were preparing the descriptions. It is assumed that a subject won't record a prompt, without having a look at the target images forming the respective password. This assumption is reasonable because the main purpose of recording the descriptions (prompt) is to make a copy of the correct prompt of the target images, which can be successfully used to authenticate in the future.

It might be argued that without any restrictions on how the prompts are written, an adversary might be able to use the description to guess the image successfully. The study reported here is first of its kind, and as discussed in Chapter 2 we are not aware of any existing experimental protocol and methods that could be used to conduct such a study. Hence, we did not want to put restrictions that would make it difficult for the subjects to record a written prompt of their passwords. In fact, the study reported in Dunphy et al. (2008) had used a protocol, where contributors were asked to describe a face, as if they were telling it to a friend. It is maintained that there is quite a bit of difference in making a reminder for oneself, and one for a friend. However, in this study the subjects were instructed to make a prompt for themselves, which will help them to recall the password in the subsequent use.

It might be contended that guessability attacks are very much dependent on the decoy set of images. The review of RBGSs presented in Chapter 2 (Tables 2.6 and 2.7) suggests that all the existing studies have used different methods to choose decoy sets and, in fact, it is also not known which decoy set selection method is suitable in terms of usability or security. This is a separate research problem, and not the focus of the current chapter. In the absence of a standard decoy set selection method, the decoy images for a target image were randomly chosen (most popular as shown in Tables 2.6 and 2.7) from the same collection, which is used during the registration process. We acknowledge that randomly chosen decoy images may make some image types more guessable than the others. This might be the case with any decoy set selection approach. The research on decoy set selection is sparse, but if a secure and usable method is found, the experiment described in this chapter can be repeated.

4.3 Analysis of the Descriptions

In this section, we discuss the process followed to collect the password descriptions from the respective account holders (subjects who participated in US1).

4.3.1 Stage 1

Most subjects in US1 drew sketches (with/without annotations) of the target images forming the password, instead of providing a textual description (Instruction I2 in Section 4.2.4). According to the statistics presented in Table 4.3, subjects drew sketches of the target images in the case of Mikon, doodle and object passwords more than the art passwords. The table also shows that the subjects may draw their RBGS passwords, in case they need to make a copy of the same to aid memorability. However, this study cannot validate whether subjects will prefer to record their prompts as sketches rather than text, in the current experimental settings used for GS1. A separate study with a different experimental framework and instructions would be suitable for investigating the most preferred way to record prompts for RBGS passwords, which should also take into account the context for recording the prompts.

	Number of RBGS Passwords having sketches for			
Category	*4 (T) images (A)*	*3(T) images (B)*	*2 (T) images (C)*	*1 (T) image (D)*
Mikon	67	21	8	4
Doodle	76	24	0	0
Art	0	0	21	79
Object	78	15	4	3

Table 4.3: Statistics showing RBGS passwords described as sketches, T means target images

All the subjects were using RBGS passwords and recording a written prompt of the respective passwords for the first time in their life. Hence we believe that they found it easier to draw sketches, rather than giving a textual description, which is also a time-consuming task. The results also demonstrate that subjects wrote textual descriptions, only when the

target image is difficult to draw, as in the case of art passwords. Hence the coping strategy to aid memorability may depend upon the visual complexity of the image type used as the RBGS password.

If it is assumed that passwords with three or all the four target images that were described as sketches (i.e. category A and B) are guessable, then according to statistics presented in Table 4.3, 88% of the Mikon, 100% of doodle and 92% of object passwords can be guessed easily by an attacker. This assumption is reasonable, since it is easier to guess an image using the corresponding sketches. Based on this assumption, 12 Mikon, 0 doodle, 100 art and 7 object passwords will be available for GS1, since the study required textual descriptions. These numbers would have been very low to conduct a guessability study with seventy attackers in the case of Mikon, doodle and object passwords. We were aware that such a scenario may arise based on our pilot studies, but did not have an explicit idea of the aforementioned low statistics, while the descriptions were being given by the subjects in US1. In order to address this challenge, we decided to make the subjects record their prompts as textual descriptions (no sketches), which is further discussed in Section 4.3.2.

4.3.2 Stage 2

When the subjects in US1 drew sketches to record their RBGS passwords, they were asked to write down a textual description for each target image forming the respective passwords (Instruction I2 in Section 4.2.4). The protocol (i.e. using textual description) was not changed to ensure, adequate number of passwords for each condition are available for the guessability study. The descriptions of the twenty most memorable passwords in each condition were selected for the guessability study. Most memorable meant that the login success percentage of these passwords were higher in week 2 (US1), compared to the remaining passwords (in each condition). It is reasonable to assume that passwords which are memorable are likely to have an elaborative encoding in the memory because they are meaningful to the user. A meaningful password, which is memorable, might be described accurately enough to be guessed by an attacker. Hence it is maintained that using the twenty most memorable passwords for the guessability study, would help to assess, whether memorable RBGS passwords can be guessed using their respective textual descriptions.

Figures 4.1 – 4.4 presents a sample password of each image type together with the description of the target images provided by the respective account holder (in italics).

(Left to right) ***Image 1:*** *A blue woodpecker;* ***Image 2:*** *One light green and one dark green tree with blue stars;* ***Image 3:*** *two people boxing, one in red and other in blue;* ***Image 4****: A globe with green color and a red heart in the right side*

Figure 4.1: Mikon password descriptions

(Left to right) ***Image 1:*** *A hut with two windows and one door;* ***Image 2:*** *A boat on the sea;* ***Image 3:*** *SUN;* ***Image 4:*** *Music symbol*

Figure 4.2: Doodle password descriptions

(Left to right) ***Image 1:*** *Two people – one man and one women, golden color;* ***Image 2:*** *Red color both light and dark and yellow color at top center;* ***Image 3:*** *Red background and then tint of black and a tree in the center;* ***Image 4:*** *Chinese art with 3 people and the middle one wearing black and red*

Figure 4.3: Art password descriptions

(Left to right) ***Image 1:*** *A silver bridge with a tint of pink at night;* ***Image 2:*** *A football stadium;* ***Image 3:*** *A yellow drink served with lime and mint leaves;* ***Image 4:*** *A red man holding one dice and 5 dices in the image*

Figure 4.4: Object password descriptions

4.4 Results

The independent variables in the guessability study (GS1) were the four different image types used (condition): Mikon; doodle; art; object. The dependent variables and the corresponding results are discussed below.

4.4.1 Performance of the Attackers in all Conditions

Firstly, MLSP (mean login success percentage) of each attacker in each condition is calculated using (Eq. 4.1). Then MLSP for each condition, i.e. SP2 is calculated using (Eq. 4.2). The descriptive statistics for the measure SP2 is presented in Table 4.4.

$$S_i = \frac{\textbf{Number of passwords identified correctly}}{\mathbf{4}} \% \qquad \text{(Eq. 4.1)}$$

Each user (i) had to guess four passwords in each condition.

$$\mathbf{SP2(j)} = \sum_{i=1}^{70} S_i / \mathbf{70} \qquad \text{(Eq. 4.2)}$$

Each condition (j) had i=1 to70 participants (repeated measures design).

SP2 for all the conditions was not normally distributed, as assessed by a *Shapiro-Wilk test*. A non-parametric test for independent measure was used to test the statistical significance because the attacks were partially independent, even though each of the attackers performed in all the conditions. The attacks were partially independent because all the attackers did not

guess the same password. According to the experimental protocol, each group guessed a different set of passwords in each of the four conditions as shown in Table 4.2.

Condition	SP2%	Standard Deviation (SD)	Standard Error (SE)	Median	Range
Mikon	52.14	19.38	2.31	50	25-100
Doodle	80.00	17.34	2.07	25	50-100
Art	42.50	17.21	2.05	50	25-75
Object	88.57	16.85	2.01	100	50-100

Table 4.4: Descriptive statistics for performance of the attackers in GS1

A *Kruskal Wallis* test confirmed that SP2 was significantly affected by the type of images used as the password [$\chi 2\ (3) = 133.08$, $p < 0.001$)]. The *Mann Whitney* post hoc test showed significant differences between all pair of conditions, except doodle-object. According to the descriptive statistics presented in Table 4.4 and the statistical tests, the order of decreasing guessability is: *Object* ≥ *Doodle* > *Mikon* > *Art*

4.4.2 Login Success Percentage of Each Group

The mean login success percentages (MS1) of each group (1-5) in each condition in the guessability study are presented in Table 4.5. Since there are more attackers (70) than the passwords (20), the statistical results for just one measure (performance of attackers in all conditions) might be considered deceptive. Hence, MS1 for each group (in each condition) was measured to make the analysis more conservative, and devoid of any deception.

Condition	Group 1		Group 2		Group 3		Group 4		Group 5	
	MS1	SE	MS1	SE	MS1	SE	MS1	SE	MS1	SE
Mikon	51.78	4.87	62.5	4.34	50	5.24	48.21	6.12	48.21	4.87
Doodle	82.14	3.13	78.57	5.77	82.14	4.08	82.14	4.85	75	5.24
Art	46.42	5.17	42.85	4.85	41.07	4.97	39.28	4.31	42.85	4.08
Object	91.07	4.97	89.28	4.31	89.28	4.31	87.5	4.34	85.71	5.05

Table 4.5: Descriptive statistics for login success percentage of each group

MS1 for each of the five groups in each condition was not normally distributed, as assessed by a *Shapiro-Wilk test*. This measure analyzes the statistical difference between groups in each condition (i.e. the within group measure). Given the use of the independent design with five groups in each condition and the non-normal distribution of the data, a *Kruskal Wallis test* was used to analyze the statistical significance. The statistical test results confirmed that the variation in MS1 between the five groups in case of Mikon, doodle, art and objects was not statistically significant. In other words, the guessability of the different groups in each condition was not significantly affected, although each group guessed different passwords using the corresponding password descriptions. It may be argued that the ability of the subjects to describe their RBGS passwords may be different, which would significantly affect the guessability. But, the results presented in Table 4.5 as well as the statistical tests, did not reveal any significant performance differences between different groups in each condition.

4.4.3 Number of Passwords Guessed

The results reported in Section 4.4.1 do not reveal much on the actual number of passwords guessed during the attack. In this section, the guessability of the passwords has been discussed in terms of the percentage of the attackers guessing one, two, three and four passwords (Table 4.6), as well as the total number of passwords guessed during the attack.

Conditions	**Number of attackers guessing**			
	1 password	**2 passwords**	**3 passwords**	**4 passwords**
Mikon	14	40	12	4
Doodle	0	11	34	25
Art	30	31	9	0
Object	0	7	18	45

Table 4.6: Password guessing trend

For example, according to Table 4.6: 1 password was guessed by 14 attackers (1password ×14 =14 times); 2 passwords were guessed by 40 attackers (2passwords ×40 =80 times); 3 passwords were guessed by 12 attackers (3passwords ×12 =36 times); 4 passwords were guessed by 4 attackers (4passwords ×4=16 times). According to the experimental protocol, in total 20 passwords were attacked 280 times (20 passwords × 14 attckers). Hence the success

rate for Mikon condition is (14 + 80 + 36 + 16)/280, which is same as SP2% of Mikon in Table 4.4 (52.14%).

- *Mikon:* 20% of the attackers guessed at least one password, 57.1% of the attackers guessed at least two passwords, 17.1% guessed at least three passwords and 5.7% guessed all the four passwords. The analysis reveals that each of the 20 Mikon passwords (100%) was guessed at least once during the guessability attack**.**

- *Doodle:* 15.7% of the attackers guessed at least two of the four passwords, 48.6% guessed three passwords and 35.7% were able to guess all the four passwords. The analysis reveals that each of the 20 doodle passwords (100%) was guessed at least once during the guessability attack**.**

- *Art:* 42.9% of the attackers guessed only one password, followed by 44.3% of the attackers guessing two passwords and 12.9% guessed three passwords. None were able to guess all the four passwords; Out of 20, ten (50%) of the art passwords were guessed at least once in the guessability study.

- *Object:* 10% of the attackers guessed two passwords, followed by 25.7% guessing three passwords and 64.3 % guessed all the four passwords. The results revealed that each of the 20 object passwords (100%) was guessed at least once, during the guessability attack.

All the subjects in US1 had produced descriptions for each of their RBGS passwords twice (as discussed in Section 4.3). In this context, further analysis shows that:

- 90-100% of the passwords guessed during the guessability attack in case of Mikon, doodle and objects had three-four target images described as sketches, during the Stage 1 (Section 4.3.1) of description collection in US1;
- 40% of the art passwords guessed during the guessability study had one-two target images described as sketches in the Stage 1 of description collection in US1;

Please note that the descriptions used in the guessability study were collected from Stage 2 (Section 4.3.2) only, i.e. textual descriptions.

4.4.4 Passwords with Denotative Descriptions

The subjects in US1 were asked to choose a category (from a list given to them) for the textual descriptions provided for each target image forming their respective passwords. The results obtained from the subjects are presented in Table 4.7. The list given to the subjects, once they had written a description for each of the sixteen target images are given below.

- Associative: describing the elements in the image by means of association to other objects based on their personal knowledge and perception;
- Emotive: describe the way the elements in the image make them feel, something personal;
- Denotative: describing the elements in the image or distinct components in the image;
- Other: any other category not listed above

The categories were influenced from the literature presented in Chapter 2 (Section 2.1.3). Sturken & Cartwright (2012) suggested that images can have two levels of meaning: (1) d*enotative*; (2) *connotative.* In this context, connotative meaning is represented by the associative and emotive categories listed above.

	Number of descriptions categorized as			
	Associative	Emotive	Denotative	Other
Mikon	28	40	332	0
Doodle	39	12	349	0
Art	0	70	330	0
Object	30	24	346	0

Table 4.7: Categorization of descriptions

The statistics in Table 4.7 highlights that the percentage of images having denotative descriptions in the case of Mikon is (87%), doodle (87.25%), art (82.5%) and object (86.5%). These results show that most subjects described the elements in the target image, while recording a written prompt of their RBGS passwords, in the current experimental settings. Denotative descriptions are interesting in the sense that they can help an attacker to guess the target images. Further analysis reveals that, the number of passwords having denotative descriptions for all the four target images is more than 60% for each of the image types (highest 75% in case of objects), as shown in Table 4.8.

	Passwords with denotative descriptions for			
Categories	**4 images (4i)**	**3 images (3i)**	**2 images (2i)**	**1 image (1i)**
Mikon	62	28	0	0
Doodle	66	25	5	0
Art	64	18	10	0
Object	75	10	8	0

Table 4.8: Number of passwords having denotative descriptions

The statistics in Table 4.9 shows that, all the 20 Mikon, doodle and object as well as the 10 art passwords which were guessed in GS1, have denotative descriptions for at least three-four target images.

	Passwords used in guessability study having denotative descriptions for			
	4 images (4i)	**3 images (3i)**	**2 images (2i)**	**1 image (1i)**
Mikon	11 (guessed)	9 (guessed)	0	0
Doodle	15 (guessed)	5 (guessed)	0	0
Art	6 (guessed)	4 (guessed)	10 (not guessed)	0
Object	13 (guessed)	7 (guessed)	0	0

Table 4.9: Number of passwords with denotative descriptions used in GS1

4.5 Discussion

This chapter presented a study (GS1) that examined the guessability of RBGS passwords, using the corresponding textual descriptions, which were provided by the respective account holders. The results obtained in GS1 demonstrated that all the twenty Mikon, doodle and object passwords were guessed at least once using the corresponding password descriptions, whereas only 50% of the art passwords were guessed. Hence, textual descriptions given by the subjects in US1 were effectively used to guess RBGS passwords in the given experimental set-up.

The statistics presented in Table 4.3 shows that most subjects did not record art passwords in the form of sketches, unlike Mikon, doodle and object passwords. This could be attributed to the fact that art images are difficult to draw and visually complex, compared to the other

image types. GS1 also shows that it is easier to guess RBGS passwords, if the textual descriptions are denotative (Table 4.9). The next paragraph will give a brief overview of some cognitive theories that would help to analyze the results obtained in GS1.

Visual complexity is an essential characteristic which distinguishes various image types (Renaud, 2009). According to Szekely & Bates (1999), the ability to describe an image would depend on its familiarity and visual complexity. According to the visual search process reported in (Wolfe, 1994; 2003), the representation of an image in the human brain is matched to the images that are present in the challenge set, to find the target image. According to Gilchrist & Harvey (2000), the viewer/describer of an image will search for the objects or features that illustrate its meaning. Hence, some components of an image may be omitted in the description, and an image which is visually less complex as well as more meaningful will be processed quickly. Greisdorf & O'Connor (2002) suggested that a picture is processed in the human brain at three hierarchical levels: (1) the primitive features are sensed in terms of the color, shape and texture; (2) objects in the images are isolated and identified; (3) there is an inductive interpretation.

The aforementioned cognitive theories suggest that the effectiveness of guessing an image would depend upon the ability of an attacker, to translate the description into a correct representation of the target image. This would depend upon the description quality, which in turn will depend on the meaningfulness and visual complexity of the image. In the case of textual descriptions, the attackers have static information. A textual description may certainly omit some details and may not be an accurate representation of the target image. The results in GS1 could be explained using the above theories. Art images are visually most complex (i.e. contain different colors, shades, texture and lot of information), and therefore it is difficult to describe clearly, which makes them the least guessable. But, the object images are distinctive and meaningful (i.e. images one knows about - familiar, uses in daily life), and therefore they are easier to describe, which makes them highly guessable.

Dunphy et al. (2008) is the only known study that has examined the guessability of face passwords using verbal descriptions. The descriptions for the guessability study were collected from nine male and female contributors, who never used the faces as their passwords. The descriptions used in GS1 were written by the respective account holders (subjects who took part in US1), in order to simulate a real life scenario. The results reported

in Dunphy et al. (2008) showed that out of 158 authentication attempts made by 56 participants, only 13 (8%) were successful, which is much lower compared to the results reported in GS1.

4.6 Study Limitations

One deficiency with the kind of experiment reported in this chapter is that it relies on unskilled participants, i.e. they are not necessarily representative of the people who would be trying to break into the system. On the other hand, GS1 examined the likelihood of guessing RBGS passwords using their corresponding textual descriptions, which the legitimate users have recorded in case they forget the password. They may misplace the recorded description or share it with someone they know. In this case someone other than the legitimate user, not necessarily a skilled hacker may try to log into the system. In terms of improvement, the same study could be re-done with a different population other than students. Further improvements in the context of GS1 will be discussed in Chapter 7 (Section 7.3.3).

Most textual descriptions given by the subjects in US1 were denotative. This could be due to the strict experimental set up and protocol used in GS1, which required the subjects to provide textual descriptions. Hence the chapter does not claim that subjects preferred to write denotative descriptions of target images. However, the results did show that it is easier to guess the RBGS passwords using the denotative descriptions. We do not claim that denotative descriptions are more guessable than non-denotative ones, as the experiment was not designed to demonstrate it.

GS1 was conducted with twenty passwords for each image type, which were most memorable after week 2 in US1. Hence another limitation is that the passwords used in GS1 may have influenced the guessability results. In this context, Table 4.8 shows that (82% - 91%) of the Mikon, doodle, art and object passwords had denotative descriptions for at least three to four target images. Table 4.9 further shows that all the Mikon, doodle, art and object passwords having denotative descriptions for three-four target images were guessed in GS1. Since most passwords were recorded using denotative descriptions, it is likely that the guessability results in the given experimental setting would be similar, if any other password from the collection was used.

In the absence of any related study of this kind, it was impossible to produce a flawless experimental design. But we ensured that:

- The descriptions are written by the *accountholders,* who were actually using the images as passwords. This ensured that the account holders are describing their own passwords and not merely images;
- Clear login and task instructions were provided to the attackers;
- A large sample size was used and a proper experimental protocol was selected;
- Pilot studies were performed to ensure the experimental design was appropriate and feasible to study the phenomenon of guessing RBGS passwords using written descriptions;
- Statistical tests were properly chosen and used to test significance in data.

4.7 Conclusion

We identified a potential research problem, i.e. *guessability of RBGS passwords using their textual descriptions,* from the literature review presented in Chapters 2. The research problem reflected one of the human factors in security, i.e. the vulnerability of RBGS password to the recorded textual descriptions. The focus of the research presented in this chapter was also to find an image type, which is least guessable using textual descriptions. This will ensure that even if users share/record their RBGS passwords, it won't be easily guessed.

The results obtained in GS1 show that users recorded RBGS passwords in the form of sketches, rather than textual descriptions, in spite of the instructions given to them. However, more studies need to be conducted with different experimental set-ups, to understand various coping mechanisms to record/share RBGS passwords. The results demonstrate that subjects in US1 described the elements in the target images (denotative descriptions) forming their respective passwords. The results also show that all the passwords (used in GS1), which were recorded using denotative descriptions, are guessed successfully by the attackers. Hence, denotative descriptions would increase the guessability of RBGS passwords. However, further studies need to be conducted to validate this claim. We conclude that in the experimental set up used in GS1, art passwords are most resistant to guessability using textual descriptions.

The issue of vulnerability of image passwords to descriptions warrants several levels of investigation. A user may create passwords using mnemonic strategies. A mnemonic strategy may help to produce meaningful and clear descriptions, which can be easily guessed by an adversary. Hence, the vulnerability of RBGS passwords (created using a mnemonic strategy) to recorded descriptions also need to be investigated. A guessability study (GS2) with mnemonic passwords is reported in Chapter 5. This will further advance the research in the context of the guessability of RBGS passwords using descriptions.

Chapter 5
A Study of Multiple Story Passwords

This chapter reports a usability study (US2), which was conducted with 80 subjects, who were required to create four RBGS passwords, each using a mnemonic strategy and recall them every week, for a period of four weeks. A guessability study (GS2) which was conducted with 70 participants to examine the vulnerability of RBGS story passwords to written descriptions is also reported in this chapter. This chapter addresses Objectives 3 and 4 (corresponding to the Stages 4 and 5 respectively), which have been already discussed in Chapter 1 (Section 1.4). The contents presented in this chapter have been published in the proceedings of the 12th Annual Conference on Privacy, Security and Trust (PST 2014).

5.1 Introduction

The techniques used to retrieve information from the memory are called *mnemonics* (Parkin, 1993). Mnemonics help to convert the information in hand to a form that can be better retained in the memory for future use. In this chapter, the image passwords created using a mnemonic strategy will be referred to as *story passwords.*

The use of mnemonic strategy to choose password images was first studied in Davis et al. (2004), which reported a single-password usability study, comparing the story passwords (visual cue- object images) with face passwords, and actually found that the former were harder to remember. One potential reason for this negative result was that nearly 50% of the story users reported choosing no mnemonic strategy, despite the instructions given to them. Moncur & Leplatre (2007) explored the memorability of multiple story passwords, again using object images as the visual cue. According to the statistics reported in Moncur & Leplatre (2007), 76 participants were assigned story passwords, but the dropout rate was almost 65%, which made it difficult to get concrete results.

In Chapter 3 of this thesis, the post study questionnaire results (Section 3.5.4) suggest that many subjects in US1 employed a mnemonic strategy to create and remember their multiple RBGS passwords, in spite of no advice given to them. However, US1 did not consider the effect of employing a mnemonic strategy on the effectiveness (memorability) and efficiency (registration and authentication time) of the RBGS passwords. Hence, this chapter reports a user study (US2) to examine the effectiveness and efficiency of multiple story passwords in

RBGS, which will address Objective 3 that has been discussed in Section 1.4. The chapter also explores the vulnerability of story passwords to descriptions provided by the respective account holders, to address Objective 4 that has already been presented in Section 1.4.

5.1.1 Terminologies

For clarification, the definitions of aspects related to the usability and guessability study reported in this chapter, which will be used frequently are as follows.

- *Account holder:* A legitimate user who has registered with an image password and proves the authority using the same password.
- *Description/ written description*: A description of the target images forming the RBGS story password, which is provided by an account holder. The descriptions can take the form of words (text) or sketches (with/without annotations), provided by the respective account holders only.
- *Subject:* An account holder who has participated in the usability study (US2). The subjects who took part in US2 are different from that of US1.
- *Attacker:* A participant in the guessability study (GS2), who is trying to guess the target images forming a RBGS password, using the corresponding descriptions given by the respective subject. The attackers participating in the guessability study (GS2) were different to the subjects who took part in the usability study (US2).

5.1.2 Contributions

The main contributions of this chapter are:

- The effectiveness and efficiency of the multiple story passwords is examined, and compared to the existing studies that have reported the usability of multiple graphical passwords;
- The guessability of the story passwords using the corresponding descriptions given by the subjects who took part in usability study (US2) is also examined;

5.2 Usability Study of Story Passwords (US2)

A usability study (US2) was conducted to investigate the effectiveness in terms of memorability, and the efficiency in terms of the registration time and the time taken for successful authentication attempts, of multiple story passwords. The number of story passwords each subject had to remember was four (same as that of US1).

5.2.1 Recruitment of the Subjects

In order to recruit the participants (subjects) for this usability study (US2), emails were sent to the first and second year student email lists in a university. The mail comprised of the same components as discussed in Section 3.4.1.

A total of 128 subjects volunteered to take part in the study, by responding to our email. Out of these 128, 5 subjects were randomly approached to take part in a pilot study over a period of two weeks. The aim of the pilot study was to ensure that:

- Subjects can use the instructions easily, to successfully register and authenticate in the RBGS;
- The instructions given in the context of employing mnemonic strategies to create their passwords is understandable;
- They can clearly understand the tasks from the documentation provided to them and the emails that were framed for the purpose of the study;
- The post-study questionnaire was interpreted correctly, and they can clearly understand the terminologies used to frame the questions;

Once the pilot study was completed, an email was further sent to the remaining 123 subjects, to confirm their participation. 89 subjects responded to the mail and agreed to participate by signing the consent form, once they were given all the relevant documentation (registration and login process, detailed description of the task each week). The remaining 34 subjects, either did not respond to our email or decided not to take part in the study. Of the 89 subjects who confirmed their participation, 9 subjects opted out of the study, during the registration stage. Hence the usability study was completed by 80 subjects. The dropout rate here, i.e. those subjects who did not complete the study, after confirming their participation is 10.2 %. For the purpose of the research reported in this chapter, the demographic information of all

the subjects who completed the study (i.e. 80 subjects) has been reported in Section 5.2.2. The results reported in Section 5.3 do not include the subjects who did not complete (i.e. 9 subjects who dropped out of US2). Table 5.1 summarises the number of subjects who responded during each stage of the recruitment process.

Stage	**Subjects responded**	**Comments**
Email invitation	128	
Pilot Study	5	123 subjects left for US2
Confirmation to participate	89	34 did not respond
Study completed	80	9 dropped out

Table 5.1: Recruitment information for the user study US2

5.2.2 Subject Demographics

The usability study was conducted with 80 subjects (Female: 28; Male: 52; age range: 19-24 years), who were undergraduate students studying various degree programmes as given below:

- Civil Engineering-17;
- Humanities-19;
- Information technology - 15;
- Physical/Chemical Science-14;
- Earth Sciences-15;

None of the subjects were experts in the field of usable security, or studying this topic as a part of their curriculum. Ethics approval *(Ethics no 00942)* was granted by the college ethics committee to conduct the usability study (US2).

5.2.3 Study Protocol

The experimental framework used for US2 is similar to the ones reported in the literature (Moncur & Leplatre, 2007; Chiasson et al., 2009). In most of the existing (single or multiple

password) usability studies with GASs, subjects had to first complete the registration process followed by a number of authentication attempts, which is generally known as the training stage. This is followed by memorability test stage(s), which are conducted after a considerable gap (1-2 weeks), without any practice session in between the two stages. A similar protocol is followed in US2, which is further discussed in Section 5.2.4.

US2 was conducted over a period of five weeks using an independent style design with four conditions as given below.

- *Mikon:* register and authenticate with four Mikon images as password.
- *Doodle:* register and authenticate with four doodle images as password.
- *Art:* register and authenticate with four art images as password.
- *Object:* register and authenticate with four object images as password.

Each condition was allocated the same number of subjects (i.e. twenty). Each subject was randomly assigned to exactly one condition. Each subject was also given:

- an instruction sheet explaining the registration and authentication stages;
- one task sheet in each week, which contained the respective week's task, i.e. number of login sessions for each password.

The RBGS prototypes used in US2 were the same as discussed in Chapter 3 (Section 3.3), but with the following changes:

- The names of all the hyperlinks representing the four specific accounts were changed to: *Office Email; Online Shopping; Personal Email and Social Network*. This was done to add more meaningful contexts to the individual accounts;
- All the four image collections for each of the image types were made distinct;
- The Mikon collection for the third link (see Table 3.1 in Chapter 3) was changed, i.e. images with annotations were replaced by images without any annotations. The aforementioned change is reasonable, as it would be easier to guess images (with annotations), using their corresponding descriptions.

5.2.4 Study Framework

The user study comprised of a single training session and four retention stages as given below.

Training stage

In the first week, each subject was asked to register with four passwords, for one of the conditions assigned to them. For example, the subjects assigned to the Mikon condition used Mikon images as their password. Each subject created four Mikon passwords, using four distinct image collections (each having 150 different images), and authenticated using the respective passwords. Each password comprised of four Mikon images drawn from one collection. Similarly, subjects in the doodle, art and object conditions used their respective types as passwords. The training stage consisted of four lab sessions, performed by the subjects in the presence of the experimenter. This ensured that each password was created on a different day in the same week.

The subjects were advised to use a mnemonic strategy to create all their passwords to aid memorability. The subjects were explained the meaning of the term mnemonic strategy using some examples such as, selecting target images of the same colour, target images defining a mood. Each subject was also instructed that a mnemonic strategy should help them to aid memorability of multiple passwords, but should not help anyone else to guess the target images forming the password. The instruction was given to make sure that subjects don't choose a mnemonic strategy, which would compromise the security of the system easily. The aforementioned instruction could be considered as a RBGS password advice to the respective account holders.

Upon completing the registration stage, each subject had to login five times (rehearsal), with each password and then answer a short questionnaire. The system would display the correct target images forming the password, in the case of three failed login attempts (not necessarily continuous), as this was a training session for the subjects, to rehearse their passwords. Due to the voluntary nature of participation, each lab training session involved (1 password × 5 rehearsals), to alleviate the risk of subject attrition. Moreover, the login statistics (i.e. successful attempts and login time) obtained from this stage was not considered to analyse the results. The short questionnaire asked the subjects to write down the mnemonic strategy used to create each of their passwords, and provide a description (maximum twenty five

words) for each of the four target images forming the respective passwords. The idea behind the questionnaire was to gather evidence that a mnemonic strategy was employed to create the RBGS passwords. Based on the responses to the questionnaire, we found that all the subjects have employed a mnemonic strategy to create their RBGS passwords. The results corresponding to the type of mnemonic strategy employed by the subjects has been further discussed in Section 5.3.3. The instructions given to the subjects for writing their descriptions were exactly the same as in US1 (Chapter 4, Section 4.2.4).

The training stage was completed in person by the subjects in front of the experimenter to ensure:

- they did not face any problems while registering and logging into the system;
- the descriptions written by the subjects were not used as a coping mechanism to remember the password images in the subsequent use;
- subjects understood the tasks to be completed in the weeks to follow.

Retention stages

Memorability tests were carried out in week 2 (gap of 7 days after creating the last password in week 1), 3, 4 and 5 respectively. In each test, subjects had to login with each of their four passwords five times. The order of logging into the accounts was randomized for each week. This stage was not completed in the presence of the experimenter. The subjects were allowed to complete the authentication sessions online from their home/accommodation. At this stage, a lockout policy was implemented for four failed login attempts (not necessarily continuous). There were no practice sessions in between any of the memorability tests. On completing the memorability test in week 5, each subject was asked to answer an online exit questionnaire.

5.3 Usability Study Results

The independent variables in US2 were the four different image types Mikon; doodle; art; object. The dependent variables and the corresponding results are discussed below.

5.3.1 Effectiveness

The mean login success percentage (SP3) of each subject in each condition was calculated using Eq. 5.1. SP3 did not consider the successful attempts made during week 1 (training).

$$SP3 = \frac{Number\ of\ successful\ login\ attempts}{Total\ number\ of\ login\ attempts}\% \qquad \text{(Eq. 5.1)}$$

The descriptive statistics for the measure SP3 are presented in Table 5.2 and box plots are shown in Figure 5.1.

Conditions	**SP3 (%)**	**SD**	**SE**	**Median**	**Range**
Mikon	63.25	6.5	1.46	65.35	51.42-74.28
Doodle	61.85	5.66	1.26	63.21	49.28-69.28
Art	44.6	4.85	1.08	44.28	35.71-54.28
Object	75.4	3.43	0.76	75.71	67.14-81.42

Table 5.2: Descriptive statistics for mean login success percentage (SP3) in US2

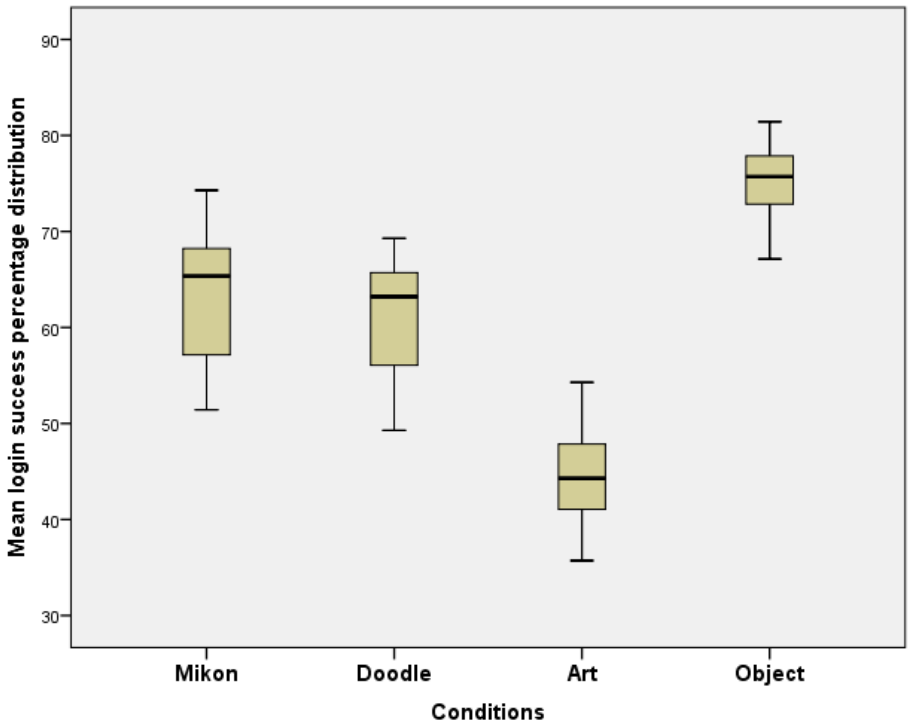


Figure 5.1: Box plot representation for mean login success percentage (SP3) in US2

A *Shapiro-Wilk* test confirmed that SP3 for all the conditions was normally distributed. An independent measure ANOVA confirmed significant differences between all the conditions. A Tukey post hoc test confirmed significant differences between each pair of conditions, except in the case of Mikon-doodle (p=0.64). According to the descriptive statistics presented in Table 5.2 and the significance test results, the order of decreasing memorability is *Object* > *Mikon* ≥ *Doodle* > *Art.*

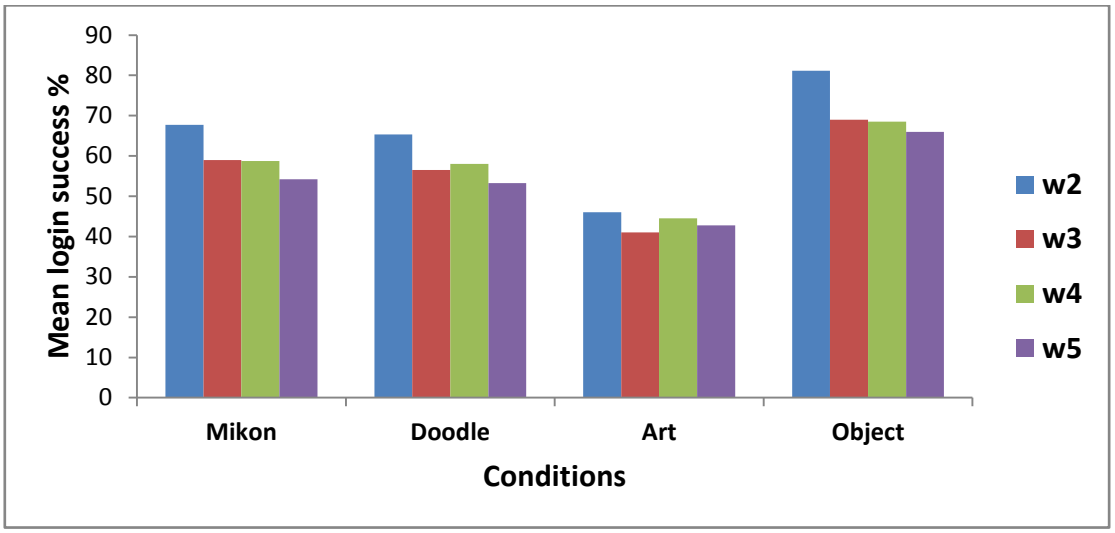


Figure 5.2: Weekly mean login success percentages for each condition in US2

Figure 5.2 presents the mean login success percentages from week (w) 2 to week 5, for each condition. The results indicate that the mean success percentage drops from week (w) 2 to 5: 13.93% in the case of Mikon; 12.06% in the case of doodle; 3.25% in the case of art; and 15.12% in the case of object passwords. The decrease in memorability in the case of Mikon, doodle and object passwords are similar. But in the case of art passwords (lowest SP3), the difference in the mean login success percentage (week 2 - 5) is very low compared to the other three image types.

5.3.2 Efficiency

The mean registration time for creating four story passwords (RegT2) and mean authentication time for successful login attempts (AuT2) were assessed to examine the efficiency of the RBGS.

(A) Registration time

The registration time for a story password is the time taken to go from screen 3 to screen 5 during the registration process, as shown in Figure 3.5 in Chapter 3. This would also include the time taken to employ a mnemonic strategy. The mean registration time (RegT2) for each condition is calculated using Eq. 5.2 as given below.

$$RegT2 = \frac{1}{4}\sum_{i=1}^{4} \begin{matrix} Registration\ time\ for\ password\ (i), \\ where\ i\ denotes\ password\ 1,2,3,4 \end{matrix} \quad \text{(Eq. 5.2)}$$

The descriptive statistics for RegT2 in each condition is presented in Table 5.3. The box plots for the distribution of registration time in each condition are shown in Figure 5.3.

Conditions	**RegT2 (sec)**	**SD**	**SE**	**Median**	**Range**
Mikon	72.05	2.48	0.55	71.5	67.5 - 76.25
Doodle	70.7	1.86	0.41	71	66.75 - 73.25
Art	86.11	1.54	0.34	86.25	83.75 - 88.5
Object	59.68	1.69	0.37	60	57.25 - 62

Table 5.3: Descriptive statistics for mean registration time (RegT2) in US2

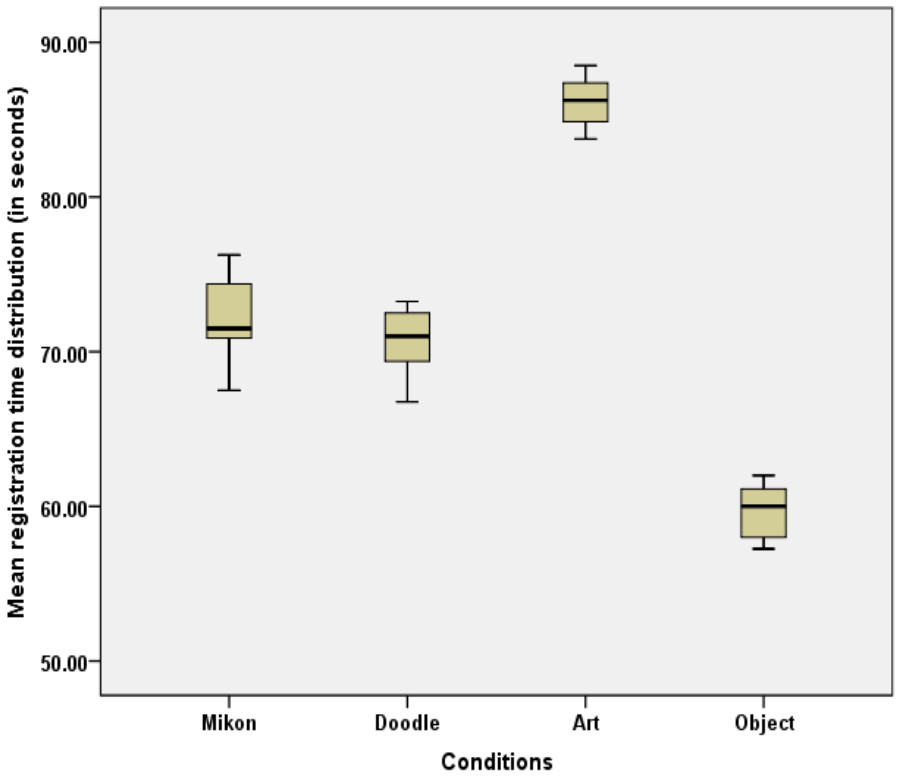


Figure 5.3: Box plot representing the registration time distribution in US2

A *Shapiro-Wilk* test confirmed that RegT2 for all the conditions were normally distributed. An independent measure ANOVA confirmed significant differences between all the conditions. A Tukey post hoc test confirmed significant differences between each pair of conditions, except in the case of Mikon-doodle (p=0.13). According to the descriptive statistics presented in Table 5.3 and the statistical tests, the order of increasing RegT2 is: *Object* < *Doodle* ≤ *Mikon* < *Art*

(B) Authentication time

The authentication time for a story password is the time taken to go from screen 2 to the success notification screen, during the authentication process as shown in Figure 3.6 in Chapter 3. The mean time for successful authentication attempts (AuT2) for each condition is calculated using Eq. 5.3, where z represents the total number of successful login attempts.

$$AuT2 = \frac{1}{z}\sum_{n=1}^{z} Login\ time\ for\ successful\ login\ (n) \quad \text{(Eq. 5.3)}$$

The descriptive statistics for AuT2 in each condition is presented in Table 5.4. The box plots for the distribution of the authentication time in each condition are shown in Figure 5.4.

Condition	**AuT2 (sec)**	**SD**	**SE**	**Median**	**Range**
Mikon	20.55	2.91	0.6	20	16-25
Doodle	21.95	2.68	0.65	22.5	18-26
Art	26.65	2.6	0.51	26	21-32
Object	20.35	1.75	0.39	20.5	18-24

Table 5.4: Descriptive statistics for mean authentication time (AuT2) in US2

A *Shapiro-Wilk* test confirmed that AuT2 for each of the conditions did not have a normal distribution. A non-parametric Kruskal Wallis test confirmed no significant difference between the conditions. A Mann Whitney post hoc test confirmed that there are no significant differences between each pair of conditions. According to the descriptive statistics presented in Table 5.4, the order of increasing AuT, without any significant difference is: *Object* ≤ *Mikon* ≤ *Doodle* ≤ *Art*

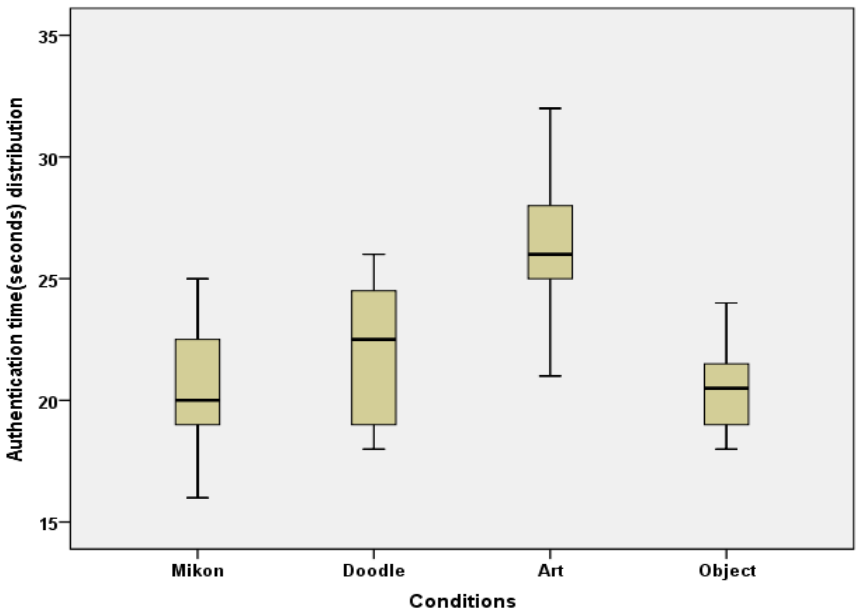


Figure 5.4: Box plot representing the authentication time distribution in US2

5.3.3 Categories of Mnemonic Strategies

Each subject was asked to categorize all the four mnemonic strategies used to create their story passwords. The categories to choose from were:

- personal, something personally related to the subjects;
- meaningful, a meaningful strategy but not personally related;
- Other, i.e. any random strategy which subjects do not believe to be either personal or something meaningful to them. Since a random strategy could be anything and would differ from one user to another, we did not classify this category further.

Condition	Personal	Meaningful	Other
Mikon	40	32	8
Doodle	48	24	8
Art	16	24	40
Object	60	8	12

Table 5.5: Statistics regarding categories of mnemonic strategy chosen by the subjects in US2

The result presented in Table 5.5 confirms that all the subjects in US2 employed a mnemonic strategy to create their password. According to these statistics:

- 50-75% of the mnemonics were personal, in the case of Mikon, doodle and object passwords, whereas only 20% were personal in the case of art passwords;
- 40% were meaningful in the case of Mikon, 30% in case of doodle and art, whereas only 10% in case of objects;
- 50% of the mnemonics were neither personal nor meaningful in the case of art passwords, 10% in case of Mikon and doodle, and 15% in case of Objects.

Overall 51.25% of the mnemonics were personal, 27.5% were meaningful, and 21.25% were neither personal nor meaningful. These results demonstrate that subjects tended to choose mnemonics which were personally related to them and this was also dependent upon the type of image used as the password.

5.3.4 Exit Questionnaire Results

All the subjects in US2 were asked to complete an online exit questionnaire, once they had finished their tasks in week 5. The results of the questionnaire are discussed below.

(A) User ratings: Each subject was asked to rate certain aspects (A-C) on a scale of 1 to 5, 1 being the poorest rating (highly negative), and 5 being the best rating (highly positive). The chosen aspects were:

(A) Ease of creating four mnemonic strategies;

(B) Ease of relating the mnemonic strategies to the target images forming the password or vice-versa;

(C) Ease of remembering four mnemonic strategies.

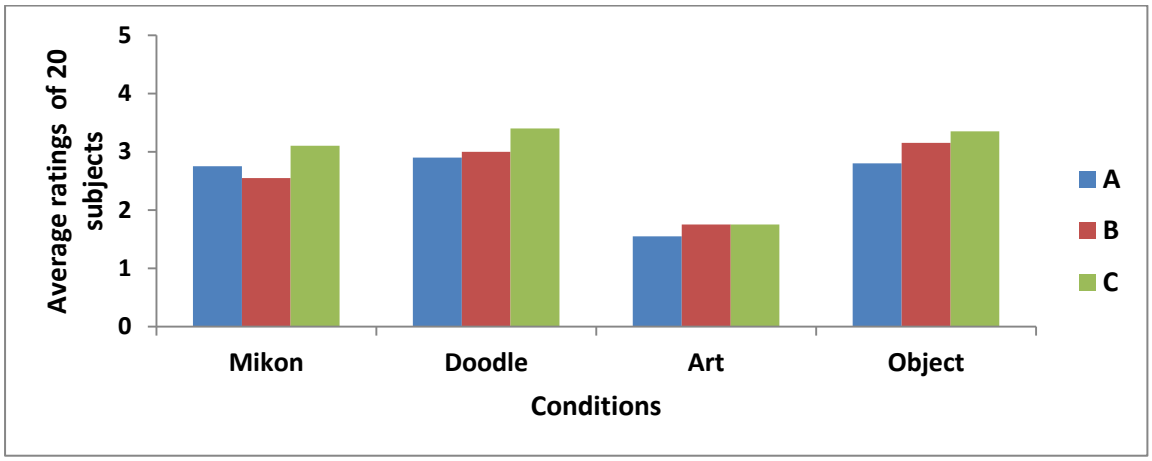


Figure 5.5: Mean ratings for each aspect obtained from the subjects

The results presented in Figure 5.5 demonstrate that subjects across all the conditions found it difficult to create four mnemonic strategies (average rating below 3). The same trend was observed in the case of other aspects (i.e. B and C) too. The subjects in the art condition gave the lowest ratings for all the three aspects, confirming that they found it difficult to create and remember the story passwords. This is in line with the effectiveness and efficiency results reported in the Sections 5.3.1 and 5.3.2 respectively.

(B) Reason for unsuccessful authentication: Each subject was asked to choose one or more reason for not being able to authenticate successfully from the following Type options:

- Type 1. Interference/Confusion between the four mnemonic strategies;
- Type 2. Forgot the strategy completely/ partially;
- Type 3. Remembered the strategy, but confusion with the target images;
- Type 4. Forgot both the strategy and target images;
- Type 5. Other reasons.

Options	**Mikon**	**Doodle**	**Art**	**Object**	**Total Responses for each Type**
Type 1	20	18	5	18	61
Type 2	12	10	13	12	47
Type 3	10	12	19	9	50
Type 4	6	5	10	5	26
Type 5	0	0	0	0	0
Total Responses for each condition	48	45	47	44	**Total Responses all conditions 184**

Table 5.6: Responses to reason for unsuccessful authentication in US2

The responses given by the subjects are presented in Table 5.6. For example, as the reasons for not being able to authenticate successfully, 20 (all) subjects in the Mikon condition chose Type 1, 12/20 subjects also chose Type 2, 10/20 chose Type 3, 6/20 subjects chose Type 4 and none of the subjects selected Type 5. Hence the total number of responses for this condition was 48 (i.e. 20 + 12 + 10 + 6). Since each subject could choose one or more types, the total number of responses for each condition was neither 20 (each subject selecting one type only) nor 100 (each subject selecting all the types). The total responses for each Type

(i.e. sixth column Table 5.6) are the sum of all the responses obtained for Type 1 in each condition. For example, out of 184 responses in total, 61 (i.e. 33%) responses obtained across all the four conditions, attributed Type 1 for unsuccessful authentication.

According to Table 5.6, the subjects assigned to Mikon, doodle and object conditions chose the same reasons for authentication failure in the exit questionnaire. 90-100% of the subjects in Mikon, doodle and object condition confused one mnemonic strategy with the other for a specific account (Type 1), but 95% (19/20) of the art users had problems associating the target images to the mnemonic strategy (Type 3).

Out of 184 responses (Table 5.6) received from the exit questionnaire:

- 33% confused one strategy with the other for a specific account (Type 1 error);
- 25.5% forgot the strategy completely/partially because it was randomly chosen (type 2 error);
- 27.1% were not able to recognize the target images, They believed that remembering four strategies and sixteen target images put excessive load on their memory (Type 3);
- 14.1% forgot both the strategy and corresponding target images because they felt the task was mentally demanding (Type 4).

The results implied that Types 1, 2 and 3 errors, predominantly made it difficult to remember multiple story passwords.

(C) *Improvements:* Each subject was asked to choose one or more improvements (A1-A4) from the list given below, which could make this system more usable:

- A1- Mnemonics should be typed and stored in the system together with the target images and displayed during authentication;
- A2- The user can write the mnemonics in any language (including English), when these are stored in the system;
- A3- The image collection during registration should provide more choices to select the target images;
- A4- The number of images in each challenge set should be decreased (e.g. 10-12 images) to make the system more usable.

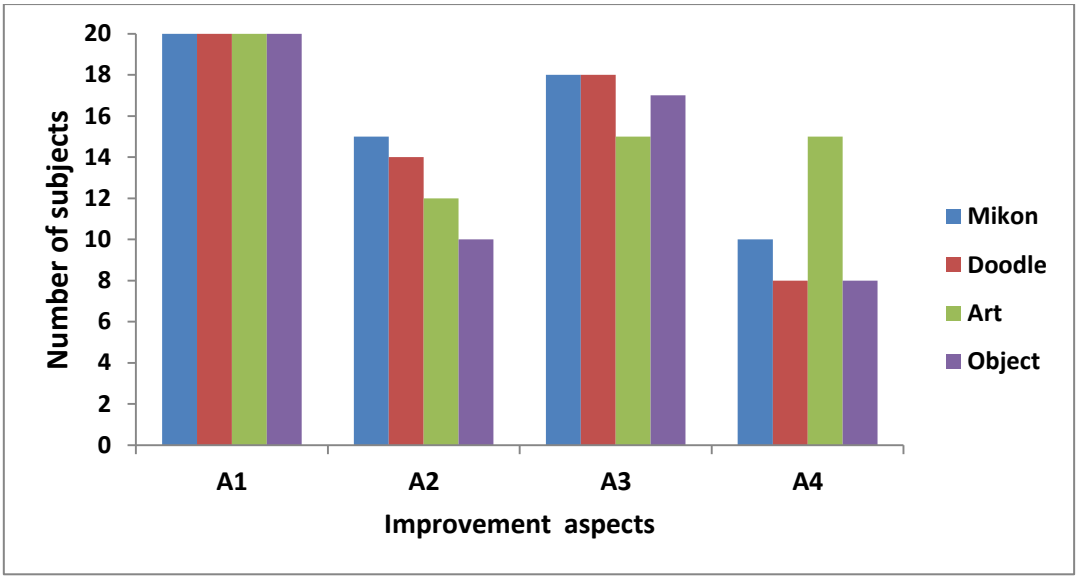


Figure 5.6: Suggested improvement responses

According to results shown in Figure 5.6, all the subjects in each condition (100%) believed that mnemonics should be typed and stored in the system, so that it is not required to remember them in the subsequent use. On average 63.75% of the subjects were in favour of writing the mnemonics in any language. 85% of the subjects felt that the registration stage should provide a huge collection to select the four target images because this would provide more choices to the user. 51.25% of the subjects believed that decreasing the number of images may enhance the usability of the system.

5.4 Guessability Study of Story Passwords (GS2)

The extent to which a story password can be guessed using its corresponding description has not been examined in any of the existing studies. A guessability study (GS2) was conducted to investigate, *whether it is possible to guess story passwords using their corresponding descriptions, provided by the account holder, in a given experimental setting?* GS2 differed to the guessability study (GS1) reported in Chapter 4 (GS1) in certain aspects as given below:

- Descriptions in the form of both sketches (with/without annotations) and/or words were acceptable in GS2. The subjects in US2 were not compelled to write textual descriptions, in case they drew sketches of the target images, in spite of the instructions given to them;
- Each password used in GS2 was created using a mnemonic strategy.

The threat model for GS2 is same as that of GS1, which has been already discussed in Chapter 4 (Section 4.1.2). The threat model presented in Chapter 2 (Section 2.7.3) also applies to this study.

5.4.1 Recruitment of the Attackers

In order to recruit the participants (attackers) for this guessability study (GS2), emails were sent to the third year, fourth year as well as the postgraduate students email lists in the university. The components of the mail were same as in Section 4.2.1. A total of 77 attackers volunteered to take part, by responding to our email. Out of 77, 5 attackers were approached to take part in a pilot study, which was conducted over a period of four days. The main aim of the pilot before conducting GS2 was the same as discussed in Section 4.2.1, i.e. to ensure that the attackers understood the tasks, relevant documentation and the instructions provided to them easily. Once the pilot study was completed, an email was further sent to the remaining 72 attackers, to confirm their participation. 70 attackers responded to the mail and agreed to participate in the study, by signing the participation consent form, once they were given all the relevant documentation (login process, detailed description of the guessability tasks) in the context of this study. The remaining 2 attackers did not respond to our email and hence were not contacted further. The guessability study was completed by 70 attackers. Table 5.7 summarises the responses received during each stage of the recruitment.

Stage	**Subjects responded**	**Comments**
Email invitation	77	
Pilot study	5	72 left for actual GS1
Confirmation to participate	70	no response from 2
Study completed	70	No one dropped out

Table 5.7: Recruitment information for the study GS2

5.4.2 Demographic Information of the Attackers

The study was conducted with 70 attackers (female: 32; male: 38; age range: 20-26 years), who were studying various degree programmes as given below:

- Computing Science -22;
- Psychology/related field – 19;
- Economics/Business Administration -12;
- Bioinformatics – 17.

The attackers were neither experts in usable security nor studying the topic as a part of their curriculum. The attackers in the guessability study were distinct from those who took part in the usability study (US2). The same ethics number as US2 was granted by the college ethics committee to conduct the guessability study (GS2).

5.4.3 Analysis of Descriptions

Most subjects in the usability study drew sketches (with/without annotations) of the target images instead of giving a word description (refer to Appendix D for an example). Table 5.8 reveals that the subjects preferred to draw sketches of the target images in the case of Mikon, doodle and object passwords more than the art passwords, in the given experimental set-up.

	Number of Passwords with Sketch for			
Category	**4 images (A)**	**3 images (B)**	**2 images (C)**	**1 image (D)**
Mikon (M)	42	15	11 M1-M11	12 M12-M23
Doodle (D)	58	12	8 D1-D8	2 D9-D10
Art (A)	0	0	4 A1-A4	6 A5-A10
Object (O)	45	17	12 O1-O12	6 O13-O18

Table 5.8: RBGS password described using sketches in US2

If it is assumed that passwords with three or all the four target images described as sketches are guessable, then according to the statistics presented in Table 5.8 (out of 80 passwords created for each image type in US2), 71.25% of the Mikon, 87.5% of doodle and 77.5% of the object passwords can be guessed easily by an attacker. This assumption is reasonable,

since it will be easier to guess a password using the corresponding sketches. Based on the assumption that, passwords with three or all the four target images described as sketches are guessable, *23 Mikon (M1-M23), 10 doodle (D1-D8), 80 art (A1-A80) and 18 (O1-O18) object* passwords were available to conduct the guessability study.

5.4.4 Guessability Study Framework

The experimental framework for this study is similar to the one described in Chapter 4. In GS2, seventy attackers were divided into five groups (G1-G5), i.e. fourteen in each group. Each attacker in a group had to guess four passwords for each image type using the corresponding descriptions, the same as the other attackers in that group. Since the number of passwords that could be used in the guessability study were less for certain image types (as demonstrated in Table 5.8), it was not possible to give distinct passwords to each group. The password allocation for each group has been illustrated in Table 5.9 and discussed below.

Condition	**G1**	**G2**	**G3**	**G4**	**G5**
Mikon	M1-M4	M5-M8	M9-M13	M14-M17	M18-M20
Doodle	D1,D2, D9, D10	D3,D4, D9, D10	D5,D6, D9, D10	D7,D8, D9, D10	D7,D8, D9, D10
Art	A1,A2 A5, A6	A3,A4 A7, A8	A9,A10 Aw1,Aw2	Aw3,Aw4 Aw5,Aw6	Aw7,Aw8 Aw9,Aw10
Object	O1,O2 O13, O14	O3,O4 O15,O16	O5,O6 O17,O18	O7,O8 O9,O10	O11,O12 O9,O10

Table 5.9: Password allocation for the guessability study (GS2)

- Mikon: All the groups received four distinct passwords for this condition. Hence the guessability study was conducted with twenty distinct passwords.
- Doodle: Each of the groups (G1-G4) were allocated two distinct passwords from category C (refer to Table 5.8 for the categories), and the same two passwords from category D. G5 received the same passwords as that of G4. Hence the guessability study was conducted with ten distinct passwords

- Art: G1 and G2 received two distinct passwords, both from category C and D respectively. G3 received two distinct passwords from Category D and two randomly chosen (Aw) distinct passwords from the remaining collection obtained from the usability study, which were described in words. Each of the groups G4 and G5 were allocated four distinct passwords, randomly chosen from the remaining collection obtained from US2. Hence the guessability study was conducted with ten passwords described using words only and ten passwords from Category C and D respectively.
- Object: Each of the groups (G1-G4) were allocated two distinct passwords from category C and category D respectively. G5 was allocated two distinct passwords from category C, but the remaining two passwords were the same as that in G4, chosen from category D.

GS2 was conducted online, so the experimenter communicated with the attackers via email, once the attackers gave their consent. The same communication protocol was followed as given in Chapter 4 (Section 4.2.3).

5.5 Guessability study results

The independent variables in the guessability study (GS2) were the four different image types used (condition): Mikon; doodle; art; object. The dependent variables and the corresponding results are discussed below.

5.5.1 Performance of the Attackers

The mean login success percentage (MLSP) of each attacker in a condition is calculated using (Eq. 5.4). Then MLSP for each condition, i.e. SP4 is calculated using (Eq. 5.5). The descriptive statistics for the measure SP4 are presented in Table 5.10.

$$S_i = \frac{Number\ of\ passwords\ identified\ correctly}{4}\ \%\quad \text{(Eq. 5.4)}$$

Each user (i) had to guess four passwords in each condition.

$$\mathbf{SP4(j)} = \sum_{i=1}^{70} S_i /_{\mathbf{70}} \qquad \text{(Eq. 5.5)}$$

Each condition (j) had i=70 attackers (repeated measures design).

Condition	SP4 (%)	SD	SE	Median	Range
Mikon	77.5	18.62	2.22	75	50-100
Doodle	77.5	17.10	2.04	75	50-100
Art	40.71	13.57	1.62	50	25-75
Object	81.07	17.25	2.06	75	50-100

Table 5.10: Descriptive statistics of the performance of the attackers (SP4) in GS2

SP4 was not normally distributed for each condition as assessed by *Shapiro Wilk test*. A non-parametric test for independent measure was used to test the significance because the guessability attacks were partially independent, even though each of the attackers performed in all the conditions (repeated measures). The guessability attacks were partially independent because all the subjects did not guess the same password, even though subjects in each of the five groups guessed the same passwords as shown in Table 5.9. A *Kruskal Wallis* test showed that SP4 was significantly affected by the type of images used as password [$\chi 2$ (3) = 131.47, $p < 0.001$)]. The *Mann Whitney* post hoc test did not show any significant difference between each pair of conditions, except Mikon-art, doodle-art and object-art conditions. According to the descriptive statistics presented in Table 5.10 and the significance test results, the order of decreasing guessability is: *Object* ≥ *Mikon* ≥ *Doodle* > *Art*

5.5.2 Login Percentage of Each Group

The mean login success percentages for each group (1-5) in each condition obtained from the guessability study are presented in Figure 5.7.

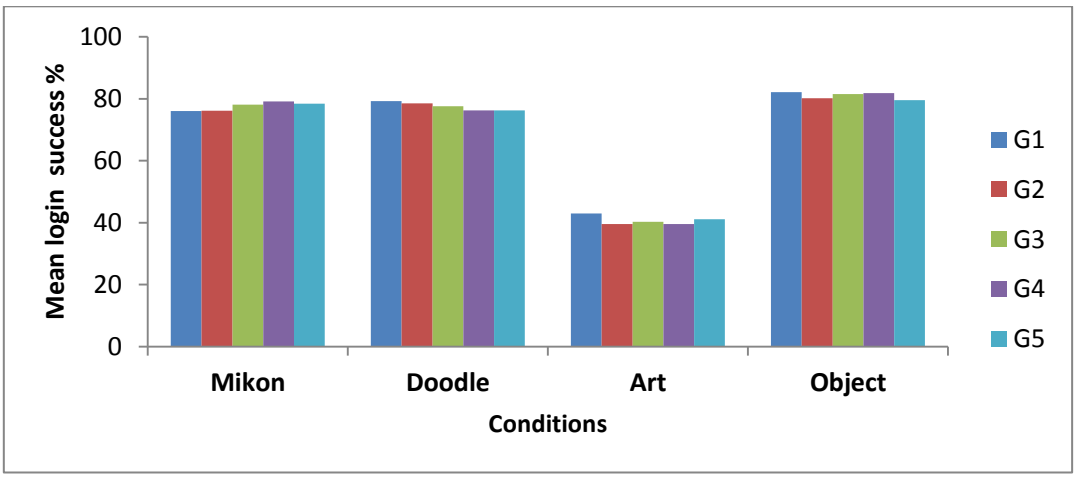


Figure 5.7: Mean login success percentages for each group in GS2

The results demonstrate that all the groups in each condition had similar login percentages, which was also found to be statistically insignificant using a *Kruskal Wallis* test. The statistics presented in Figure 5.7 and the significance test result show that the guessability of the different groups in each condition was not significantly affected, though some groups guessed different passwords using the corresponding password descriptions.

The results revealed that most groups guessed passwords to the same extent in each condition. Hence attackers did find it easier to guess Mikon, doodle and object passwords compared to the art, which is in line with the SP4 statistics presented in Table 5.10.

5.5.3 Number of Passwords Guessed

In this section, the guessability of the story passwords will be discussed in terms of the percentage of the attackers guessing one, two, three and four passwords, as well as the total number of passwords guessed during the attack. Figure 5.8 presents the box plot distribution of number of passwords guessed by the attackers in each condition.

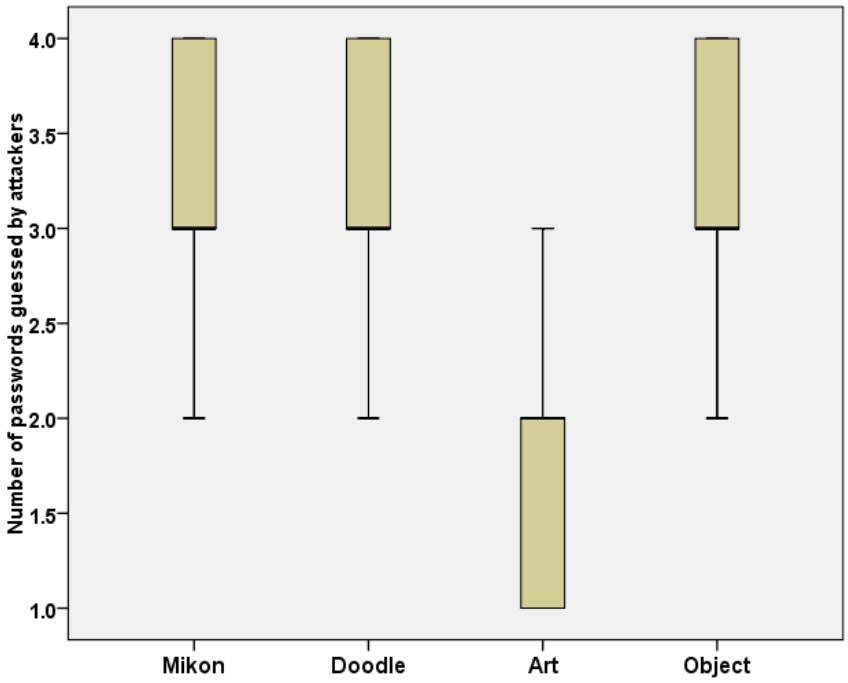


Figure 5.8: Box plots representing number of passwords guessed in each condition

- *Mikon:* 22.8% of the attackers guessed at least two passwords, 44.28% guessed at least three passwords, 32.85% guessed all the four passwords. These results

demonstrate that each of the 20 Mikon passwords (100%) used in the guessability study were guessed at least once during the guessability attack.

- *Doodle:* 18.57% of the attackers guessed at least two of the four passwords, 52.85% guessed three passwords and 28.57% were able to guess all the four passwords. The results demonstrate that each of the 10 doodle passwords (100%) was guessed at least once during the guessability attack**.**
- *Art:* 41.42% of the attackers guessed only one password, followed by 57.14% of the attackers guessing two passwords and 1.4% guessed three passwords. None of the attackers were able to guess all the four passwords; Out of 20, ten of the art passwords (50%) were guessed in the guessability study. All the passwords guessed in the study had either one or two of the target images forming the password, described as annotated sketches. The passwords which were described using words (only) were not guessed at all.
- *Object:* 14.28% of the attackers guessed two passwords, followed by 47.14% guessing three passwords and 38.57 % guessed all the four passwords. The results suggest that each of the 18 object passwords (100%) was guessed at least once, during the guessability attack.

5.6 Comparing the Results

In the following sub-sections, the results obtained from the usability (US2) and guessability (GS2) studies are compared to the existing research that has reported similar studies.

5.6.1 Comparing US2 Results

In this section, we will compare the results obtained in US2 to the similar multiple graphical password studies (reported in the literature) and US1 (user study reported in Chapter 3).

US2 – Moncur & Leplatre (2007)

According to the statistics presented in Table 2.5 (Section 2.6), the login success percentage was highest in the case of group 3, i.e. participants who were asked to use a mnemonic strategy to remember their object passwords (RT1- 92%, RT2- 14%, RT3 – 20%, Mean- 42%). The results obtained in US2 show superior performance, viz. 75.6 % mean login

success percentage in the case of object (story) passwords, compared to the mean login success of group 3 in Moncur & Leplatre (2007). The experimental protocol adopted in US2 also ensured that each of the subjects in US2 employed a mnemonic strategy to create their passwords, which could have influenced superior results compared to Moncur & Leplatre (2007).

US2 – Chiasson et al. (2009)

The results of the memorability tests conducted two weeks after the registration stage in Chiasson et al. (2009), reported that the success rate of click-based passwords was 57%, and text-based password was 70%. In case of story passwords, the success rate is highest in the case of object passwords (75.4%), followed by Mikon (63.25%), then doodle (61.85%) and lowest in the case of art passwords (44.6%). The mean registration time for click-based passwords was 43.9sec, which is lower than the story passwords, varying between 59.68 sec (object) and 86.11 sec (art). The mean authentication time for click-based password varied from 15.1sec to 47.0sec after two weeks, which is higher than the story passwords, varying between 20.35sec (object) and 26.65sec (art).

US2 – Everitt et al. (2009)

The study reported in Everitt et al. (2009) found that participants using four different system-issued face passwords each week had a failure rate of 15.23% after four weeks. In US2, the lowest failure rate is in the case of the object passwords (24.4%) and highest in the case of the art passwords (55.4%), which shows that the story passwords had an inferior performance compared to face passwords. The mean login time reported in Everitt et al. (2009) was 29.7 sec, which is more than the story passwords (20.35- 26.65 sec).

US2 – Hlywa et al. (2011)

The results reported in *Hlywa et al. (2011)*demonstrated slightly better performance, viz. mean success rate of 78.33% in study 1 and much superior performance, i.e. mean success rate of 95% in study 2 for object passwords, compared to the login success results reported in US2 (Table 5.2). In relation to the mean authentication time, the performance of the object passwords in US2 (Table 5.4) is superior compared to the statistics reported in Hlywa et al. (2011) for both the studies, as shown in Table 2.5. The login performance with face passwords in Hlywa et al. (2011) was better in study 2 (Table 2.5) compared to the results

reported in US2 (Table 5.2). The mean authentication time of face passwords in both the studies reported in Hlywa et al. (2011) were much higher compared to all the conditions in US2 (Table 5.4).

US2 - US1

The memorability results reported in Chapter 3 demonstrate that the successful login percentage after a period of eight weeks for Mikon was 74.17%; doodle - 67.04%; art - 54.9%; object - 77.3%. The same trend is observed in the results obtained in US2; however the login success percentage decreased in each condition (Table 5.2). The mean registration time (Table 5.3) and authentication time (Table 5.4) reported for story passwords follow the same order as in US1 (Tables 3.5 and 3.6). However, there is a difference in the quantitative values reported in both the studies, which is favourable towards US1. Hence, the effectiveness (memorability) results obtained from both the studies US1 and US2 provide evidence that, mnemonic strategies do not enhance the memorability of multiple RBGS passwords.

5.6.2 Guessability of Story Passwords

This chapter also reported a guessability study that was conducted with 70 participants to examine the vulnerability of story passwords in RBGSs to written descriptions. The guessability of Mikon, doodle and object passwords was 100%, whereas art passwords had a guessability of 50%. The results show that all the story passwords that have one or two target images recorded as sketches (annotated/ non-annotated) were guessed during the guessability attack. If it is assumed that the story passwords having three or four target images recorded as sketches are guessable, then the overall guessability (80 passwords created in US2) of Mikon, doodle and object passwords will be 100%, making them highly insecure to use. Seventy (out of 80) art passwords created in US2 had textual descriptions for at least three to four target images. In GS2, ten passwords which had textual descriptions for three-four target images were not guessed at all. But, these results do not provide sufficient evidence to claim that the remaining 60 passwords having textual descriptions for three to four target images cannot be guessed. Hence we maintain that in the current experimental setting, all the passwords which had one to two target images recorded as sketches were successfully guessed.

GS2 – Dunphy et al. (2008)

The results reported in Dunphy et al. (2008) found that, out of 158 authentication attempts made by 56 participants, (8%) were successful. The guessability of story passwords was much higher as discussed in section 5.5.3 and the statistics presented in Table 5.10, compared to Dunphy et al. (2008).

GS2 – GS1

In GS1 the descriptions given to the attackers took the form of words (text) only. However, in GS2 the descriptions comprised of sketches (annotated/ non-annotated) as well as words. The results presented in Tables 4.4 and 5.10 suggest that the overall performance of the attackers was almost the same, except for the Mikon passwords, in both GS1 and GS2 respectively. In the case of Mikon passwords, the guessability performance was better in GS2 (77.5%) compared to GS1 (52.14%). The high guessability of Mikon in GS2 can be attributed to the fact that most Mikon images were recorded as sketches, which made it easier to guess them. However, the results discussed in Sections 4.4.3 and 5.5.3 also suggest that all the Mikon, doodle and object passwords were guessed at least once during the attack, in both the guessability studies GS1 and GS2 respectively. In both GS1 and GS2, only 50% of the art passwords were guessed. Moreover, it was also found that the target images forming the art passwords were the least recorded as sketches in both GS1 (stage 1) and GS2.

5.6.3 Limitations of the Studies

In US2, two limitations in the field raised by Biddle et al. (2009) have been addressed: (1) the usability of multiple RBGS story passwords was examined; (2) the registration of the passwords was split into multiple sessions (i.e. different days in the same week). The limitations in the usability study (US2) were the same as in US1, which has been discussed in Chapter 3 (Section 3.7.3).

The limitation in the guessability study (GS2) in the context of the unskilled attackers is same as that of GS1, which has been discussed in Chapter 4 (Section 4.6). In the context of the descriptions provided by the subjects who took part in US2, it is not known, whether the

descriptions were influenced by the mnemonic strategy employed to create and remember the RBGS passwords. A short questionnaire study with the subjects, once they finished recording their descriptions would have helped to assess it. We believe that if the descriptions provided by the subjects are influenced by the mnemonic strategy employed to create the respective passwords, then it would make RBGS passwords highly guessable, if the descriptions are compromised. In the current experimental set-up subjects in US2 were asked to record a description (maximum 25 words) for each of their passwords. However in the real world, instead of writing a description of the target images, subjects might have recorded the mnemonic strategy used to create and remember the respective passwords. Hence, subjects in US2 should have been explicitly asked about the coping mechanism they would employ in real life to remember the multiple story passwords. This information would have helped to better understand the coping mechanism that will be used in a real life scenario. However, we believe that these limitations do not invalidate the results obtained in GS2 because the main aim was to examine, whether RBGS story passwords can be guessed by a third-party using the descriptions, which are provided by the RBGS password owners; the experimental set-up used in GS2 did investigate and found answers to the research question.

5.7. Conclusion and Recommendation

The results obtained in US2 (Table 5.2) and the comparisons made with other studies (Table 2.5 and 3.4) in Section 5.6.1, clearly suggests that mnemonic strategies do not enhance the memorability of multiple RBGS passwords. The effectiveness (Table 5.2), efficiency (Tables 5.3 and 5.4) as well as the exit questionnaire results (Section 5.3.4) confirmed that subjects found it difficult to create multiple mnemonic strategies, and use them effectively to remember their respective passwords. Moreover, using multiple mnemonic strategies caused interference in the human memory, making password authentication a mentally demanding task.

GS2 found that most subjects recorded their story passwords by drawing sketches (annotated or non-annotated) of the target images (Table 5.8), which made it easier for the attackers to guess the passwords. However, art passwords performed the best compared to the other image types, in the sense of being the least guessable and the least recorded as sketches (Table 5.10 and Section 5.5.3). We believe that the issue of RBGS password descriptions, i.e.

ability to write down and describe RBGS passwords is a fairly new area of research in the field. In the future, if RBGSs become widely adopted, then explicit selection criteria for such systems may be based upon the ability to write down, describe and share the password credentials, as well as the extent to which such credentials can be guessed, if revealed by the owner.

Based on the results of the exit questionnaire (improvements, i.e. Figure 5.6), a hint-based authentication system is proposed, which will be further discussed in the Chapter 6. Two empirical studies are also reported to evaluate the performance of the proposed system.

Chapter 6
Passhint Authentication System

In the context of the existing interest in image passwords, this chapter presents a novel authentication system, i.e. Passhint Authentication System (PHAS) to address the problem of remembering multiple such passwords. The chapter also reports, a usability and a guessability study, which was conducted with 40 subjects. The contents of this chapter have been published in the proceedings of the ACM CHI Conference on Human Factors in Computing Systems 2014.

6.1 Introduction

The studies reported in the literature (Moncur & Leplatre, 2007; Everitt et al., 2009; Hlywa et al., 2009), as well as Chapters 3 and 5 have examined the memorability of multiple RBGS passwords and found that the user's performance deteriorates over time. In the context of the existing interest in image passwords, a *Passhint Authentication System (PHAS)* is presented as a potential solution to address the problem of remembering multiple RBGS passwords.

6.1.1 Terminologies

For clarification, the definitions of aspects related to the usability and guessability study reported in this chapter which will be used frequently are given below.

- *Account holder:* A legitimate user who has registered with a PHAS password and proves the authority by using the same password.
- *Subject:* An account holder who has participated in the usability study (US3).
- *Attacker:* A participant in the guessability study (GS3) who is trying to guess passwords in PHAS.

6.1.2 Contributions

The main contributions of this chapter are:

- A novel authentication system PHAS is presented and its performance is investigated in terms of the memorability of multiple passwords, registration time and login time for successful authentication attempts;

- The superior performance of PHAS is demonstrated by comparing the results with existing studies;
- The guessability of the passwords in PHAS is also examined.

6.2 Cognitive Theories

The techniques used for aiding recall from memory, i.e. retrieving information, are called *cues or mnemonics,* which have certain characteristics (Parkin, 1993):

- Cues are not directly connected to the information that has to be learned, but have some meaning and structure, which typically varies from one individual to the other;
- Cues also form a meaningful association, between what is to be learned (new information), and what is already stored in the long term memory (LTM).

LTM can be classified as (Parkin, 1993):

- *Explicit,* where all the memories are consciously available (declarative in nature, for example, recalling statistics formula);
- *Implicit,* where the memory is unconscious and unintentional, (non-declarative in nature, for example, performing a specific task/action).

Cognitive studies have shown that it is easier to remember an event or personal experience, when it is strongly encoded in the LTM, i.e. explicit in nature (Kolb & Whishaw, 2003). Hence a cue is generally related to the explicit memory, which can be classified as (Rovee-Collier et al., 2001):

- Episodical memories of events in one's personal life;
- Flashbulb memories, recollections that are vivid, as if snapshots of moments in life;
- Context and perception dependent memories, which involve dynamic searching for the best interpretation of the available data and thus going beyond the immediately given evidence;
- State dependent memories related to the internal "state" of the user and other physiological factors, while creating the cues.

In PHAS, the images chosen by a user are stored in the LTM because the individual not only selects them, but gives a hint for each one of them. Since users would have to focus their

attention to choose the images and create hints, this should help to process as well as encode the images in the memory. The hints will act as cues while recognizing the images in the future, and are likely to enhance the memorability. A usability study (US3) is reported in Section 6.4, which will examine the effectiveness of cues in PHAS.

According to the discussion on guessability of images presented in Chapter 2 (Section 2.1.3), an image can be interpreted in different ways by each viewer. An image can be guessed easily in PHAS, if the hint given by a user denotatively describes the elements in it. But if the hint is connotative, where the user relates it to something personal (such as an episode in life), a sign or state (how it makes them feel), a context (an idea or event that only has relevance to them), then it might be difficult for an attacker to guess, without being aware of the relation between the hint and the image. A guessability study (GS3) is reported in Section 6.5 which will examine the vulnerability of hints in the case of PHAS passwords.

6.3 Design of PHAS

6.3.1 Registration

A user must select four images from the given collection presented by the system, and then provide a hint for each image to complete the registration. A collection comprising of 150 images is presented on the screen, as six sets of 25 images in the form of 5×5 grids. The collection for each image type (Mikon, doodle, art and object) is drawn from the archives having 450 distinct images of the respective image type, which were used in US1 (Section 3.2). A user can browse from one set to another using a 'change set' button on the web page. A user can choose all the four images from a single set or each image from a different set. The hints can be in any language (maximum limit of six words), but must be typed in English characters. The four target images together with their corresponding hint form a PHAS password. Figure 6.1 shows an art password in PHAS.

Figure 6.1: Sample art password in PHAS

6.3.2 Authentication

Authentication is a four step process. At each step, a challenge set consisting of 15 decoy images and 1 target image is displayed in a 4×4 grid, together with the associated hint (Figure 6.2). The user has to recognize and select the target image with the help of the hint at each step.

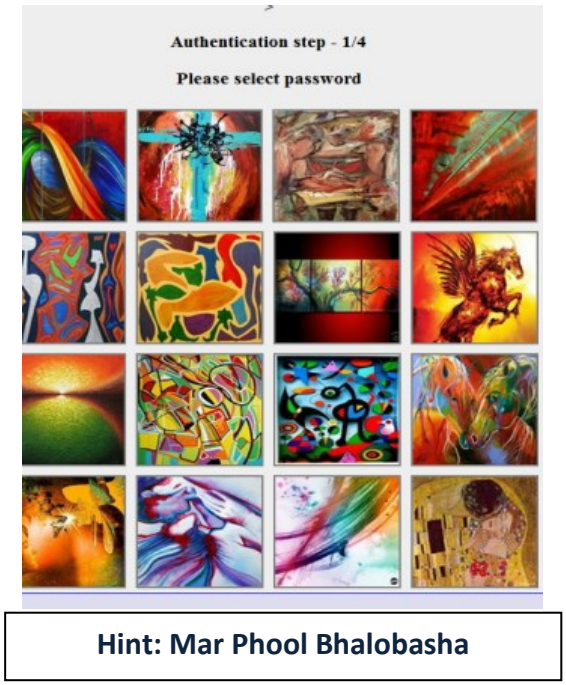


Figure 6.2: A challenge set screen in PHAS

The challenge set configuration developed for PHAS is as follows:

- The 15 decoy images for each of the four authentication steps will be distinct and never comprise of one of the target images ;
- The challenge sets for a user will not change when the web page is reloaded using the refresh button of the browser. This protects against an intruder guessing the target image, either by refreshing the web page or logging into the system continuously;
- If a wrong image is selected at any step, the target images won't be displayed in the subsequent challenge sets (steps). In such a scenario, 16 decoy images (without the target image), different from the original challenge set will be displayed, with the original hint;
- The result of the login will only be displayed, once the last step is completed. A lockout policy is implemented for four failed login attempts.

PHAS is different from the story passwords reported in Davis at al. (2004) as well as the passwords used in Chapter 5 of this thesis. In PHAS, users give a hint for each target image and can use any strategy to do so. They do not need to create a story or a pattern. The users do not need to either remember or reproduce the hints at any stage because these are stored in the system, and displayed with the challenge set to enhance the memorability.

In order to conduct the usability and guessability studies, four PHASs (prototypes) were developed. In PHAS, each password comprised of four images of the same image type, together with the corresponding hints. Each prototype used a different image type (Mikon, doodle, art and object) as the password. The image collection used in each prototype comprised of 150 images of a specific image type, which were drawn from the collections used in US1. For example, Mikon prototype had a collection comprising of 150 Mikon images which were used to: (1) create the passwords during the registration; (2) generate the challenge sets for each target image.

6.4 Usability Study (US3)

A usability study was conducted with 40 subjects (Female: 22; Male: 18; age range: 19-24 years), who were undergraduate students studying various degree courses such as engineering, management, arts and humanities. In order to recruit the participants (subjects) for this usability study (US3), emails were sent to the student email lists in a university. All the subjects were regular internet users, but none were experts in either computer usability or security. They had different nationalities *(N): N1-15; N2-10; N3-4; N4-3; N5-4; N6-4.* Ethics approval *(Ethics ref no CSE 01199)* was granted by the college ethics committee to conduct both the usability and guessability studies.

6.4.1 Usability Study Experiment Design

This lab-based study investigated the *usability* (effectiveness in terms of the memorability and efficiency in terms of the password registration and authentication time) of multiple passwords in PHAS. An experimental framework similar to the one reported by Moncur & Leplatre (2007) and Chiasson et al. (2009) was used to conduct US3. The experimental framework used to conduct US3 is discussed below.

Stage 1(Day 1): Each subject was asked to register with four passwords, one of each image type (Mikon, doodle, art and object). Each password consisted of four target images and the associated hints. Each subject was given an instruction sheet that explained both the registration and login processes. They were told that their passwords were for four specific accounts: *banking; online shopping; personal email and social networking*. This ensured that participants had a context to use, in differentiating their multiple passwords. The subjects were not given any information or suggestion as to the strategy to be adopted for creating the passwords and corresponding hints. The instructions regarding the hints were:

- They should be one or more words long, in any language, but must be typed in English characters (maximum length for each hint being six words). The restriction imposed on the length of the hints ensured that these are not too long;
- Each hint should be something which will help a legitimate user to recognize the target images at a later date and, ideally, not be useful to anyone else trying to guess the target images.

A distraction task was given to the subjects after registering with each password. These were also used in the study reported in Chiasson et al. (2009) and are intended to clear the working textual memory and verbal memory, when multiple passwords are created simultaneously in a single session. The distraction tasks in US3 included listening to songs, watching funny videos, solving word puzzles and answering a quiz about the University. Each distraction task lasted for about 8-10 minutes.

Each subject had to authenticate three times, with each of their passwords. The system displayed the password in the case of three failed authentication attempts. Finally, the subjects were asked to categorize the hints they had used for each of the target images into one of the selected types of explicit memory (episodic, flashbulb etc.) discussed in Section 6.4.2. The login success was not analysed for this stage.

Stage 2- Retention test (14 days after Stage 1): Each subject was asked to authenticate three times using the hints. There were no practice session in between stage 1 and 2. This experimental design helped to examine the usability of multiple PHAS passwords, when they have not been used for a considerable period of time.

6.4.2 Usability Study Results

The independent variables in US3 are the four different image types Mikon; doodle; art; object. The dependent variables and the corresponding results are discussed below.

(A) Effectiveness

The mean login success percentage (SP5) of each subject in each condition (image type) is calculated using Eq. 6.1.

$$SP5 = Successful\ attempts / Total\ attempts \quad \text{(Eq. 6.1)}$$

Table 6.1 presents the descriptive statistics for the measure SP5.

Conditions	**SP5%**	**SE**	**SD**
Mikon	95	3.49	22.07
Doodle	95	3.49	22.07
Art	97.5	2.5	15.81
Object	97.5	2.5	15.81

Table 6.1: Descriptive statistics for mean login success percentage in US3

SP5 for each of the conditions was not normally distributed as assessed by the *Shapiro-Wilk* test. A *Friedman* test showed that there is no significant difference between the conditions (χ^2 = 0.667, df =3, p=0.88). *Wilcoxon* post hoc test did not show any significant difference between each pair of conditions. Out of 160 passwords only 6 (3.75%), were not memorable. This demonstrates the effectiveness of PHAS, when the users have to remember multiple passwords.

(B) Efficiency

The mean registration time (RegT3) and mean authentication time (AuT3) was calculated to assess the efficiency of PHAS.

Registration time: The descriptive statistics for RegT3 are given in Table 6.2. The box plots for the RegT3 distribution is shown in Figure 6.3. The RegT3 in each of the conditions was normally distributed as assessed by the *Shapiro-Wilk test*. A *Repeated Measure ANOVA* was chosen to examine the statistical significance. Since the estimate of the sphericity (0.78)

was greater than 0.75, a Huynh-Feldt correction was used. The result showed no significant difference between the RegT in each of the conditions (F= 1.372, p=0.258, df = 3). The post hoc comparisons also demonstrated no significant differences between all pairs of conditions. According to the descriptive statistics presented in Table 6.2, the decreasing order of RegT3 is*: Doodle* ≥ *Mikon* ≥ *Art* ≥ *Object,* but this is not statistically significant.

Conditions	**RegT3 (seconds)**	**SE**	**SD**
Mikon	57.78	1.42	9.01
Doodle	58.03	0.87	5.14
Art	56.88	0.86	5.46
Object	55.63	0.86	5.47

Table 6.2: Descriptive statistics for mean registration time in US3

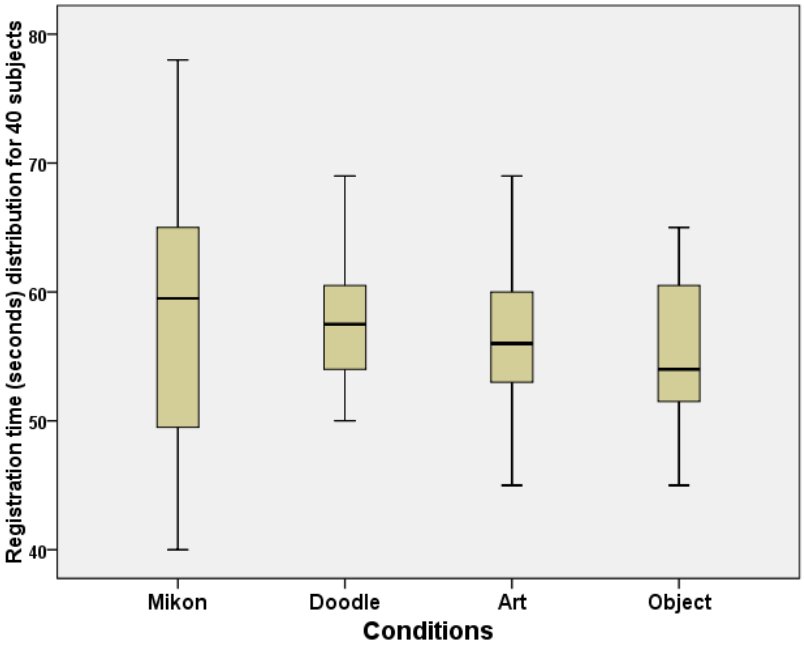


Figure 6.3: Box plot showing the registration time distribution in US3

Authentication time: The mean authentication time (AuT3) of the passwords for the successful authentication attempts in each condition is reported in Table 6.3. The box plots for the authentication time distribution in each condition are shown in Figure 6.4.

Conditions	AuT3 (seconds)	SE	SD
Mikon	15.88	0.57	3.65
Doodle	17.15	0.44	2.80
Art	13	0.35	2.25
Object	13.57	0.33	2.09

Table 6.3: Descriptive statistics for mean authentication time in US3

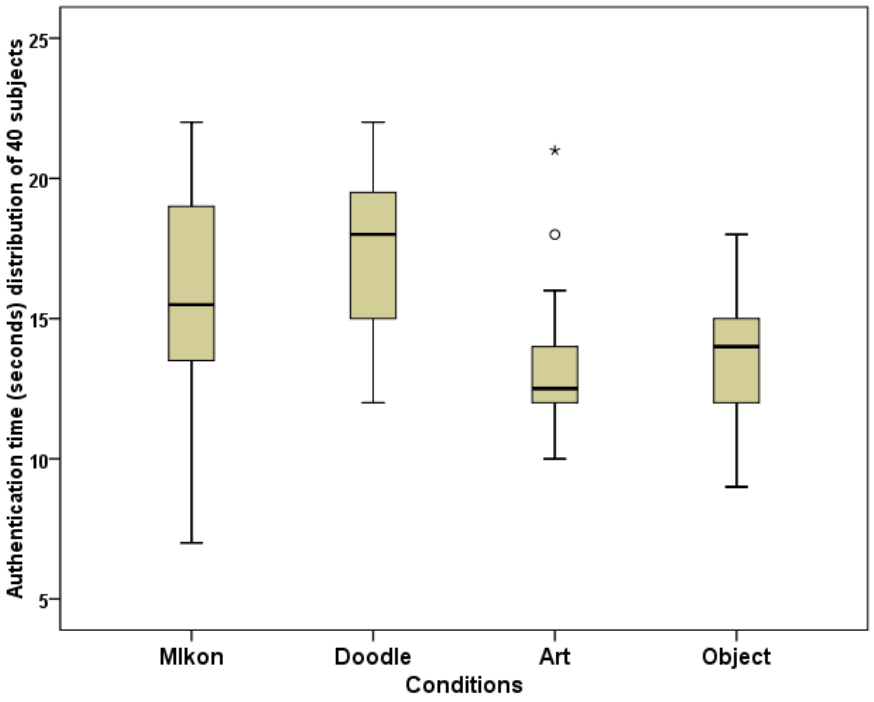


Figure 6.4: Box plots showing distribution of the authentication time in US3

AuT3 for each of the conditions was not normally distributed as assessed by the *Shapiro-Wilk* test. A *Friedman* test showed that there is no significant difference between the conditions (χ^2 = 37.36, df =3, p<0.001). *Wilcoxon* post hoc test showed significant difference between all pairs of conditions, except Doodle-Mikon and Object-Art. According to the descriptive statistics presented in Table 6.3 and the significance tests, the decreasing order of AuT3 is: *Doodle* ≥ *Mikon* > *Object* ≥ *Art.* The box plots show that the art passwords have two outliers, i.e. two subjects taking longer time to authenticate than the majority of the sample population. In both cases (outliers), subjects were not able to find the association between the hint and some of the target images forming the password. This might be due to nature of the

hints provided by these subjects. Since, we did not ask the subjects the reason for the delay in authenticating; this aspect is not discussed further, in the context of the art passwords in PHAS.

(C) Categories of Hints

All the subjects were asked to categorize each of their hints. The categories were: episodic memory; flashbulb memory; sign/context; descriptive knowledge; randomly chosen. The details and explanations for all the categories were provided to the subjects, which were similar to the ones described in the cognitive theory (Section 6.2). Figure 6.5 shows the responses given by 40 subjects for each hint created by them.

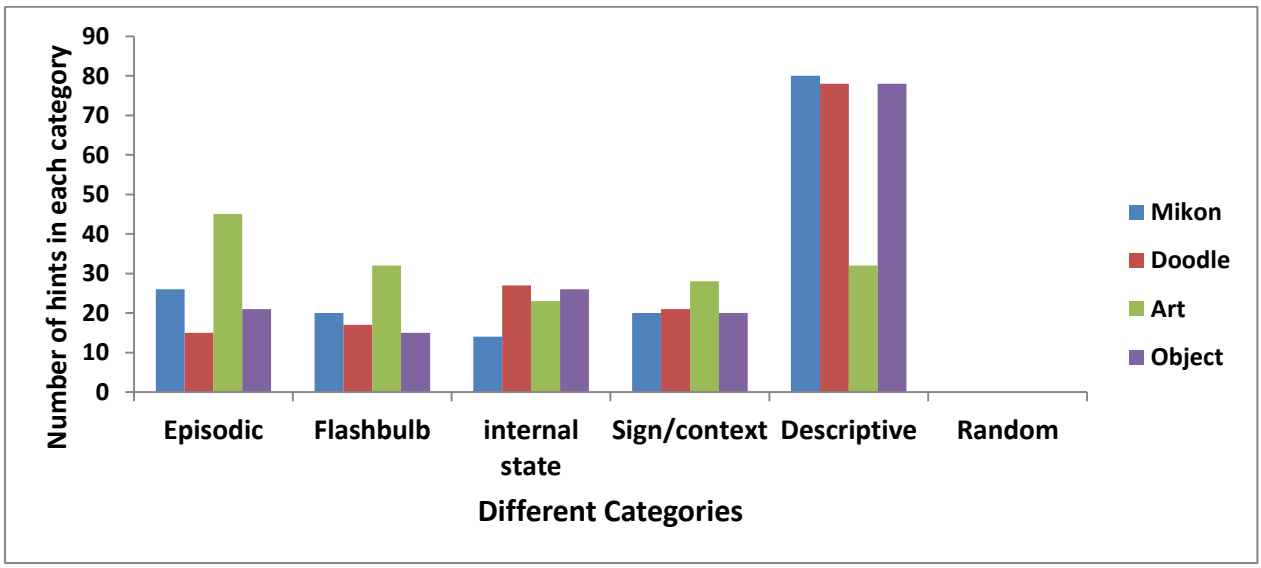


Figure 6.5: Responses given by the subjects in US3 in context to hint categorization

Please note that each password in PHAS is composed of four target images and each image has a hint. The number of images having descriptive hints in the case of art passwords was (32/160), which is lower than compared to Mikon (80/160), doodle (78/160) and object (78/160). Further analysis revealed that the number of passwords having descriptive hints for all the target images is also considerably lower, in the case of art passwords than the others, as shown in Table 6.4. Descriptive hints are interesting in the sense that they can not only enhance memorability, but can also help an attacker to guess the target images. In this context, the results reported in Chapter 4 (Table 4.9) show that RBGS passwords can be effectively guessed using denotative descriptions of the target images.

	Passwords having descriptive hints for			
	4 T images	**3 T images**	**2 T images**	**1 T image**
Mikon	8	8	7	10
Doodle	7	8	8	10
Art	1	3	5	9
Object	8	8	8	6

Table 6.4: Passwords having descriptive hints, T denotes target images

Figure 6.6 (below) presents a number of art images together with their hint and the corresponding hint category.

(i)
Hint: Table
Category: Descriptive

(ii)
Hint: Serene
Category: Flashbulb

(iii)
Hint: Interiors
Category: internal state

(iv)
Hint: Rajasthani
Category: sign

(v)
Hint: Sistine
Category: Episodic

(vi)
Hint: crying
Category: context

Figure 6.6 (i to vi): Art images in PHAS with hints and respective categories

6.5 Guessability study (GS3)

A lab-based study was also conducted to examine the extent to which passwords in PHAS can be guessed using the corresponding hints (the additional component used to enhance the memorability). The study was conducted on the same day as the retention test (Stage 2) of the usability study.

6.5.1 Guessability Study design

The usability study was conducted with subjects from different nationalities, who gave hints in different languages. In the usability study, 160 passwords (160×4=640 target images) and their corresponding hints were created. Almost, 90% of hints for the target images were either in the mother tongue or the national language of the subject, creating the hints. Therefore, it was necessary to have two treatments for each password:

- *Treatment 1 (T1)* was a guessing attack by an attacker with the same nationality as that of the subject;
- *Treatment 2 (T2) was* a guessing attack by an attacker whose nationality differed.

Each attacker was asked to guess 4 Mikon, 4 doodle, 4 art and 4 object passwords in PHAS. Each attacker was given:

- a task information sheet;
- sixteen usernames to log into the system;
- nationalities of the account holders;
- access to all online resources like translators, web images etc.

Each attacker had four chances to guess a password. The attackers were not put under any time limit for the task. Figure 6.7 shows the distribution of the guessing trials (number of login attempts). For example, Nationality 1 (N1) had 15 subjects. Hence, the collection obtained from US3 had 15 passwords of each image type for the nationality N1.

- In T1, each of the 15 passwords was guessed by at least 2 attackers from N1, making it 30 trials for each image type. The total number of trials in T1 was *40 passwords × 2 subjects = 80 (for each image type)*.

- In the case of T2 (N1), 6 passwords were attacked by only one attacker; whereas 9 passwords were attacked by two attackers, having a different nationality (N2-N6).

In T2, passwords were attacked by one, two or three attackers, unlike T1 where each password was attacked by exactly two attackers. The difference in the number of subjects among all the nationalities, and the experimental protocol that each attacker must attack four passwords of each image type, made it difficult to have equal number of trials for each password in T2.

Treatment	Nationality	Total passwords	Each password guessed by
T1	N1	15	2 subjects
T2	N1	9	2 subjects
T2	N1	6	1 subjects
T1	N2	10	2 subjects
T2	N2	10	2 subjects
T1	N3	4	2 subjects
T2	N3	3	3 subjects
T2	N3	1	2 subjects
T1	N4	3	2 subjects
T2	N4	2	3 subjects
T2	N4	1	2 subjects
T1	N5 & N6	4	2 subjects
T2	N5 & N6	4	2 subjects

Each attacker guessed 4 passwords for each image type

In T1 each password was guessed by at least 2 attackers ***(from the same nationality).***

In T2 each password was guessed by one, two or three attackers ***(from a different nationality).***

Figure 6.7: Distribution of guessing trials

6.5.2 Guessability Study Results

(A) Performance of the attackers from the same nationality (T1)

The mean login success percentage (SP6) of each password in each condition (image type), for T1 is calculated using Eq. 6.2. The descriptive statistics are presented in Table 6.5.

$$\sum_{i=1}^{40} \frac{P_i\left(^{successful\ attempts}/_{total\ attempts}\right)}{40}$$, *P_i represents subjects, i=1 to 40* (Eq. 6.2)

Conditions	SP6 (%)	SE	SD
Mikon	6.25	3.19	20.21
Doodle	11.25	4.19	26.52
Art	2.50	1.74	11.03
Object	16.25	4.86	30.77

Table 6.5: Descriptive statistics showing the performance of the attackers in treatment T1

	Password guessability percentage in T1		
Conditions	**0%**	**50%**	**100%**
Mikon	36	3	1
Doodle	33	5	2
Art	38	2	0
Object	32	6	2

Table 6.6: Password guessability distribution in treatment T1

Table 6.6, analyses the guessability of passwords in each condition for T1. In T1, two attackers attempt to guess each password.

- 0% guessability denotes that the password has not been guessed by any attacker;
- 50% means that the password was guessed by only one of the two attackers;
- 100% means that the password was guessed by both the attackers.

For example, in the case of Mikon, 36 passwords were not guessed, 3 were guessed by one of the two attackers and one was guessed by both the attackers.

SP6 was not normally distributed for each condition as assessed by the *Shapiro Wilk test*. A non-parametric independent measure test was used to examine the significance. A *Kruskal Wallis test* showed that there is no significant difference between the conditions (p=0.055). A *Mann-Whitney* post hoc test (*Bonferroni correction*, at a 0.008 level of significance) did not show any significant differences too. The decreasing order of guessability, for the measure SP6 is: $Object \geq Doodle \geq Mikon \geq Art$

(B) Performance of the attackers from different nationality (T2)

The mean login success percentage (SP7) of each password in each condition is calculated using Eq. 6.2. The descriptive statistics for SP7 is presented in Table 6.7.

Conditions	**SP7 (%)**	**SE**	**SD**
Mikon	10.83	4.48	28.38
Doodle	10.00	4.84	30.82
Art	1.25	1.74	11.03
Object	17.91	5.00	31.06

Table 6.7: Descriptive statistics for the performance of the attackers in treatment T2

Table 6.8 analyses the guessability of the passwords in each condition for T2. In T2, each password was attacked by a minimum of one and a maximum of three attackers.

- 0% guessability denotes that the password has not been guessed by any attacker;
- 33.33% denotes that password was guessed by one of the three attackers.
- 50% means that the password was guessed by only one of the two attackers.
- 66.66% means that password was guessed by two of the three attackers.
- 100% means that password was guessed by both or all the three attackers.

For example, in the case of Mikon 34 passwords were not guessed, one was guessed by 1 out of 3 attackers, one was guessed by 1 out of 2 attackers and four were guessed by all the attackers.

	Password guessability percentage in T2				
Conditions	**0%**	**33.33%**	**50%**	**66.66%**	**100%**
Mikon	34	1	1	0	4
Doodle	36	0	0	0	4
Art	39	0	1	0	0
Object	29	0	7	1	3

Table 6.8: Password guessability distribution in treatment T2

SP7 was not normally distributed for each of the image types as assessed by a *Shapiro Wilk test*. A *Kruskal Wallis* test showed significant differences between all the conditions (p=0.015). The *post hoc* test did not show any significant difference between each pair of conditions (p>0.008), except the art-object (p=0.002). The decreasing order of the guessability is same as that of SP6: *Object* ≥ *Mikon* ≥ *Doodle* > *Art.*

(C) Comparing T1 and T2

A *Wilcoxon signed rank* test showed that the variation between SP6 and SP7 in each condition is statistically insignificant. This is also evident from Table 6.9, which shows that the overall guessability is almost similar for both T1 and T2, in each condition. Hence we conclude that the overall guessability of the image types in PHAS is *Object* > *Doodle* ≥ *Mikon* > *Art*.

	T1	**T2**
Total Password guessed	21/160	22/160
Mikon	4/40	6/40
Doodle	7/40	4/40
Art	2/40	1/40
Object	8/40	11/40

Table 6.9: Comparing T1 and T2

In Figure 6.8, the passwords which were guessed during the guessability attack are further analysed. The horizontal axis represents the image types and the vertical axis represents the password numbers. For example, password number 28 in all the conditions belonged to one and the same subject (account holder). The meaning of different symbols is as follows.

- ***Diamond:*** all the passwords of an account holder were successfully guessed.
- ***Circle:*** three passwords of an account holder were successfully guessed.
- ***No shape:*** only one password of an account holder was successfully guessed.
- ***Red numerals:*** passwords successfully guessed both in T1 and T2.
- ***Black numerals:*** passwords guessed in T2 but not in T1.
- ***Green numerals*:** passwords guessed in T1 but not in T2.

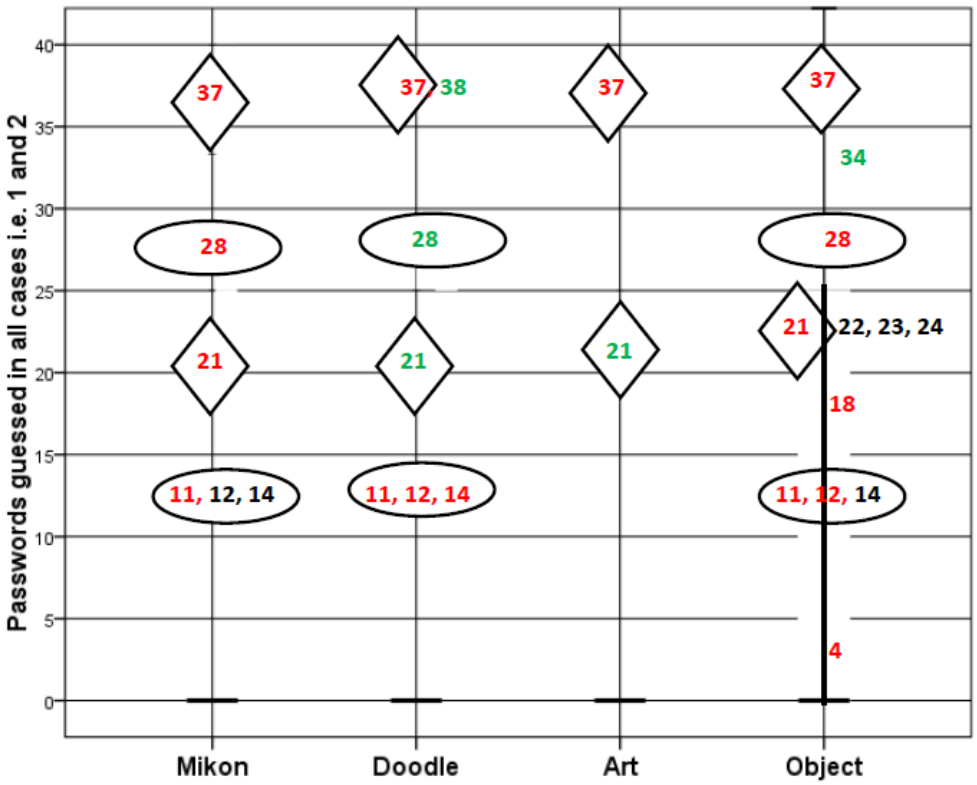


Figure 6.8: Analysing password guessability in PHAS

In the case of Mikon all the four passwords guessed in T1 were also guessed in T2 (4/6). 4/7 doodle passwords guessed in T1 were the same as those in T2 (4/4). In the case of art passwords, one out of 2 passwords guessed in T1 was the same as that in T2. 7/11 object passwords guessed in T2 were same as 7/8 passwords guessed in T1. Almost 50% of the passwords that were guessed in T1 were the same as the ones guessed in T2.

6.6 Discussion

The performance of PHAS passwords in US3 and GS3 is compared to the existing studies that have reported similar experiments.

6.6.1 Comparing the Performance of PHAS with Multiple Password Studies

This section will compare the results of the PHAS usability study (US3) to the existing studies that have explored the cognitive demands of using multiple graphical passwords (including US1 nnd US2). Table 2.5 presented in Section 2.6, provides a summary of the results reported in the literature that have explored the usability of multiple graphical

passwords. The login success rate immediately after creating the passwords (training stage) is not considered in the comparisons, as they tend to make the mean values higher. The results obtained in US3 are also compared to the single password studies discussed in Chapter 2 (Table 2.3).

US3 – Moncur & Leplatre (2007)

According to the statistics reported in Moncur & Leplatre (2007)which has been also presented in Table 2.5, the login success rate (RT1) for group 0 dropped to almost 5% after 2 weeks, group 1 was around 11%, group 2 was less than 10%, group 3 was around 14% and group 4 was almost 12%. These are considerably less than the login success rate of PHAS (96.25%), after the same period of time. The passwords in Moncur & Leplatre (2007) were issued by the system, so the registration time was not reported. The login time was not reported.

US3 – Chiasson et al. (2009)

In Chiasson et al. (2009), 57% (15/26) of the participants were able to recall their click-based passwords successfully, two weeks after the registration. In the case of PHAS, 85% of the participants (34/40) could remember all their four passwords, after two weeks; the remaining 15% (6/40) failed to recall just one of their four PHAS passwords. The mean registration time for six click-based passwords was 43.9 seconds (approximately), which is lower than PHAS passwords, viz. 57.03 sec. The mean authentication time for click-based passwords varied between 15.1 sec and 47.0 sec. The mean authentication time of four PHAS passwords was 14.9 sec (lowest in the case of art images, i.e. 13 sec), which is lower compared to the click-based passwords, after the same period of time.

US3 – Everitt at al. (2009)

Everitt at al. (2009) demonstrated that participants accessing four different face passwords each week had a failure rate of 15.23% after a month, when each password was used once a week. PHAS had a failure rate of 3.75% after two weeks, without any practice in the preceding weeks, which demonstrates better performance compared to Everitt at al. (2009). In Everitt at al. (2009) , the passwords were issued by the system, so the registration time was not reported. The mean authentication time of PHAS passwords was 14.9 sec (best in the case of art images, i.e. 13 sec), which is better than the time reported in (Everitt et al., 2009), i.e. 24.27 sec.

US3 – Hlywa et al. (2011)

PHAS passwords demonstrated superior memorability performance (login success – 96.25%), compared to the performance of each image type used in Hlywa et al. (2011), as shown in Table 2.5 in Chapter 2. Hlywa et al. (2011), reported the best success rates (S1: 78.33% and S2: 95%) for the object passwords. The mean authentication time of PHAS passwords (57.03 sec) is much better than the passwords used in Hlywa et al. (2011), as shown in Table 2.5 in Chapter 2. In this context, it should be taken into account that the number of challenge sets used in Hlywa et al. (2011) for each authentication session is more than that of PHAS, which might have influenced the login time in favour of PHAS.

US3 – US1 and US2

In both the studies, i.e. US1 and US2, multiple (i.e. four) object passwords were the most memorable compared to the other image types, as shown in Table 3.4 (mean login success 77.31%) and Table 5.2 (mean login success 75.4%) respectively. However, the memorability of multiple (i.e. four) PHAS passwords (mean login success 96.25%) is more than the statistics reported in both US1 and US2. The best mean registration time was reported in the case of object passwords (70.61 sec in US1 and 59.68 sec in US2), which is more than each of the image types used in PHAS, as shown in Table 6.2. The best authentication time was also recorded in case of object passwords (18.28 sec in US1 and 20.35 sec in US2), which is also more than each of the image types used in PHAS, as shown in Table 6.3. Hence the results obtained in US3 demonstrated superior performance for PHAS passwords, compared to the RBGS passwords examined in both US1 and US2.

US3 – RBGS single password studies (Table 2.3, Chapter 2)

The mean login success rate (96.25%) of multiple PHAS passwords obtained from US3 is either similar, or better than the statistics reported in the existing studies in the literature that have explored the usability of a single RBGS password (login success varies between 72% and 100% as shown in Table 2.3 in Chapter 2). The registration time of PHAS passwords (Table 6.2) is more than the statistics reported in the existing single password studies. However, the authentication time of PHAS passwords for each of the image types (Table 6.3) is either similar or less, compared to the statistics reported in the existing RBGS password studies (Table 2.3).

6.6.2 Guessability of PHAS Passwords

In the guessability study (GS3), out of 160 passwords, only 27 were guessed. The results showed that *six Mikon, seven doodle, two art and twelve object passwords* were guessed. According to Table 6.4, subjects gave the most number of descriptive hints in the case of object passwords, followed by doodle and Mikon passwords and the least for art passwords. Hence the results obtained from GS3 shows that descriptive hints would increase the guessability of passwords in PHAS, which would make the passwords insecure to use. These findings complement the results obtained in GS1 (Chapter 4 – Tables 4.8 and 4.9), which showed that Mikon, doodle, art and object passwords can be easily guessed using descriptive descriptions. Overall in GS3, two subjects had each of their four passwords successfully guessed (Figure 6.8). The Mikon, doodle and art passwords of four subjects were guessed successfully. There were seven subjects whose only one password was guessed successfully.

	Art passwords having hints for			
Nature of hints	**4 images**	**3images**	**2 images**	**1 image**
Episodic	9	3	0	0
Flashbulb	8	0	0	0
state	1	3	5	0
Sign	6	1	0	1
Descriptive	1	3	5	9

Table 6.10: Categorization of hints in art passwords (PHAS)

The guessability study demonstrated that art passwords were the least guessable (only two passwords were guessed) in PHAS. Further analysis revealed that the two art passwords that were guessed had descriptive hints. It was also found that these art passwords belonged to subjects (21 and 37, diamond shape in Figure 6.8) whose all passwords were guessed, because they gave descriptive hints for all their passwords. Art images were the least guessable because most of the hints had connotative meanings that were either derived from

the episodic memory, flashbulb memory or represented a sign, internal state of the subjects, when creating the hints, as shown in Table 6.10.

6.6.3 Limitations of US3 and GS3

We acknowledge that the usability study (US3) reported in the chapter may not mirror an ideal real-life usage. It is very unlikely that users will create four passwords in succession and be asked to recall all of them after two weeks, without using them during the two weeks gap. However, we believe that this rigorous experimental framework, i.e. evaluating the memorability of multiple passwords without any practise sessions between the registration and retention stages, is an useful way to examine the effectiveness of PHAS. The results obtained from such a study can provide proof of principle for conducting field studies with PHAS in the future In US3 the creation of PHAS passwords was not spread across multiple sessions, which could be done in the future, to study the effect of retroactive interference (Baddeley, 1997). It may also be a good practice to make the subjects familiar with the authentication system beforehand, so that their behaviour is more natural and novelty effects might be avoided.

One deficiency with the kind of guessability experiment (GS3) reported in this chapter is that it relies on unskilled participants, i.e. they are not necessarily representative of the people who would be trying to break the system. The aim of the guessability study was to examine, whether hints (an additional component used to enhance memorability) will help make the target images guessable, in the first instance. Hence we maintain that the result obtained from the guessability experiment provides proof of principle for similar large scale studies in the future. Another limitation of GS3 is all the attackers did not guess the same password and there were considerable variation in the hints provided by the subjects (US3). Hence an attacker could have been hindered to guess a password correctly by being assigned one whose hints were not useful. Hence in this scenario the attacker is almost reduced to a random guess. A post-study questionnaire investigating the ease of guessing the target images using their respective hints, reason for unsuccessful attacks and approach used for guessing the target images would further help to better understand the phenomena of guessability in PHAS.

6.7 Conclusion

Multiple passwords in PHAS are easy to remember compared to existing GASs, which is confirmed from the results presented in Section 6.4.1 (percentage of successful login attempts, i.e. SP5), under a high cognitive load condition (i.e. to remember four passwords, after a period of two weeks). The hints for each target image created by the account holders simplify the recognition of the images, by making the retrieval process effective as well as cohesive. This is achieved by adding a context in which the images are encoded in the memory. Hints impose a meaning and structure by associating the images with what is already stored in the long term (i.e. explicit) memory.

The discussion presented in Section 6.6.1 provides evidence that the authentication time of PHAS passwords is also lower compared to the results reported in existing studies with other GASs. The only deficiency of PHAS is that the registration time is high compared to the statistics reported in the other studies. The registration time is high because the users not only select the target images, but type hints (which act as cues) for each one of them, which make the passwords memorable. Hence there is a trade-off between the effectiveness and efficiency. But, this trade-off may be acceptable since PHAS is able to solve the most critical issue, i.e. the memorability problem in remembering multiple passwords.

The statistics reported in the guessability study and the discussion presented in the section 6.6.2 provides substantial evidence that descriptive hints will make a PHAS password guessable. In this context, art passwords were found to be the least guessable compared to the other image types. Out of 40 only two art passwords were guessed because the hints for these passwords (two guessed art passwords) were descriptive (described elements in the image).

The chapter provides proof of principle that PHAS using art passwords could solve the problem of remembering multiple passwords and also offer an acceptable level of security to the users. Further scope of research with PHAS has been discussed in Chapter 7 (Section 7.3).

Chapter 7
Conclusions and Future Work

This chapter concludes the thesis by revisiting the thesis statement presented in Chapter 1 (Section 1.4), followed by our final position regarding the claims made in the thesis statement, and possible future work.

7.1 Thesis Statement Revisited

Multiple image passwords are memorable, and cannot be guessed using a description of the target images forming the password, given the current state-of-the-art in recognition-based graphical authentication systems (RBGSs).

The thesis statement has two components: (1) memorability of multiple RBGS passwords; (2) guessability of RBGS passwords using descriptions. Each of these components will be discussed in Sections 7.1.1 and 7.1.2, and is further illustrated in Figure 7.1.

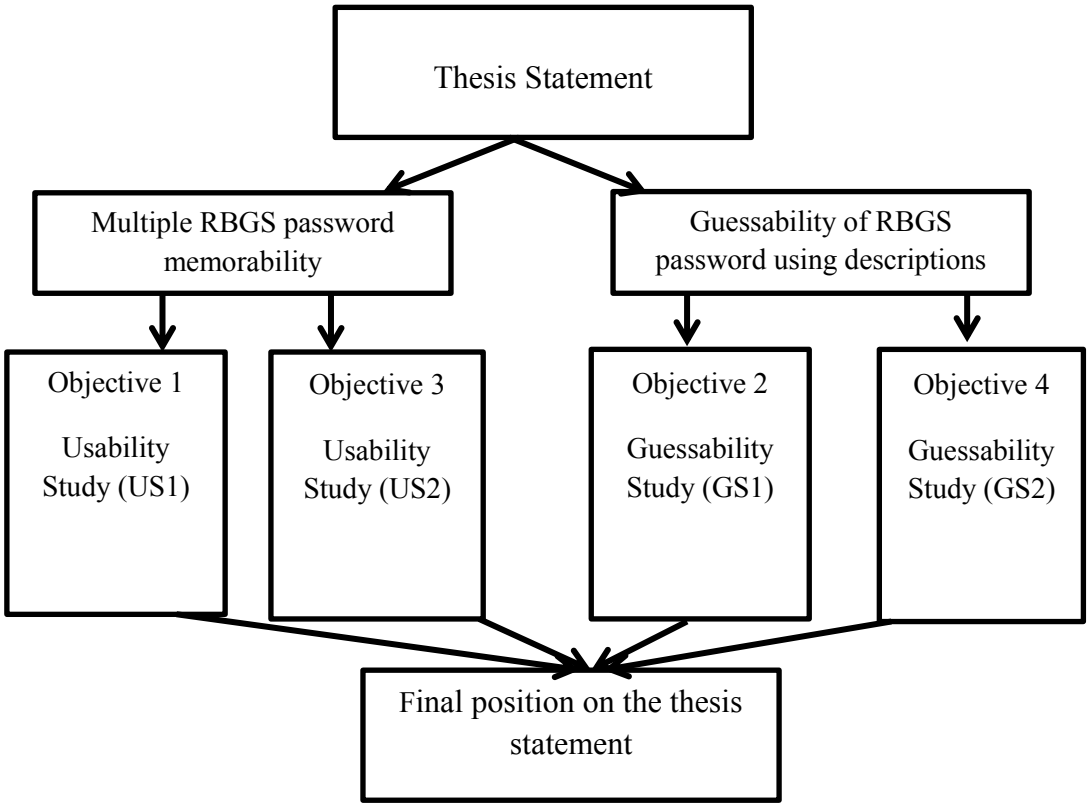


Figure 7.1: Research approach to examine the thesis statement

Study	Mean Login Success %				Registration time (sec)	Login time (sec)
(Moncur & Leplatre, 2007) 5 system assigned object passwords	***Groups***	***RT1***	***RT2***	***RT3***	***Mean***	Not reported as passwords were issued by the system	Not reported
	(0) PINs	14	5	5	8		
	(1) OP (object password)	54	11	12	26		
	(2) OP + background color (bg)	66	10	14	29		
	(3) OP + mnemonic	92	14	20	42		
	(4) OP + bg + mnemonic	82	12	18	37		
(Chiasson et al., 2009) 6 user selected click based password or text password	***Conditions***	***Recall1***	***Recall2***	***Mean***		43.9 MCP	15.1 in recall 1 to 47.0 in recall 2
	MTP- Multiple Text password	68	70	69			
	MCP- Multiple Click password	95	57	76			
(Everitt et al., 2009) system assigned face passwords	***Conditions***	***After five weeks***		***After four***		Not reported as passwords were issued by the system	***Login time (sec)***
	(4) 4 passwords, all used once in each week	84.77		85.71			24.27 – 31.71
	(5) 4 passwords, a distinct password used every week 4 times	97.5 (Week 5 data only)		0			26.88 – 28.22
(Hlywa et al., 2011) 2 or 3 system assigned passwords	***Conditions/ Image types***	***Study 1 (S1)***		***Study 2 (S2)***		Not reported as passwords were issued by the system	***S1***	***S2***
	Objects	78.33		95			31.03	22.55
	Faces	63.33		87			41.45	35.96
	Houses	38.33		Not used (NU)			83.06	NU
US1 RBGS passwords (4 passwords of a single image type)	***Conditions/ Image types***	***SP1 % (mean % for 7 weeks)***				***RegT1 (sec)***	***AuT1 (sec)***	
	Mikon	74.17				72.18	19.52	
	Doodle	67.04				75.42	22.16	
	Art	54.90				84.22	24.56	
	Objects	77.31				70.61	18.28	
US2 RBGS story passwords (4 passwords of a single image type)	***Conditions/ Image types***	***SP3 % (mean% for 4 weeks)***				***RegT2 (sec)***	***AuT2 (sec)***	
	Mikon	63.25				72.05	20.55	
	Doodle	61.85				70.7	21.95	
	Art	44.6				86.11	26.65	
	Objects	75.4				59.68	20.35	
US3 PHAS (one password of each image type) – 4 in total	***Conditions/ Image types***	***SP5 % (mean % after 2 weeks)***				***RegT3 (sec)***	***AuT3 (sec)***	
	Mikon	95				57.78	15.88	
	Doodle	95	**Mean for 4 passwords: 96.25%**			58.03	17.15	
	Art	97.5				56.88	13	
	Objects	97.5				55.63	13.57	

Table 7.1: Summary of results US1, US2, US3 and all the existing multiple graphical password studies (Table 2.5)

7.1.1 Memorability of Multiple RBGS Passwords

The user studies reported in Chapters 3 and 5 corresponding to the Objectives 1 and 3 (discussed in Section 1.4, Chapter 1), have investigated the memorability of multiple RBGS passwords. The position of this thesis on the memorability of multiple RBGS passwords is based on the conclusions derived from each of the usability studies (US1 and US2), and the same is discussed below. Table 7.1 summarises the results obtained from all the usability studies (US1, US2 and US3) reported in this thesis and reported in the existing multiple graphical password studies.

RQ1: Whether multiple RBGS passwords in the current state-of-the-art are memorable, in a given experimental setting?

This research question was investigated and answered in Chapter 3. According to the statistics presented in Table 7.1, the results obtained in US1 demonstrated that the object images are the most usable in the sense of being more memorable and less time-consuming to employ, Mikon images are close behind, but doodle and art images are inferior.

The statistics presented in Table 7.1 (column name- mean login success %) also shows that the object passwords were the most memorable (mean login success percentage of 77.31%) compared to the memorability statistics reported in other multiple graphical studies (existing literature), except Everitt et al. (2009). However in US1, the object passwords were regularly used over a period of eight weeks and the frequency of usage may differ in a real life scenario, which might decrease their memorability. Moreover, the results of various single password studies presented in Chapter 2 (Table 2.3) clearly demonstrated that the mean login success of single graphical passwords range between 85% and 100% (in most studies), which is much higher than the results obtained in US1. Hence we conclude that the users will find it difficult to remember multiple RBGS passwords and their performance will deteriorate with the increasing number of passwords.

In the context of the efficiency, the statistics presented in Table 7.1 (column name – registration time) demonstrate that the mean registration time of RBGS passwords in US1 varied between 70.66s (object) to 84.44s (art). The registration time was more than the click-based passwords (43.9 sec) reported in Chiasson et al. (2007). The mean authentication time for the RBGS passwords in US1 varied between 18.28 sec (object) and 25.56 sec (art),

whereas the variation in the case of click-based passwords was between 15.1 sec and 47.0 sec, and that of the system-issued passwords (Hlywa et al., 2011) was between 24.27 sec and 31.71 sec.

Based on the results obtained in US1 (under the given experimental set-up) and the comparisons made with other studies that had reported the use of both multiple and single graphical passwords, we conclude that multiple RBGS passwords are difficult to remember and time consuming to employ with the current state-of-the-art.

RQ3: Whether the memorability of multiple RBGS passwords improves by employing a mnemonic strategy, to choose the passwords during the password registration stage, in a given experimental setting?

This research question was investigated and answered in Chapter 5 (Section 5.6.1). The memorability results obtained from the usability study (US2) reported in Chapter 5 demonstrated that the mean login success percentages of RBGS story passwords varied between 44.6% (art) and 75.4% (object), which is lower compared to the statistics reported in US1, i.e. between 54.90% (art) and 77.31% (object), as shown in Table 7.1.

The statistics presented in Table 7.1 also show that the mean registration time of story passwords (US2) varied between 59.68 sec (object) and 86.11 sec (art). However in US1, the mean registration time varied between 70.61 sec (object) and 84.44 sec (art). The authentication time in case of story passwords (US2) varied between 20.35 sec (object) and 26.65 sec (art), while the variation ranged from 18.28 sec (object) to 24.56sec (art) in US1.

These results clearly demonstrate that the object passwords were the most usable both in US1 and US2, while art passwords were the least usable, though there is a difference in the statistics reported in the respective studies. However, the statistics in relation to the mean login success percent presented in Table 7.1 and discussed above clearly show that employing mnemonic strategies do not enhance the effectiveness, i.e. memorability, and efficiency of multiple RBGS passwords.

7.1.2 Guessability of RBGS Passwords Using Written Descriptions

The user studies reported in Chapters 4 and 5 corresponding to the Objectives 2 and 4 (stated in Section 1.4, Chapter 1) investigated the guessability of RBGS passwords using their

corresponding descriptions, which were provided by the respective account holders. The position of this thesis on the guessability of RBGS passwords using their corresponding descriptions is based upon the conclusions derived from each of the guessability studies (GS1 in Chapter 4 and GS2 in Chapter 5), and the same is discussed below.

RQ2: Whether RBGS passwords can be guessed using their corresponding textual descriptions provided by the respective account holders, in a given experimental setting?

This research question was investigated and answered in Chapter 4 of this thesis. A guessability study (GS1) was conducted using the password descriptions that were provided by the subjects (respective account holders), who participated in US1. In spite of the instructions that the password descriptions should be textual (i.e. written in words), subjects in US1 drew sketches of the target images in the case of Mikon, doodle and object passwords more than the art passwords, which has been shown in Table 4.3. However, the subjects were also asked to provide a textual description for each of the target images forming the respective passwords, and these descriptions were eventually used in the guessability study.

The results obtained in GS1 (Section 4.4.3) showed that all the Mikon, doodle and art passwords used in the guessability study were successfully guessed, at least once during the guessability attack. However, only 50% of the art passwords were guessed in GS1. The statistics presented in Table 4.7 demonstrated that more than 80% of the target images for each of the image types were recorded using denotative descriptions (i.e. described the elements in the image). In the context of the denotative descriptions, Table 4.9 shows that each of the Mikon, doodle and object passwords that were used in the guessability study had denotative descriptions for at least three to four target images. Ten art passwords that were not guessed in the guessability study had denotative descriptions for at least two target images. Hence the denotative descriptions are more likely to increase the guessability of the target images in RBGS passwords. These results suggest that textual descriptions were effectively used to guess RBGS passwords in the experimental set-up used for GS1.

RQ4: Whether RBGS passwords created by employing a mnemonic strategy are guessable, using their corresponding descriptions provided by the respective account holders, in a given experimental setting?

This research question was investigated in the guessability study (GS2) reported in Chapter 5 (Section 5.4). GS2 was conducted using password descriptions that were provided by the

subjects (respective account holders), who participated in US2. The analysis of the descriptions presented in Table 5.8 (Section 5.4.3) reveals that most passwords were described using annotated/ non-annotated sketches of the target images forming the respective passwords, despite the instructions given to the subjects to write them in words. Table 5.8 (GS2) and Table 4.3 (GS1) show that most subjects recorded their RBGS password descriptions in the form of sketches, except in the case of visually complex images like art, which are difficult to draw.

The results of the guessability attacks analysed in Chapter 5 (Section 5.5.3) showed that all the Mikon, doodle and object passwords were guessed at least once in GS2, whereas only 50% of the art passwords were guessed. These results follow the same trend as reported in GS1. Hence the thesis concludes that RBGS story passwords were easily guessed using their corresponding descriptions, in the experimental set-up used for GS2.

7.1.3 Thesis Statement Validation

The discussion presented in Section 7.1.1 revealed that multiple RBGS object passwords (best performer) were memorable to similar extent in both US1 and US2. But the memorability performance of these passwords is inferior, when compared to the single RBGS password studies. However, based on the number of passwords (i.e. four) used in each of the usability studies reported in this thesis, the frequency of password usage and the experimental protocol, the highest statistic in the context of the memorability is 77.31% in US1, which did not improve even when a mnemonic strategy was added (75.4% in US2). Both these figures are less than the comparable statistics reported in the single password studies. *Hence this thesis concludes that multiple RBGS passwords are difficult to remember given the current state-of-the-art in recognition-based graphical authentication systems (RBGSs).*The discussion presented in Section 7.1.2 highlights that RBGS passwords were successfully guessed using their corresponding descriptions, in the experimental set-up used for GS1 and GS2. *Hence this thesis concludes that RBGS passwords are guessable using their corresponding descriptions.* The results of US1, US2, GS1 and GS2 do not support the claims made in the thesis statement.

Since the results of the RBGS password studies (US1 and US2) did not support the claim made regarding the memorability of multiple RBGS passwords, there was a clear need to

develop an alternative usable authentication system. In the context of the existing interest of the research community in image passwords, this thesis presented the *Passhint Authentication System (PHAS)* as a potential solution to address the problem of remembering multiple such passwords in Chapter 6.

7.1.4 Passhint Authentication

The results obtained from the usability study (US3) demonstrated that the memorability of multiple passwords in the Passhint Authentication System (PHAS) is better than in existing Graphical authentication systems (GASs), as shown in Table 7.1. The memorability performance of PHAS passwords is either better or similar to the statistics reported in the single graphical password usability studies (Chapter 2, Table 2.1, 2.2. and 2.3). Although the registration time is high, authentication time for successful attempts is either similar to or less than the time reported for previous GASs.

A guessability study (GS3) that was conducted with the same subjects who took part in US3 revealed that art passwords were the least guessable, followed by Mikon, doodle and objects, in that order. The results obtained from GS1, GS2 and GS3 strongly suggest the use of art passwords in PHAS, as it would be more resistant to being guessed using the corresponding password descriptions and hints, compared to the other image types.

The usability (US3) and guessability (GS3) study results offer proof of principle that multiple PHAS passwords are highly memorable and art images are the least guessable in PHAS. The thesis does not claim that PHAS is flawless. Hence, a number of large scale field studies need to be conducted examining various other usability and security aspects, before PHAS could be practically deployed.

7.2 Thesis Contributions

The research presented in this thesis contributes to original ideas and knowledge in the field of RBGSs. The main contributions of this thesis are enumerated below.

Identifying the research problem: Chapter 2 identifies an important limitation in the field of GASs, i.e. most usability studies have focused on the unrealistic use of a single password. In this context, Section 2.6 identifies that in the last 15 years (to our knowledge) only four

studies have explored the usability of multiple graphical passwords, and two of these studies had a high drop-out rate. The survey in Chapter 2 reveals that none of the existing studies have explored the vulnerability of RBGS passwords to descriptions, or any sort of revelation produced by an account holder, except (Dunphy et al., 2008). Hence, the thesis statement was established together with the research objectives systematically to explore the usability of multiple RBGS passwords, and their vulnerability to written descriptions.

Memorability of multiple RBGS passwords: The usability of multiple RBGS passwords has been examined in Chapter 3 with four distinct image types: Mikon, doodle, art and objects, over an online study (US1) conducted for a period of eight weeks. The results demonstrate that object images are most usable in the sense of being more memorable and less time-consuming to employ, Mikon images are close behind but doodle and art images are inferior. Another usability study (US2) is presented in Chapter 5, which examines the usability of multiple RBGS passwords when such passwords are created using a mnemonic strategy, using the same four image types as in US1. The results obtained in US2 follow the same trend as that of US1. However, the results obtained in both the studies provide concrete evidence that multiple RBGS passwords are difficult to remember, and time consuming to employ, given the current state-of-the art.

Guessability using descriptions: The vulnerability to third-party guessing of RBGS passwords, created in US1 and US2, using descriptions provided by the respective account holders is examined in Chapters 4 and 5 respectively. Both the studies show that most descriptions provided by the account holders were annotated/ non-annotated sketches of the target images forming the password. In the case of textual descriptions, these were denotative (i.e. described the elements in the image), which again helped in guessing the respective passwords. The results obtained from both the studies (GS1 and GS2) demonstrated that all the Mikon, doodle and object passwords were guessed, whereas 50% of the art passwords were guessed. It was difficult to guess art passwords using the textual descriptions and these passwords were the least amenable to sketching, compared to the three other image types. Hence these results provide evidence that art images are more resistant to being guessed using written descriptions, compared to the other image types.

Novel authentication system: A novel authentication mechanism, Passhint (PHAS), is proposed in Chapter 6. A prototype was created and two empirical lab-based studies

(usability – US3 and guessability- GS3) were conducted. The results obtained from the multiple password usability study show that PHAS have memorability advantages, over other existing GASs. The results of the guessability study (GS3) with PHAS reveal that art passwords are the least guessable, followed by Mikon, doodle and objects in that order. The results strongly suggest that the use of art passwords in PHAS, would offer usable as well as secure authentication. This thesis offers the results of the initial usability and guessability studies as a proof of principle for the Passhint system.

7.3 Future Research Directions

This thesis has contributed to the field of RBGSs and usable security literature, but has also raised further issues. In this section, a number of potential future research directions are discussed.

7.3.1 PHAS Evaluation

(A) Usability study to assess the performance of multiple art passwords in PHAS

US3 investigated the usability of four PHAS passwords, out of which one comprised of the art images. A lab-based usability study should be conducted in future to assess the performance of, 'n', number of art passwords in PHAS. This will demonstrate, whether multiple art passwords in PHAS are usable, i.e. memorable, it is easy to choose hints for multiple art passwords and efficient in terms of password creation as well as authentication time. The field of usable security also lacks comprehensive and conclusive results on text passwords, which makes it difficult to use them as benchmarks. Hence an empirical study should include the use of text passwords as the control group, using the same protocol that is used for examining PHAS passwords. This will help to systematically compare the performance of the two authentication mechanisms.

(B) Field study of PHAS

Despite the field of graphical authentication having existed for over 12 years (Biddle et al., 2009), most usability studies reported in the existing literature are lab-based. However, with the advent of ubiquitous technologies and specifically the widespread adoption of smart-phones, lab-based experiments seem to be less ecologically valid compared to the field

studies (Rogers et al., 2011). In this context, the results reported in Chiasson et al. (2007) have shown considerable discrepancies in the performance of the users in a lab-based usability study compared to a field study. This raises an important question about the combination of studies that are required to assess the viability of a specific authentication mechanism in a given context. However, we believe that the lab-based studies could be used as a proof of principle, before conducting large scale field studies.

PHAS has proven successful in terms of memorability of multiple image passwords in a lab-based study. The next logical step is to conduct a field study to examine the performance of PHAS in the real world. Such a study would help to assess the acceptability, suitability and usability of PHAS. However, it remains a challenge as to how field studies can be conducted to examine the usability of multiple graphical passwords, since most of the existing multiple graphical password studies are either lab-based or web-based, as discussed in Chapter 2.

7.3.2 Improving the Security of PHAS

(C) Varying the configuration of the system to increase the theoretical password space

Charrau et al. (2005) showed that increasing the password space, i.e. number of target images and number of challenge sets, would have a negative impact on usability, i.e. memorability will decrease and the system will be time consuming to employ. However, it is worth investigating, whether PHAS performs equally well, when the system parameters are modified to increase the theoretical password space. Such a study would help to assess the limitations of PHAS in the context of the security against guessing attacks. The aforementioned proposal was assigned to an MSc Information Security student (2013-14) as the MSc dissertation in the School of Computing Science (University of Glasgow) . The student investigated the usability of PHAS, when the number of target images is increased to six and a user is required to remember six such passwords.

(D) Additional features to increase the security of PHAS

We propose some additional features which could be implemented to enhance the security of PHAS as given below.

- *Lock out policy based on login time:* PHAS may offer more secure authentication, if the lock out policy is based not only on a definite number of failed login attempts, but a threshold value of login time. For example, once a user has used the system for α

number of times, then a timer, β, could be set for each login session. If the user is unable to complete the login session within the set time interval, then this will be recorded. After a definite number of failed attempts due to the timer expiration, the account will be locked. But, different aspects such as how to customize the timer to cater for the needs of a user and the number of attempts before the account is locked have to be considered, before this feature could be implemented in practice. Most importantly, the impact of the proposed security component on the overall usability of the system has to be considered too.

- *False Challenge Sets:* Let a user, U, select four images and give one hint for each one of them in PHAS (x_1- x_4). The system selects 15 decoy images for each of the target images (x_1- x_4), generating four challenge sets (T_1- T_4). (T_1- T_4) are the true challenge sets for the respective target images (x_1- x_4). The system would now choose four more images with their corresponding hints (F_1- F_4), which do not belong to the user U. Let the target images and the corresponding hints (F_1- F_4) belong to four different users. So (F_1- F_4) are the false challenge sets, which do not belong to the user U. In each authentication session, for the user U, the system would display *m* number of true sets selected from (T_1- T_4) and *n* number of false sets selected from (F_1- F_4). The values of *m* and *n* can either vary for each authentication session or remain constant for all the authentication sessions. Each challenge screen will have 16 images, a hint and a button named "*Ignore".* This Ignore button can be used by the legitimate user, when a false challenge set is displayed. The lockout policy may be the same, i.e. four failed login attempts. The approach for choosing the decoy images might be as follows:

  (1) 15 decoy images for each of the four challenge sets will not be a target for another challenge set;

  (2) decoy images for all the challenge sets is fixed;

  (3) sets won't change, even when the image selected is not a target (different from the configuration used in PHAS (Chapter 8)

  (4) result of the authentication will be shown once the last step is completed.

If cognitive attacks are carried out to break a PHAS password, we believe that the false challenge sets would make it difficult for the attacker to follow a lead for breaking into the system and the lock-out policy based on login time will put further pressure, making it hard

to succeed. However, rigorous usability studies need to be conducted, before these features could be adopted in practice.

7.3.3 Understanding the Topic of Descriptions

The topic of descriptions and password recordability in case of RBGSs need to be assessed at many different levels. In GS1 and GS2, the recorded descriptions of the target images forming the password were presented in the same order as they would appear in the authentication steps. However, in real life this may not be an ideal scenario. An account holder may record a prompt in any order, which might decrease the chances of successfully guessing the target images. Hence it would be worth investigating the effectiveness of guessing, when the order of the authentication steps is varied and the order of the descriptions is randomized, instead of presenting them sequentially (as in our experiments).

Another possible improvement might be to allow the account holders record their prompts as they wish. Then a study should be conducted to assess the effectiveness of guessing using the recorded prompts. This might help to understand the different approaches to record RBGS passwords, the vulnerability of each approach as well as the feasibility of each approach in a real life scenario. Moreover, the topic of descriptions in the context of the RBGSs needs to be assessed very carefully and at various levels because it relies on user behaviour, which is not only difficult to control, but varies from one individual to the other.

7.3.4 Guidelines for Designing Experiments

In the area of graphical authentication, most published results lack consistency, which makes it difficult to compare them. In context of graphical authentication systems, Biddle et al. (2009) suggested that user studies should include:

- motivation of the work, context of use and target users;
- clear description of the methodology used to conduct the usability study
- clear description of the system's design;
- Security parameters and aspects that are being investigated.

However, there are no general set of principles to design such usability studies in context of authentication mechanisms. For example, each of the multiple graphical password studies

discussed in Chapter 2 have used different experiment protocols, i.e. duration of the experiment, gap between two sessions in case of multi-session studies, number of participants, and metrics reported. Such variations in the reported studies make it difficult to systematically examine and compare the characteristics of different authentication mechanisms. Hence in order to make such studies more comprehensive and comparable, it is necessary to establish a set of common guidelines, to conduct usability as well as guessability studies with human subjects in the field of usable authentication. This set of guidelines would need to address issues such as:

- different ways of designing experiments to establish proof of principle;
- protocols to be used in the experiment, i.e. minimum duration of study, metric to be reported, interpretation of each metrics and statistical tests to be used for each metrics;
- type of training to be given before starting the experiment, length of the training and the instruction that need to be given to the uses;
- consider the potential effects of training, while interpreting the results of the study;
- establish benchmarks that could be used to compare the results obtained in a study and demonstrate its viability in a specific context.

Practitioners and experts from both the areas, i.e. HCI and security should come together and establish rules for effective experimental design, which could be adopted in the future. By taking such steps, we can edge closer to evaluating usable authentication in a systematic and comprehensive way.

7.4 Closing Remarks

Graphical passwords have been proposed as an alternative to the traditional text passwords, and the former's use is supported by cognitive theories such as the picture superiority effect (Paivio, 1986), which suggested that images, rather than words could provide a stronger foundation for the design of usable authentication systems. However, this thesis has identified that prior studies have examined the memorability of single graphical passwords (Sections 2.2 – 2.4), and studies examining the use of multiple graphical password are sparse (Section 2.6). The thesis also identified that the uptake of graphical passwords in real-world systems is low. This is likely, but not limited to:

(i) uncertainty regarding the challenges that graphical password systems might bring to the already delicate interplay between usability (for e.g. password memorability) and security;
(ii) lack of multiple password studies demonstrating the memorability of graphical passwords;
(iii) uncertainty in relation to the impact of the password coping mechanisms on the security of graphical passwords;
(iv) impact of the different image types on the usability and security of a GAS;
(v) amount of effort involved in implementation and deployment of such systems;
(vi) lack of benchmarks and consistency in reporting the graphical password schemes, which makes it difficult to compare different schemes and potentially assist their adoption.

In spite of the likely issues highlighted above, there exist a number of commercial authentication mechanisms using graphical elements. Notably, Android PatternLock allows a user to lock his or her phone by drawing a pattern connected by dots in a grid (Android, 2011). Windows 8 picture passwords combine images and gestures for authentication (Windows, 2011). The deployment of such authentication mechanisms by major commercial players shows the interest of the community in graphical passwords. Hence, in this thesis we chose to advance our research in the field of GASs.

In the context of the memorability of multiple graphical passwords (ii, above) and impact of the different image types (iv, above), this thesis contributes two usability studies (US1 and US2) and provides evidence that multiple RBGS passwords are difficult to remember, even when a mnemonic strategy is employed to select them. This contradicts the earlier findings (Section 2.6) reported in the context of multiple password use. Moreover, the results also highlight that object passwords are the most memorable and least time consuming to employ, followed by Mikon, doodles and object images, in that order.

In relation to the password coping strategies and their possible impact (iii and iv, above), this thesis contributes two user studies (GS1 and GS2) that examined the guessability of RBGS passwords using password descriptions (prompts made by the account holders to aid memorability of such passwords in subsequent use). The results highlight that the art images were the least recorded as sketches, and are most resistant to being guessed using written

descriptions, compared to the other image types (i.e. Mikon, doodle and object) examined in this thesis. Hence the empirical studies in the context of guessability shows that the RBGS passwords can be successfully guessed using the corresponding written descriptions, but the extent of the guessability would depend upon the image type used as the password.

The general goal in this thesis was to examine the memorability of multiple RBGS passwords and their guessability using written descriptions provided by the account holders. Additionally, the thesis presents a novel PHAS (Passhint authentication system), which demonstrates superior memorability of multiple image passwords, even when compared to the results reported in the existing literature (Table 7.1), thus improving the current state-of-the-art in the field of RBGSs. We recommend the use of art passwords in PHAS, based upon the results of three guessability studies (GS1, GS2 and GS3) reported in this thesis. Hence, this thesis offers the initial results in relation to PHAS, as a proof of concept, to conduct large scale field studies in the future, to further investigate its usability and security in different contexts.

On completion of this work, and having researched RBGSs over the last three and half years, we believe that the choice of authentication system should be made as a trade-off that incorporates the understanding of the: characteristics of the user population; context of the use; usage frequency; media (image type) to be used, memorability; relevant model of the likely adversaries and possible threats. In the current state, PHAS may not become main-stream due to the lack of comprehensive, consistent and comparable research contributions. However, we recommend that PHAS is best suited to a context, where the users do not frequently authenticate (providing potential memorability benefits) and the perceived security level is not high.

References

(Total number of references in this section is 92)

Adams, A. & Sasse, M.A., 1999. Users are not the enemy. *Communications of the ACM*, 42(12), pp.41-46.

Alsulaiman, F. & El Saddik, A., 2006. A Novel 3D Graphical Password Schema. *In the Proceedings of the IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems*., pp.125-128.

Android, PatternLock., 2011. *Tafasa PatternLock Features* [Online] Available at: http://www.tafasa.com/patternlock.html [Accessed 15 February 2015]

Angeli, A.D., Coventry, L., Johnson, G. & Renaud, K., 2005. Is A Picture Really Worth A Thousand Words? Exploring The Feasibility of Graphical Authentication Systems. *International Journal of Human-Computer Studies*, 63(1-2), pp.128-52.

Atinkson, C.R. & Shiffrin, M.R., 1968. Human Memory: A Proposed System and Its Control Processes. In K.W. Spence & J.T. Spence, eds. *Advances in The Psychology of Learning and Motivation*. New York Academic Press.

Aviv, A.J. et al., 2010. Smudge Attacks on Smatphone Touchscreens. *In the Proceedings of the 4th Workshop on Offensive Technologies USENIX*., pp. 1-7.

Baddeley, A., 1997. *Human Memory: Theory and Practice*. Hove, UK: Psychology Press.

Biddle, R., Chiasson, S. & Oorschot, P.C., 2009. TR-09-09 *Graphical Passwords: Learning from The First Generation*. Technical Report. Carleton University.

Brainard, J. et al., 2006. Fourth Factor Authentication: Somebody You Know. *In the Proceedings of the 13th ACM Conference on Computer and Communications Security*. Alexandria, pp. 168-178.

Brostoff, S. & Sasse, M.A., 2000. Are Passfaces More Usable Passwords: A Field Trial Investigation. In *People and Computers XIV- Usability or Else: Proceedings of HCI*., pp. 405-424.

Brown, A.S., Bracken, E., Zoccoli, S. & Douglas, K., 2004. Generating and Remembering Passwords. *Applied Cognitive Psychology*, 18(6), pp.641-51.

Carnor, L. & Garfinkel, S., 2005. *Security and usability: Designing Systems that People Can Use*. O'Reilly Media.

Charrau, D., Furnell, S. & Dowland, P., 2005. PassImages: An Alternative Method of User Authentication. *In the Proceedings of the 4th Annual ISOne World Conference and Convention*. Las Vegas.

Chiasson, S., Forget, A., Biddle, R. & Oorschot, P.v., 2009. User Interface Design Affects Security: Patterns in Click-based Graphical Passwords. *International Journal of Information Security*, 8(6), pp. 387-98.

Chiasson, S., Biddle, R. & Oorschot, P.v., 2007. A Second Look at The Usability of Click-based Graphical passwords. *In the Proccedings of the 3rd Symposium of Usable Privacy and Security*., 2007. pp. 1-12.

Chiasson, S. et al., 2009. Multiple Password Interference in Text and Click-Based Graphical Passwords. *In the Proccedings of the 16th ACM conference on Computer and Communications Security*. New York.

Chiasson, S., Oorschot, P van & Biddle, R., 2007. Graphical Password Authentication Using Cued Click Points. *Lecture Notes of Computer Science*, pp. 359-74.

Chiasson, S., Stobert, E. & Forget, A., 2011. Persuasive Cued Clickpoints: Design, Implementation, and Evaluation of a Knowledge-based Authentication Mechanism. *IEEE Transactions on Dependable and Secure Computing*, 9(2), pp.222-35.

Chowdhury, S. & Poet, R., 2011. Comparing the usability of Doodle and Mikon Images to be Used as Authenticators in Graphical Authentication Systems. *In the Proccedings of the User Science and Engineering (i-USEr)*., pp. 54-58.

Coventry, L., 2005. Usable Biometrics. In L. Cranor & S. Garfinkel, eds. *Security and Usability: Designing Secure Systems That People Can Use*. O'Reilly Media.

Davis, D., Monrose, F. & Reiter, M.K., 2004. On User Choice in Graphical Password Scheme. *In the Proccedings of the 13th Conference On USENIX Security Symposium-Volume 13*., pp. 11-11.

Dhamija, R. & Perrig, A., 2000. Deja vu: A User Study Using Images For Authentication. *In the Proccedings of the 9th Conference on USENIX Security Sumposium*., pp. 4-4.

Dunphy, P., Nicholson, J. & Olivier, P., 2008. Securing Passfaces for Description. *In the Proceedings of the 4th Symposium on Usable Privacy and Security*., pp. 24-35.

Dunphy, P. & Yan, J., 2007. Do Background Images Improve "Draw A Secret" Graphical Passwords? *In the Proceedings of the 14th ACM Conference on Computer and Communications Security*., pp. 36-47.

Everitt, K.M., Bragin, T., Fogarty, J. & Kohno, T., 2009. A Comprehensive Study of Frequency, Interference, and Trainingo Multiple Graphical Passwords. *In the Proceedings of the 27th International Conference on Human Factors in Computing Systems- CHI*., pp. 889-898.

Field, A. & Hole, G., 2003. *How to Design and Report Experiments*. Sage Publishers.

Florencio, D. & Herley, C., 2007. A Large-Scale Study of Web Password Habits. *In the Proceedings of the 16th International Conference on World Wide Web*., pp. 657-666.

Gaw, S. & Felten, E.W., 2006. Password Management Strategies for Online Accounts. *In the Proceedings of the Second Symposium on Usable Privacy and Security*., 2006.

Gilchrist, I.M. & Harvey, M., 2000. Refexation Frequency and Memory Mechanisms in Visual Search. *Current Biology*, 10, pp.1209-12.

Goldberg, J., Hagman, H. & Sazawal, V., 2002. Doodling Our Way to Better Authentication (student poster). *In the Proceedings of the ACM Conference on Human Factors in Computing Systems*., 2002.

Greisdorf, H. & O'Connor, B., 2002. Modelling What Users See When They Look At Images: A Cognitive Viewpoint. *Journal of Documentation*, 58 (1), pp. 6-29.

Hayashi, E., Christin, N. & Perrig, A., 2008. Use Your Illusion : Secure Authentication Usable Anywhere. In *Proceedings of the 4th Symposium on Usable Privacy and Security*., pp. 35-45.

Hayashi, E., Hong, J. & Christin, N., 2011. Security Through a Different Kind of Obscurity: Evaluating Distortion in Graphical Authentication Scheme. *In the Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, CHI*., pp. 2055-2064.

Herley, C., Oorschot, P.v. & Patrick, A., 2009. Passwords: If We' re So Smart, Why Are We Still Using Them. *In the Proccedings of the Financial Cryptography and Data Security*. Barbados, pp. 230-237.

Hewett, T. et al., 1996. *ACM SIGCHI Curricula for Human-Computer Interaction*. [Online] Available at: http://www.sigchi.org/cdg/index.html [Accessed 12 June 2014].

Hlywa, M., Biddle, R. & Patrick, A.S., 2011. Facing the Facts about Image Type in Recognition-based Graphical Passwords. *In the Proccedings of the 27th Annual Computer Security Applications Conference (ACSAC'11)*. New York, pp. 149-158.

Inglesant, P. & Sasse, M.A., 2010. The True Cost of Unusable Password Policies: Password Use in the Wild. *In the Proceedings of the 28th International International Conference on Human Factors in Computing Systems*., pp. 383-392.

Jain, A., Hong, L. & Pankanti, S., 2000. Biometric Identification. *Communication of the ACM*, 43(2), pp.91-98.

Jermyn, I. et al., 1999. The Design and Analysis of Graphical Passwords. *In the Proceedings of the 8th Conference on USENIX Security Symposium*.

Khlot, R., Kumaraguru, P. & Srinathan, K., 2012. WYSWYE: Shoulder Surfing Defense For Recognition Based Graphical Passwords. *In the Proceedings of the 24th Australian Computer-Human Interaction Conference*., pp. 285-294

Klein, D.V., 1990. Foling the Cracker: A Survey of , and Improvement to, Password Security. *In the Proceedings of the 2nd Usenix Security Workshop*., pp. 5-14.

Kolb, B. & Whishaw, I., 2003. *Fundamentals of Human Neuropsychology*. Worth Publishers.

Komanduri, S. et al., 2011. Of Passwords and People: Measuring the Effect of Password-Composition Policies. *In the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*., pp. 2595-2604.

Madigan, S., 1983. Picture Memory. In J. Yuille, ed. *Imager, Memory and Cognition: Essays in Honour of Allan Paivio*. Lawrence Erlbaum Associates.

Mandler, J.M. & Ritchey, G.H., 1977. Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 3(4), pp.386-96.

Mathur, P.N., 1978. Barriers to Effective Visual Communication. *Media Asia 3rd Edition*.

Mcguigan, F., 1993. *Experimental Psychology- Methods of Research*. Prentice Hall.

Miller, S., 1984. *Experimental Design and Statistics*. 2nd ed. Routledge.

Moncur, W. & Leplatre, G., 2007. Pictures at the ATM: Exploring the Usability of Multiple Graphical Passwords. *In the Proceedings of the ACM SIGCHI*., pp. 887-894.

Nali, D. & Thorpe, J., 2004. *Analyzing User Choice in Graphical Passwords*. Technical Report TR-04-01, School of Computer Science, Carleton University.

Nelson, D.L., 1979. Remembering Pictures and Words: Appearance, Significance and Name. In L.S. Cernak & F. Craik, eds. *Levels of Processing and Human Memory*. Erlbaum, Hillsdale. pp.45–76.

Notoatmodjo, G. & Thomborson, C., 2009. Passwords and Perceptions. *In the Proceedings of the Seventh Australasian Conference on Information Security*., pp. 71-78.

Oorschot, P.v. & Thorpe, J., 2008. On Predictive Models and User-Drawn Graphical Passwords. *ACM Transactions on Information and System Security*, 10(4), pp.1-33.

Orozco, M., Malek, B., Eid, M. & El Saddik, A., 2006. Haptic-based Sensible graphical Password. In *Proceedings of Virtual Concept*., pp.1-4.

Paivio, A., 1986. *Mental Representataions:A Dual Coding Approach*. Oxford Press, UK.

Parkin, A.J., 1993. *Memory: Phenomena, Experiment and Theory*. Oxford UK: Blackwell.

Patrick, A.S., Long, A.C. & Flinn, S., 2003. HCI and Security Systems. *In the Proceedings of the CHI, Extended Abstracts (Workshops)*. Florida, 2003. ACM Press.

Pering, T., Sundar, M., Light, J. & Want, R., 2003. Photographic Authentication Through Untrusted Terminals. *IEEE Pervasive Computing*, 2(1), pp.30-36.

Poet, R. & Renaud, K., 2009a. An Algorithm for Automatically Choosing Distractors for Recognition Based Authentication using Minimal Image Types. *Ergonomics Open Journal*, (2), pp.178-84.

Poet, R. & Renaud, K., 2009b. A Mechanism for Filtering Distractors for Doodle Passwords. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(5), pp.1005-29.

Real, U.C., 2004. *The Science Behind Passfaces*. [Online] Real User Corporation (1) Available at: http://www.realuser.com/published/ScienceBehindPassfaces.pdf [Accessed 12 June 2014].

Renaud, K., 2005. A Visuo- Biometric Authentication Mechanism for Old Users. *In the Proceedings of the British HCI 2005*. Edinburgh, Renaud, K. (2005a) ", pp.167-182.

Renaud, K., 2005. Evaluating Authentication Mechanisms. In L. Carnor & S. Garfinkel, eds. *Security and Usability: Designing Secure Systems That People Can Use*. O'Reilly Media.

Renaud, K., 2007. A Process for supporting Risk-Aware Web Authentication Machanisms A Usability Perspective. *Reliability Engineering and system Safety*, 92(9), pp.1024-217.

Renaud, K., 2009a. On User Involvement in Production of Images Used in Visual Authentication. *Journal of Visual Languages and Computing*, 20(1), pp.1-15.

Renaud, K., 2009b. Web Authentication Using Mikon Images. *In the Proceedings of the World Congress on Privacy, Security, Trust and the Management of E-Business*., pp.79-88.

Renaud, K. & Angeli, D., 2004. My Password is Here! An Investigation into Visuospatial Authentication Mechanisms. *Interacting with Computers* , 16(6), pp.1017-41.

Rogers, Y., Sharp, H. & Preece, J., 2011. *Interaction Design:Beyond Human-Computer Interaction*. 2nd ed. John Wiley and Sons.

Ross, S., 1999. *Unix System Security Tools*. McGraw-Hill.

Rovee-Collier, C., Hayne, H. & Colombo, M., 2001. *The Development of Implicit and Explicit Memory*. John Benjamins Publishing.

Sasse, M.A., Brostoff, S. & Weirich, D., 2001. Transforming the Weakest Linka Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal*, 19(3), pp.122-31.

Shepard, R., 1967. Recognition Memory for Words. Sentences and Images. *Journal of Verbal Learning and Verbal Behaviour*, 6, pp.156-63.

Strauss, A. & Corbin, J., 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* Sage, Newbury Park.

Sturken, M. & Cartwright, L., 2012. *Practices of Looking: An Introduction To Visual Culture*. Oxford Press.

Suo, X., 2006. *A Design and Analysis of Graphical Password*. Master's Thesis, College of Arts and Science, Georgia State University.

SUS., 2011. *System Usability Scale*. [Online] usability.gov Available at: http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [Accessed 12 June 2014].

Szekely, A. & Bates, E., 1999. *Objective Visual Complexity as a Variable In Picture Naming*. CRL Newsletter Center for Research in Language, University of California.

Takada, T., Onuki, T. & Koike, H., 2006. Awase-e: Recognition-Based Image Authentication Scheme Using Users' Personal Photographs. *In the Proceedings of the IEEE, Innovations in Information Technology*., pp.1-5.

Tao, H. & Adams, C., 2008. Pass-Go: A Proposal to Improve the Usability of Graphical Passwords. *International Journal of Network Security*, 7(2), pp.273-92.

Tullis, T.S. & Tedesco, D.P., 2005. Using Personal Photos as Pictorial Passwords. I *In the Proceedings of the CHI, Extended Abstracts on Human Factors in Computing Systems*., pp.1841-1844.

Tullis, T.S. & Tedesco, D., 2011. Can Users Remember Their Pictorial Passwords Six Years Later. *In the Proceedings of the CHI*., pp.1789-1794.

Valentine, T., 1999. *Memory for Passfaces After a Long Day*. Technical Report. London Goldsmith College.

Varenhorst, C., 2004. *Passdoodles: A Lightweight Authentication Methods*. MIT Research Science Institute.

Weinshall, D., 2006. Cognitive Authentication Schemes Safe Against Spyware (short paper). *In the Proceedings of the IEEE Symposium on Security and Privacy*.

Weiss, R. & Luca, A.D., 2008. PassShapes - Utilizing Stroke Based Authentication to increase Password Memorability. *In the Proceedings of the NordiCHI*., pp.383-392.

Wiedenbeck, S. et al., 2005a. PassPoints: Design and Longitudinal Evaluation of a Graphical Password System. *International Journal of Human-Computer Studies*, 63(1-2), pp.102-27.

Wiedenbeck, S. et al., 2005b. Authentication Using Graphical Passwords: Effects of Tolerance and Image Choice. I*n the Proceedings of the1st Symposium of Usable Privacy and Security*., pp.1-12.

Wiedenbeck, S., Waters, J., Sobrado, L. & Birget, J.C., 2006. Design and Evaluation of a Shoulder-Surfing Resistant Graphical Password Scheme. *In th*e *Proceedings of The Working Conference on Advanced Visual Interfaces, AVI*., pp.177-184.

Windows, Picture Password, 2011. *Signing in with a Picture Password* [Online] Building Windows 8 Available at: http://blogs.msdn.com/b/b8/archive/2011/12/16/signing-in-with-a-picture-password.aspx [Accessed 14 February 2014].

Wolfe, M., 1994. Guided Search 2.0 A Revised Model of Visual Search. *Psychonomic Bulletin & Review*, 1(2), pp.202-38.

Wolfe, M., 2003. Moving Towards Solution to Some Enduring Controversies in Visual Search. *Trends in Cognitive Science*, 7(2), pp.70-77.

Appendix A
Images used in user studies

Figure A1: Sample Mikon images for My Jokes (coloured Mikons)

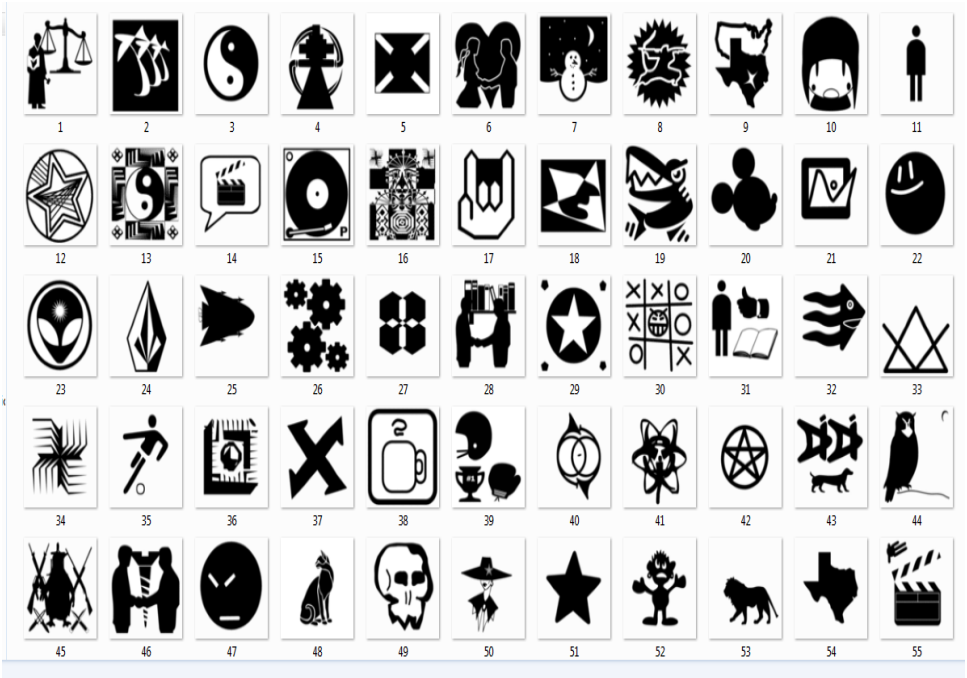


Figure A2: Sample Mikon images for My Movies (Black and white Mikons)

Figure A3: Sample Mikon images for My News (Mikons with annotations)

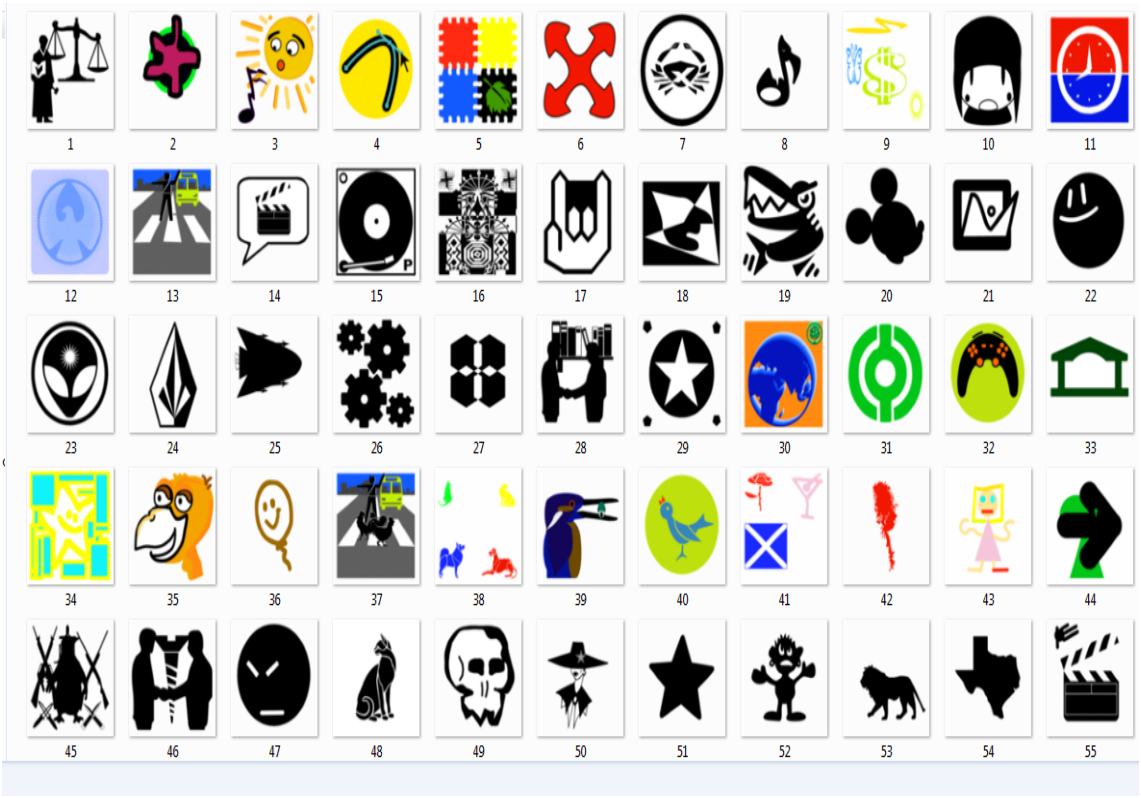


Figure A4: Sample Mikon images for My Status (combination of A1, A2 and A3)

Figure A5: Sample doodle images for My Jokes

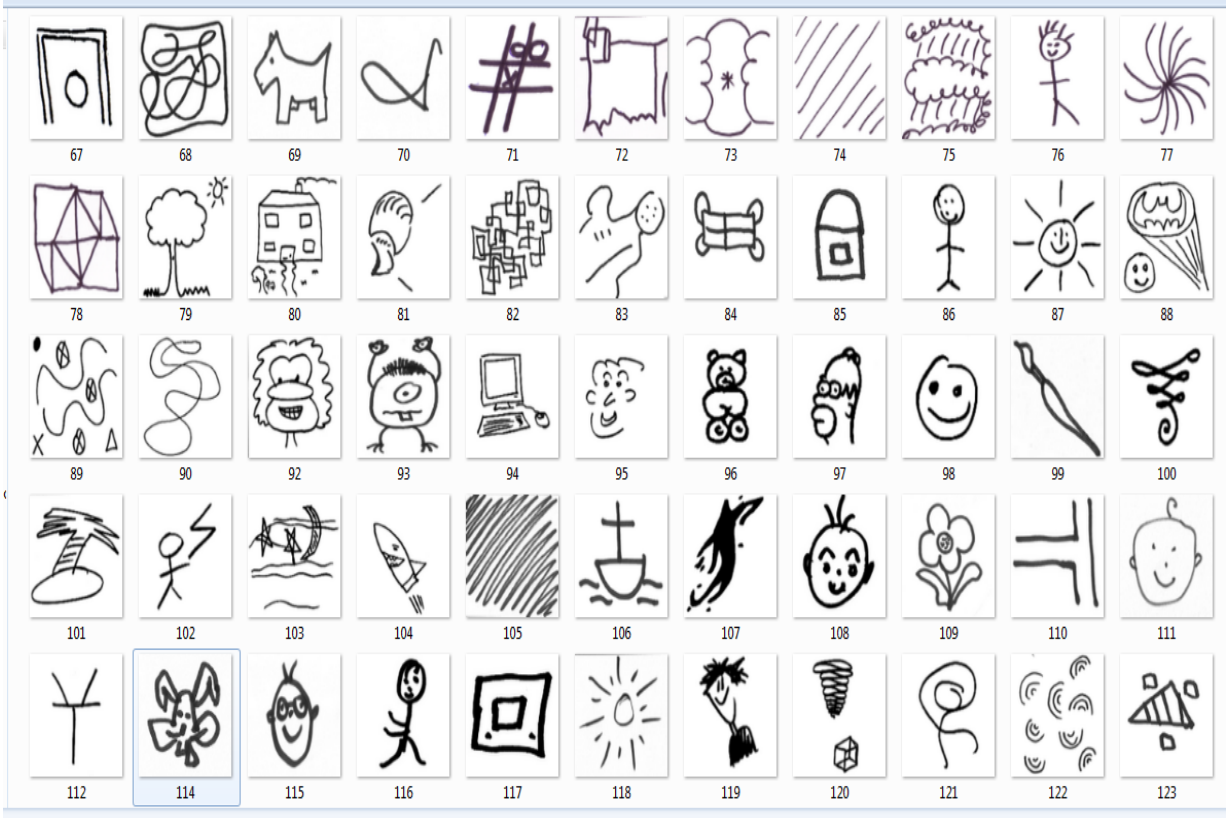


Figure A6: Sample doodle images for My Movies

Figure A7: Sample doodle images for My News

Figure A8: Sample doodle images for My Status

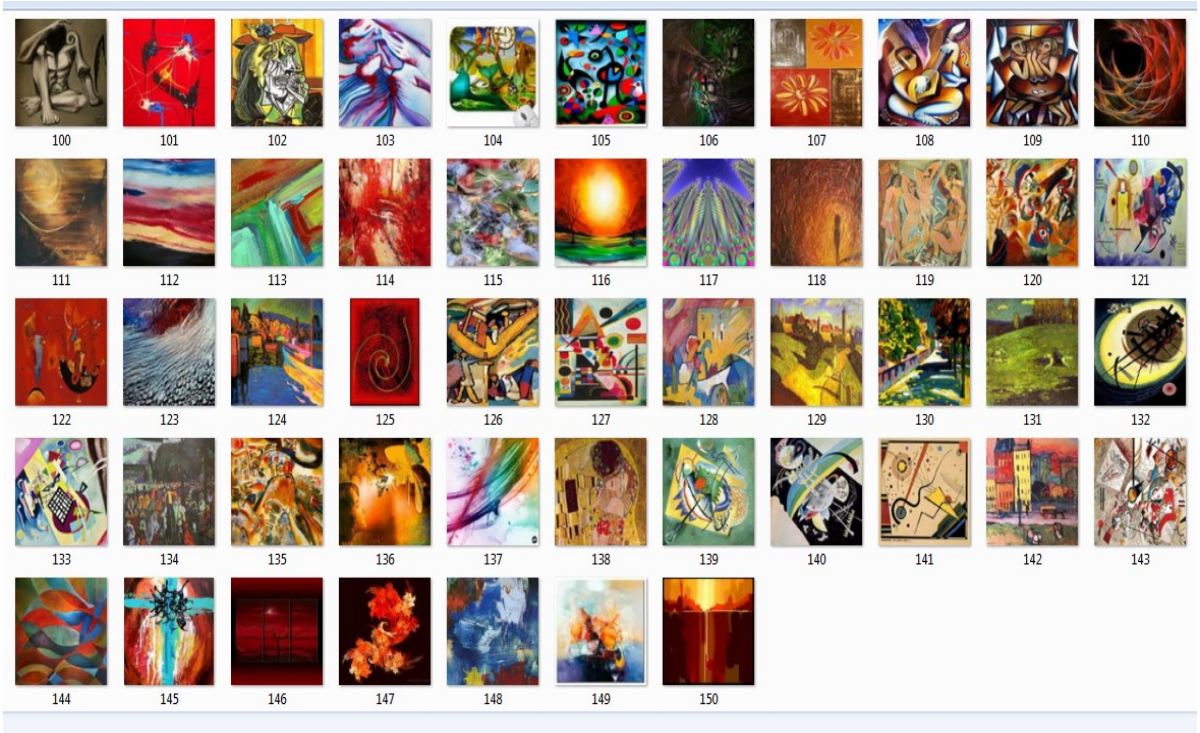


Figure A9: Sample art images for My Jokes

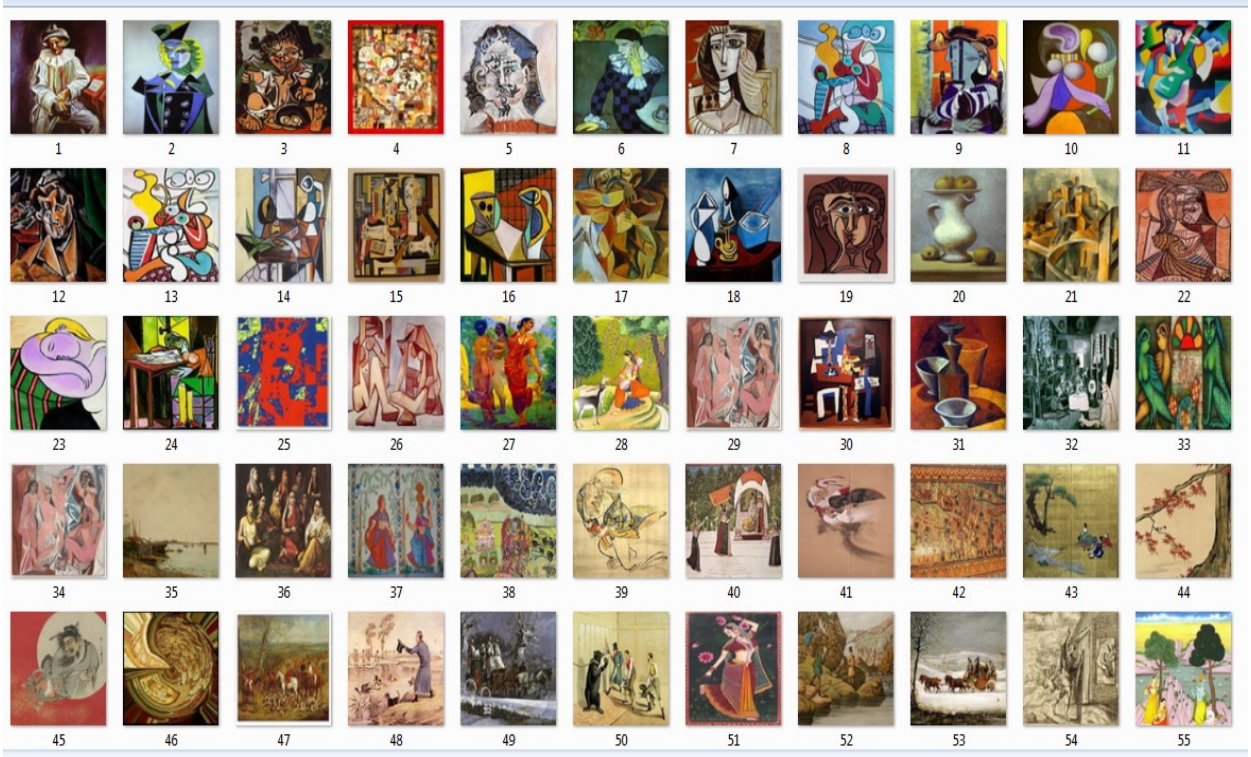


Figure A10: Sample art images for My Movies

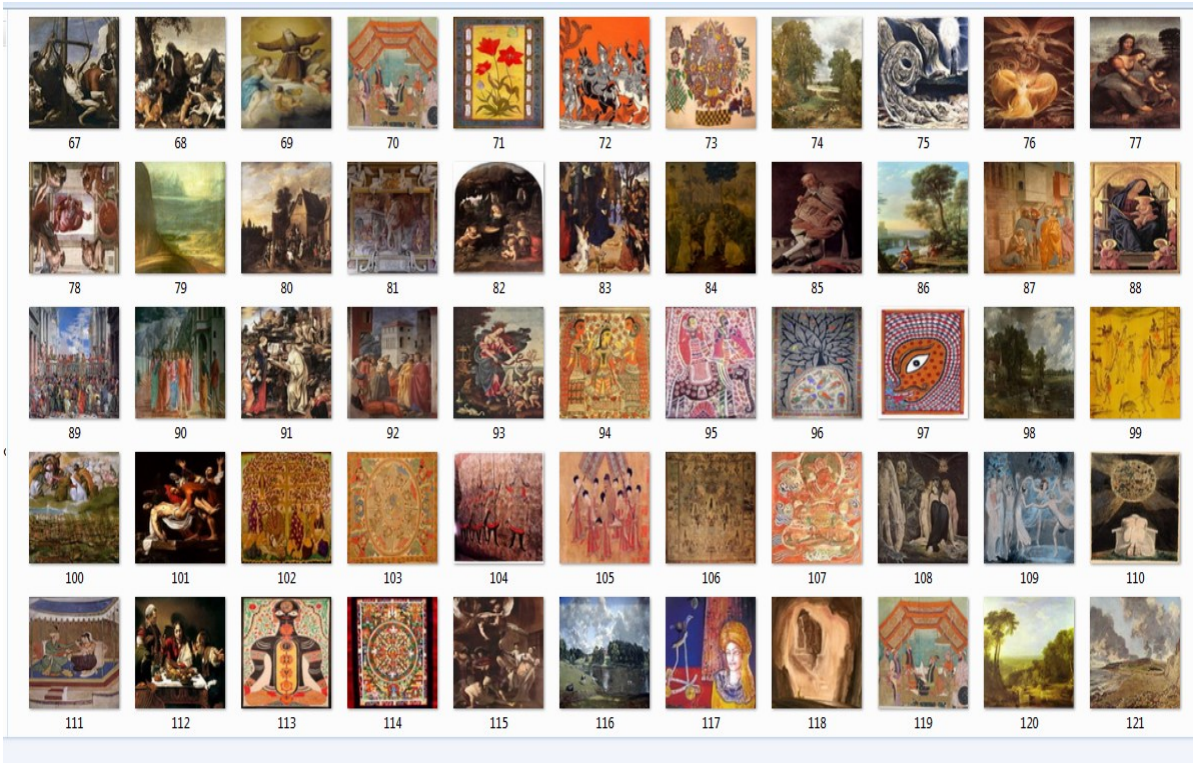


Figure A11: Sample art images for My News

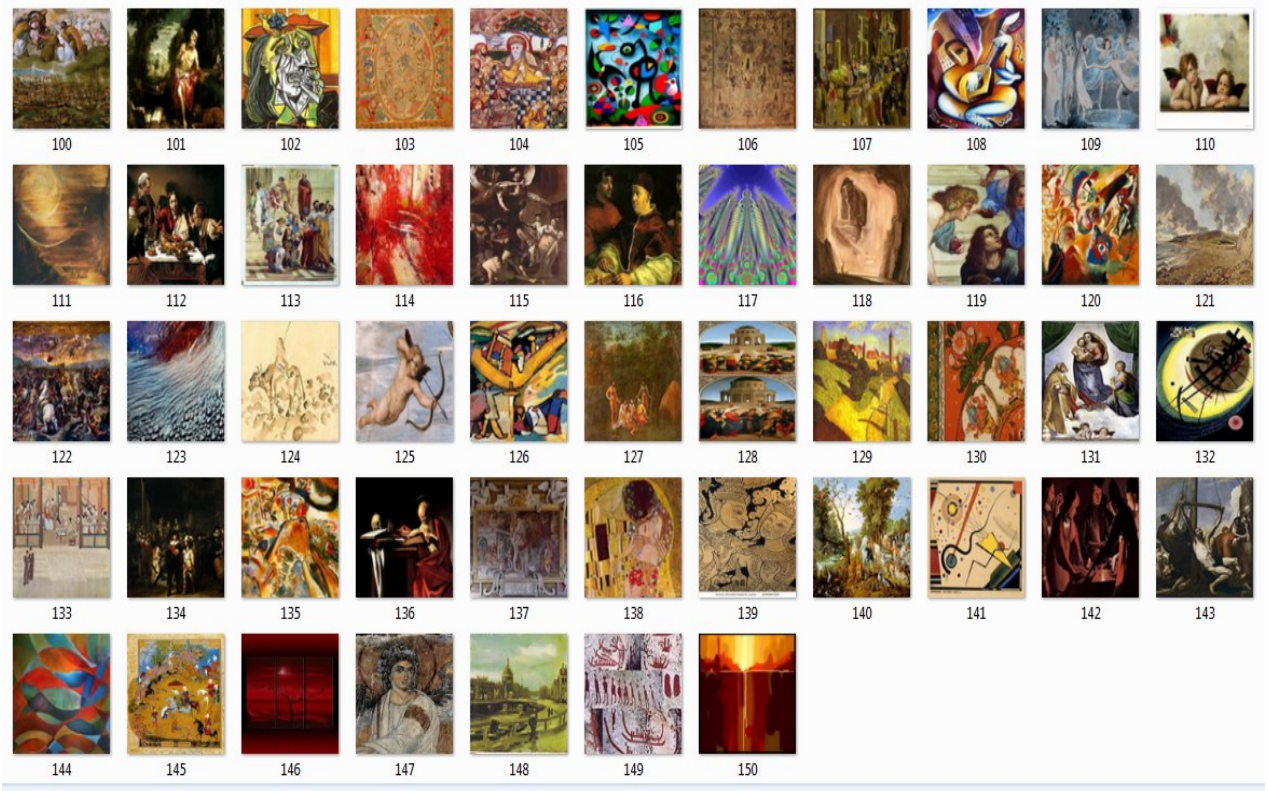


Figure A12: Sample art images for My Status

Figure A13: Sample object images for My Jokes (images of food and drinks)

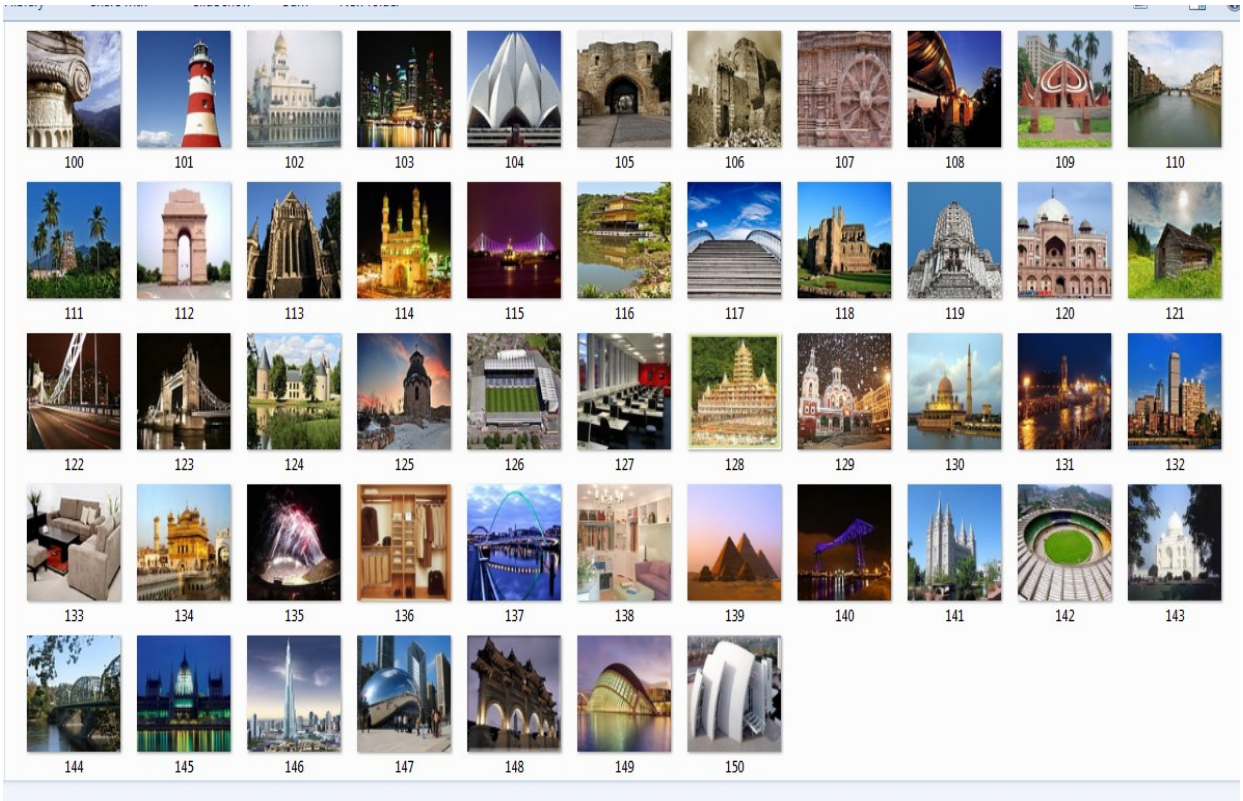


Figure A14: Sample object images for My Movies (images of buildings)

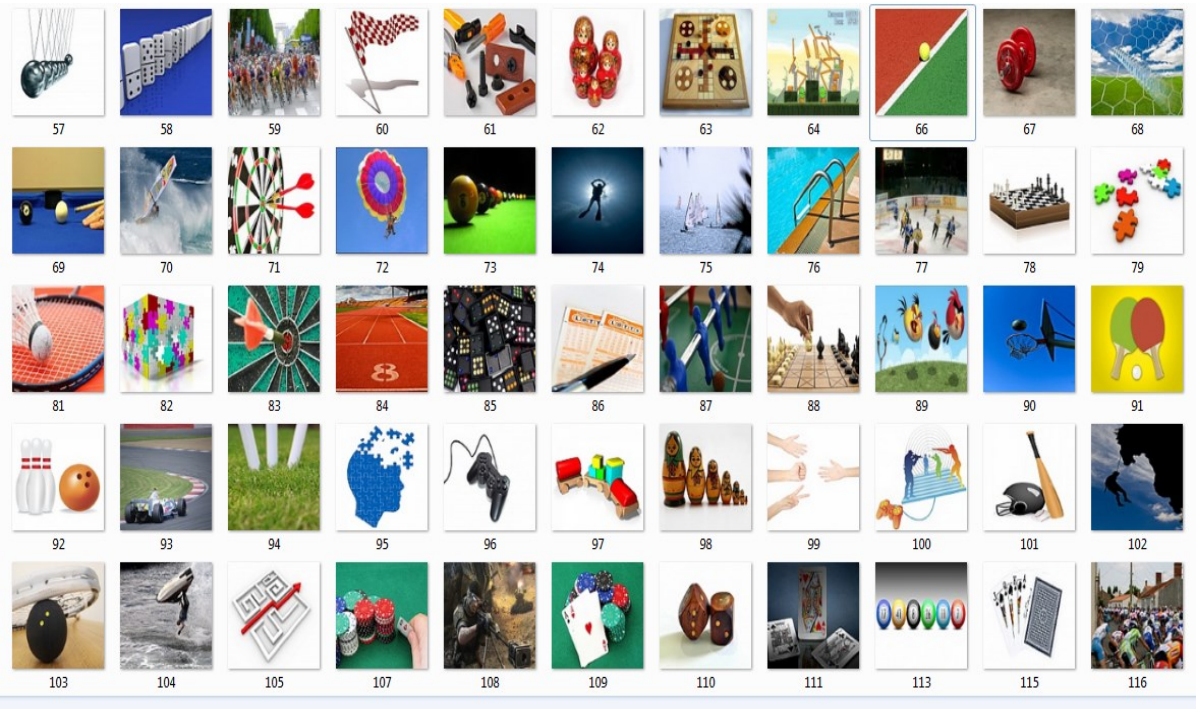


Figure A15: Sample object images for My News (images of sports and leisure activities)

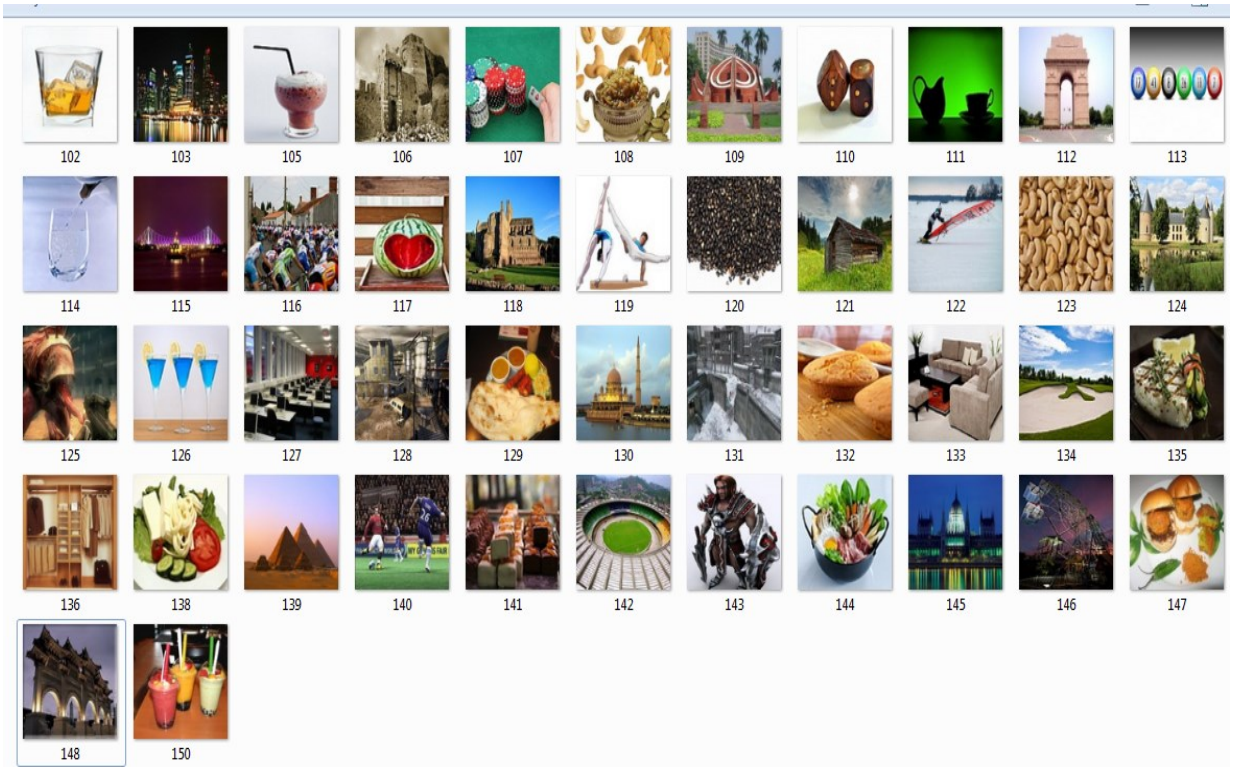


Figure A16: Sample object images for My Status

Appendix B.1
Pre-study questionnaire (sample) US1

Use of Passwords- Authentication
(Pre-study Questionnaire)

Hi

This survey is being conducted as a part of my PhD Research Project on Graphical Authentication Systems in University of Glasgow. As a part of the study, you are asked some questions about the use of passwords in your day to day life. Thank you for your time and taking part in the survey.
If you have any further questions you can mail me: soumc@dcs.gla.ac.uk
I will be happy to answer any queries from your side.

Soum Chowdhury
University of Glasgow

1)

Please specify your Gender

Male	
Female	

2)

Please specify your age

15-20	
21-25	
26-30	
31-35	
36-40	
41 or above	

3) How many distinct passwords do you have ? Please specify the number. For example Mr X has following accounts , that require him to remember a password. 2 email accounts – different passwords 2 social networking accounts – 1 distinct and 1 similar to email account 1 online shopping account – similar to social networking account The total number of passwords Mr X has is 3,	
1-3	
4-6	
7-9	
10 or above	

4) Do you tend to forget your passwords, please specify the reasons for the same	
5) What strategy do you use to **remember** your passwords	
Same passwords for different accounts	
Similar password for different accounts	
Different combination of the letters in the password	
Password somehow personally related to you	
Use a random password and write it down	
6) What other strategies do you use to remember your passwords	

7) Why do you employ these strategies to remember your passwords
8) Do you feel that text passwords are secure, considering the strategy you use to remember them. Please specify the reason for your choice too.

9) Have you ever used image as a password. For example, a system that makes you choose an image as password and during login you have to pick the password image to gain access to the system.	
Yes	
No	
10) If the answer to the above question is YES please specify where have you used such a system	
Could you please provide your email id if we need to contact you further	

Appendix B.2
Post-study questionnaire (sample) US1

Rate for 1 to 5: 1 is the lowest rating (worst) and 5 is the highest rating (very good)

1. How easy was the registration process: 1 2 3 4 5
   (Easy in the sense of selecting four images in each of your four passwords)

2. What is good about the registration process (if any)?

3. What are the problems/ difficulty you faced during registration (if any)?

Rate for 1 to 5: 1 is the lowest rating (worst) and 5 is the highest rating (very good)

- Images used in My Jokes:
- Images used in My Movies:
- Images used in My News:
- Images used in My Status:

Rate for 1 to 5: 1 is the least meaningful (worst) and 5 being most meaningful

- Images used in My Jokes:
- Images used in My Movies:
- Images used in My News:
- Images used in My Status:

Rate for 1 to 5: 1 is the lowest rating (worst) and 5 is the highest rating (very good)

1. How easy was the authentication process: 1 2 3 4 5
   (Easy in the sense of remembering four images for each password)

2. What is good about the authentication process (if any)?
3. What are the problems/ difficulty you faced during authentication (if any)?

Rate for 1 to 5: 1 is the lowest rating (worst) and 5 is the highest rating (very good)

1. Your overall rating for the images used in the system: 1 2 3 4 5
   (How satisfied were you after using the image used in the system for 8 weeks)

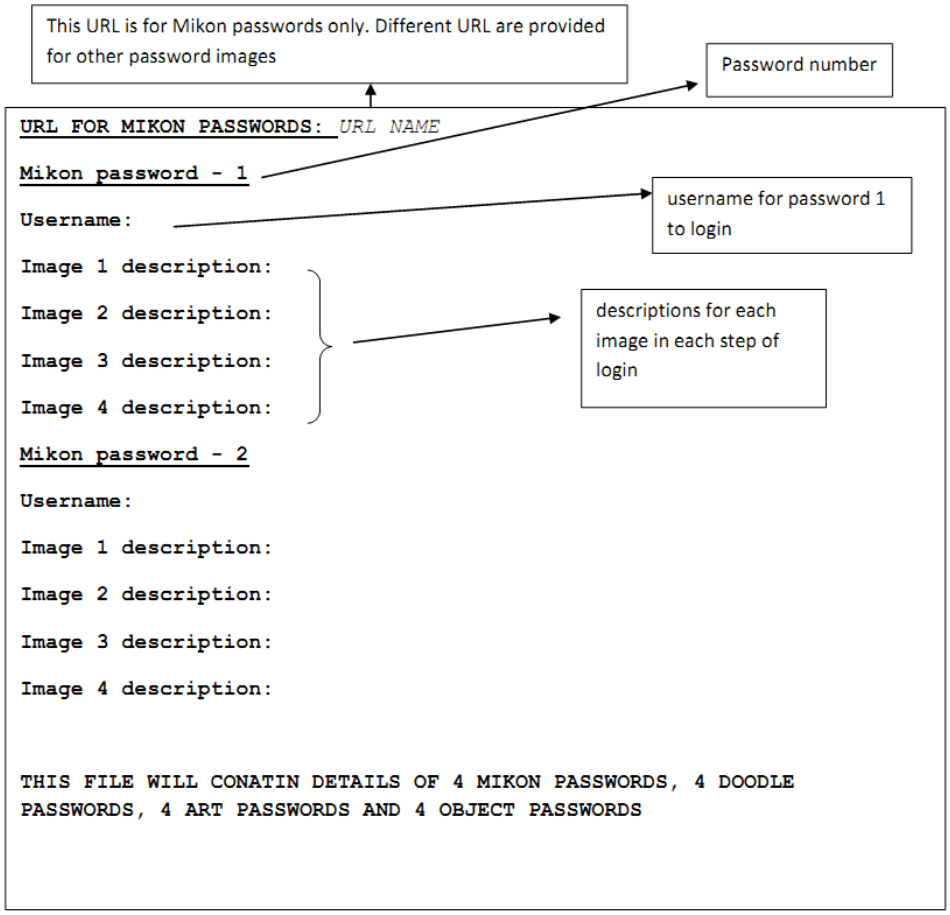
Appendix C
Task Sheet for guessability study (GS1)

TASK INFORMATION SHEET- PARTICIPANTS

The aims of this experiment is **Study the guessability of image passwords using user written descriptions**

You will try to guess 16 passwords - each password comprising of 4 images.

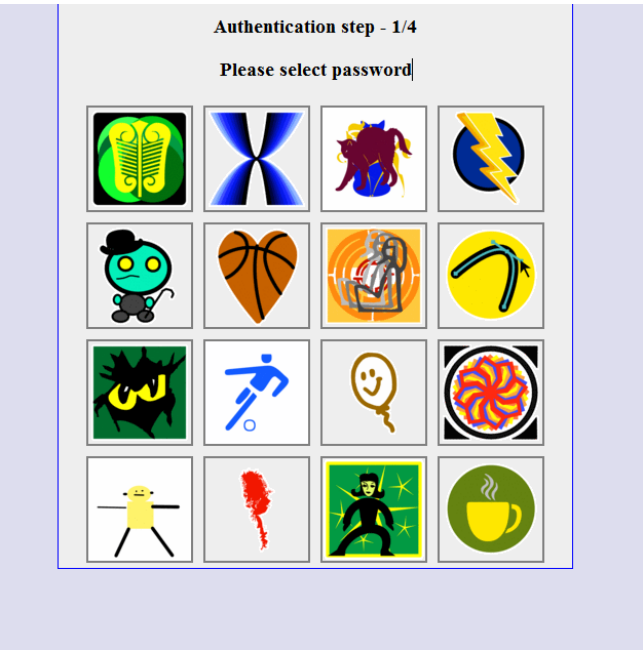
1. You will be mailed one website links together with some user names. The same mail will also give description of 16 passwords in a text file. Each username will correspond to one of the passwords. The contents of the text file have been described and explained below



2. You have to go to the URL and then enter the user name for Authentication/ Login

Step 1: Enter the username for the password-n and then click login

Step 2: Following page will open. This is the first step of authentication. In this step 16 images are displayed: one password image and 15 other images that are not your password. Guess the image according to the description given to you in the text file corresponding to password-n, image-1.



Step 2: as soon as you click on an image it will take you to the Step 2 which is exactly same as before. Guess the image according to the description given to you in the text file corresponding to password-n, image-2.

Step 3: as soon as you click on an image it will take you to the Step 3 which is exactly same as before. Guess the image according to the description given to you in the text file corresponding to password-n, image-3.

Step 3: as soon as you click on an image it will take you to the Step 4 which is exactly same as before. Guess the image according to the description given to you in the text file corresponding to password-n, image-4.

Wrong authentication: If you have selected wrong password image at any step of authentication following page will open. You can use the home button in the page to go to the home page and start login again. The success/ failure of the login will be displayed after step 4 of authentication.

You need to choose the images according to the descriptions given in the file mailed to you. No extra information will be provided to you other than the text file.

3. If your login is a failure, then you have to try again. You will be given total of 4 chances to identify each password. Once you have used all the four chances, you will move on to the next password. Once you have finished your task, you will be mailed the details of the next guessability attack.

Please feel free contact me if you have any difficulty.
Soumyadeb Chowdhury
PhD student, Computing Science Department, University of Glasgow
Email: soumc@dcs.gla.ac.uk

Appendix D
Password descriptions in GS2

Object password Description (two sketches)

Figure D1. Sample object password (US2)

Mnemonic strategy: *My favorites.*

Image 1: A room with many PCs;

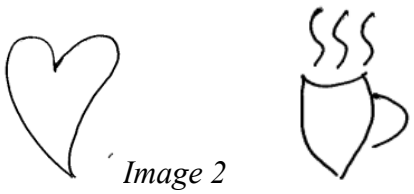


Image 2 *Image 3*

Image 4: Two people doing gymnastics

Doodle password Description (all sketches)

Figure D2: Sample doodle password (US2)

Mnemonic strategy: *Things I love.*

Mikon password Description (three sketches)

Original Passwords

Figure D3: Sample Mikon password (US2)

Mnemonic strategy: *Same color.*

Image 1

Image 2: Red square with a blue pentagon inside;

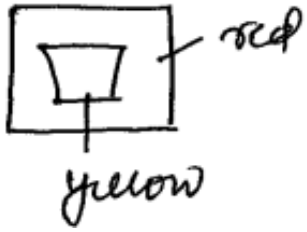


Image 3

Image 4

Art Password Description (all in words)

Figure D4: Sample art password (US2)

Mnemonic strategy: *Dance and Music.*

- *Image 1:* Two tress and two people dressed in yellow color;
- *Image 2:* Women dressed in red with a goad and green background;
- *Image 3:* A women dressed in red dancing and background is red;
- *Image 4:* A women dressed in red and orange dancing and the background is green

Appendix E
Theoretical password space computation

The theoretical password space for RBGSs depends upon the

- Number of rows in the challenge set (row)
- Number of columns in the challenge set (col)
- (row × col) = total number of images in a challenge set (total)
- Number of challenge sets (c)
- Number of target images forming a password (t)
- Order of selecting the target images (ordered/ un-ordered)

Type 1: In case of unordered selection (i.e. order of selecting the target images in the challenge set is irrelevant) and number of challenge sets is more than one, the theoretical password space is calculated as in Eq. E1 (Hlywa et al., 2011)

$$\log_2(row\ x\ col)^t \qquad \text{(Eq. E1)}$$

The calculations for each system (Table 2.4 and Table 2.5), where the selection of target images are unordered is given shown below.

- Passfaces (Real, 2004) and Faces (Davis et al., 2004)

$$\log_2(3\ x\ 3)^4 = 12.67 \text{ bits}$$

- Faces (Everitt et al., 2009)

$$\log_2(3\ x\ 3)^5 = 15.72 \text{ bits}$$

- (Hlywa et al., 2011)

$$\log_2(5\ x\ 5)^6 = 27.86 \text{ bits}$$

$$\log_2(4\ x\ 4)^5 = 20 \text{ bits}$$

- Doodle (Renaud, 2009a) and Mikon (Renaud, 2009)

$$\log_2(4\ x\ 4)^4 = 16 \text{ bits}$$

- Use Your Illusion (Hayashi et al., 2011)

$$\log_2(9)^3 = 10 \text{ bits}$$

- AWASE (Takada et al., 2006)

$$\log_2(9)^4 = 12.67 \text{ bits}$$

- Photographic authentication (Pering et al., 2003)

$$\log_2(4)^{10} = 20 \text{ bits}$$

Type 2: In case of unordered selection (i.e. order of selecting the target images in the challenge set is irrelevant) and number of challenge sets is one, the theoretical password space is calculated as in Eq. E2 (Biddle et al., 2009)

$$\log_2 \frac{(total)!}{t!(total-t)!} \quad \text{(Eq. E2)}$$

- VIP 3(Angeli et al., 2005)

$$\log_2 \frac{(16)!}{4!(16-4)!} = 10.82 \text{ bits}$$

- Dejavu (Dhamija & Perrig, 2000)

$$\log_2 \frac{(25)!}{5!(25-5)!} = 15.69 \text{ bits}$$

- Cognitive (Weinshall, 2006), for total = 110

$$\log_2 \frac{(110)!}{60!(110-60)!} = 73 \text{ bits}$$

- Convex hull (Wiedenbeck et al., 2006)

$$\log_2 \frac{(45)!}{5!(45-5)!} = 20.22 \text{ bits}$$

- WYSWYE (Khlot et al., 2012)

$$\log_2 \frac{(25)!}{4!(25-4)!} = 13.62 \text{ bits}$$

Type 3: In case of ordered selection (i.e. order of selecting the target images in the challenge set is important), the theoretical password space is calculated as in Eq. E3 (Biddle et al., 2009)

$$\log_2 \frac{(total)!}{[(total)-t]!} \qquad \text{(Eq. E3)}$$

- Story (Davis et al., 2004)

$$\log_2 \frac{(9)!}{[(9)-4]!} = 11.56 \text{ bits}$$

- (Moncur & Leplatre, 2007)

$$\log_2 \frac{(10)!}{[(10)-4]!} = 12.29 \text{ bits}$$

- VIP 1 and 2 (Angeli et al., 2005)

$$\log_2 \frac{(10)!}{[(10)-4]!} = 12.29 \text{ bits}$$

- PassImage (Charrau et al., 2005)

$$\log_2 \frac{(25)!}{[25-6]!} = 26.90 \text{ bits}$$