How Viable are Stubblefield and Simon's Inkblots as Password Cues?

Karen Renaud & Tony McBryan
University of Glasgow
{karen@dcs.gla.ac.uk}

ABSTRACT

Computer users are generally authenticated by means of a passwords which are often forgotten and written down. Replacement is expensive and inconvenient. Stubblefield and Simon [1] proposed using an inkblot as a password cue to reduce the incidence of password forgetting. Is this mechanism feasible? In this paper I will outline two experiments carried out in order to determine viability of these images as password cues.

1. INTRODUCTION

Computer users need to authenticate themselves, mostly by means of a secret password. People have to remember multiple passwords and since human memory is fallible, people forget their passwords, and need reminders or replacements.

In this paper the issue of password cueing is addressed. This term may seem to be an oxymoron since passwords are a security tool, and need to remain secret. Cues could tear a large hole in the security ostensibly maintained by the password, if not carefully chosen.

Stubblefield and Simon propose that cues could be provided in the form of an abstract inkblot-like image which makes the cue itself is so obscure and vague that it acts as a cue only to the legitimate owner of the password. The viability of this proposal can only be proven if two questions can be answered in the affirmative:

1. is the inkblot the best possible image to use as a cue?
2. will users make use of the inkblot if it is provided at authentication?

In Section 2 secret-based authentication is reviewed. Section 3 confirms the potential of images as cues and explores the kinds of images that could serve as cues. Section 4 outlines the methodology followed in order to identify the best cueing image. Section 5 presents the results and identifies the best image type in order to answer the first question posed above. Section 6 reports on the experiment that tested the use of the best image type as a cue during authentication, which answered the second question. Section 7 concludes.

2. AUTHENTICATION

In order to grant access to a restricted digital space, a two phase protocol is used: identification followed by authentication. In the face of fallible human memory and insecure communication channels this tends to fail, so that passwords have to be replaced [2], which presents problems.

The replacement process weakens the mechanism because a replacement key has to be delivered in some way and this delivery can be intercepted by an intruder. If challenge questions are used instead of a straight replacement the authentication mechanism is weakened unacceptably because of the difficulty of choosing questions. If the user has to generate the question he or she/he is equally likely to forget the question as the password. If the system has a set group of questions these need to be applicable to a wide range of users. Site owners resort to setting widely applicable questions and only a relatively superficial knowledge of the legitimate user is often required in order to discover the answers to these questions.

The replacement has to be funded. Gartner [2] claims that a single replacement costs between $15 and $30 and each employee will call about 5 times a year. A cheap alternative is simply to send people their passwords by email, but since email is seldom encrypted, this option can only be used for insecure systems, and only where people haven't forgotten their email password.

The ideal situation would be for a suitable cueing mechanism to be identified which could help users to remember their passwords hence reducing the incidence of password replacements.

2.1 Cueing Mechanisms

A cue can be defined as: a. A reminder or prompting, or b. A hint or suggestion. A cue heard by someone other than the person for whom it is intended, therefore, could produce the same association or act as the same reminder as it was intended to elicit in the target person. In authentication such a universal cue is useless since it undermines security. A cue used in an authentication setting needs to be deliberately obtuse.

Hertzum [3] proposes that users specify particular password characters which will be displayed at password entry in order to jog their memory. This idea was tested with 14 users and it did help them to remember their passwords. He notes that the defined passwords were often weak and some kind of cueing mechanism is required in order to support the use of longer and stronger passwords.

The proposed inkblot cues rely on the fact that there is strong evidence that pictures are more memorable than words, ie the *picture superiority effect* [4]. A purely representational image will not work in this secure context because what one really needs is an image that elicits a different textual association from different users so that intruders cannot confidently guess textual associations within the three strikes allowed before a lockout.

Stubblefield and Simon [1] experimented with using inkblots to assist users to form a semantic association with the textual password, which could be used as a reminder mechanism as required. They displayed 10 inkblots in a particular sequence. For each blot the user was required to enter two characters — the starting and ending character of their inkblot description. They had some success in trials of this mechanism, achieving an entropy of 4.09 bits per character. However, the cognitive load imposed on the user is significant. They do not merely provide a textual description; they have to parse it in their minds to extract the required starting and ending character, and then type that in. Stubblefield and Simon do not give demographic information about their experimental subjects but one can envisage this cognitive load being untenable for any but the most mentally agile of users. Lab-based studies often deliver unrealistic results. One can only reliably conclude the viability of a mechanism by means of a test "in the wild", where users have to actually use the mechanism to access a protected resource.

Our first question requires us to determine what kind of image could best support cueing activity in an authentication setting. We need to find out what characteristics this image would have to exhibit to facilitate superior recall and low predictability. The image descriptions would also have to be more durable than random textual passwords in order to improve the current situation.

A series of experiments were conducted in order empirically to verify the use of images in this context. Before discussing our experiments, we need first to discuss different image types and the effects of human vision on the image choice.

3. HUMAN VISION

One of the most vital of the human senses is vision. When an object is seen, the viewer will compare that object to an internal "database" of objects within his or her mind, and use past experience to match that object with the object being seen in order to identify it. Thus visual perception interacts with perceptual processes but also with memory, reasoning and communication [5].

This research considers the use of images as cues. In order to act as a cue in an authentication environment, the image must have the following characteristics:

1. *Ambiguity* — The image cue should mean different things to different people.

2. *Efficacy* — Human memory for pictures and their textual description needs to be superior to word memory so that the cueing mechanism excites a durable association. Furthermore, the textual description needs to be strong enough to act as a password.

The following two sections address these concepts in greater detail.

3.1 Ambiguity

The Gestalt psychologists formulated a set of laws of organisation that help us understand the perceptual filling-in process. The laws relate to [6]: *Closure, Good Continuation, Proximity, Similarity, Relative Size, Surroundedness, Orientation and Symmetry* and *Common Fate*. Ambiguity requires images that are vague in terms of the Gestalt laws.

We need a way to describe different candidate image types so that we can arrive at a particular description of an efficacious image type that can act as a cue. Alario and Ferrand [7] classified a number of images and propose the following norms to describe them:

- *Name agreement* — the degree to which the people agree on the name of the picture;
- *Image agreement* — the degree to which the person's mental image matches the picture;
- *Familiarity* — the familiarity of the concept being depicted;
- *Visual complexity* — measuring the number of lines and details in the picture;
- *Image variability* — indication of whether the name of an object invokes many or few images for the object.

These norms will be used in later sections to delineate the kinds of images the cueing application requires. Obviously representative images have high name agreement and this disqualifies them. We are left with the broad class of *abstract images*. If we are able to identify such a suitable image class, our next concern is the efficacy of the textual description a particular image member of that class will elicit, in terms of acting as a password cue.

In addition to abstract images, *human faces* were included in the experiment. The face is a special image as far as humans are concerned. Humans can identify thousands of faces without difficulty, suggesting limitless and durable memory for faces which could be exploited.

3.2 Efficacy

Efficacy encompasses more than one aspect:
Descriptiveness — Humans should have the ability to describe the pictures in a textual format — this is termed *picture naming*.
Strength — The text association needs to have either length or complexity, which make it harder to break.
Memorability & Durability — The text association should be durable in the sense that users are able to reproduce it perfectly after a time lapse.

3.2.1 Descriptiveness

The central premise of this research has been that we can rely on the previously-mentioned picture superiority effect *accompanied by* reliably retained textual descriptions.An abstract image is more expressive than a representative image, and does not have a simple label, but requires interpretation and verbalisation. For example, consider the process involved in assigning a name the inkblots. Rapaport [8], referring to the Rorschach inkblot verbalisation process, argues that such a process is an *"association process initiated by the inkblots as stimuli"* (p91). The results of the association process need to be converted to language, and this process is highly dependent on individual factors [9]. Hence even if two people perceive a particular image as belonging to the same semantic class they are likely to verbalise it in slightly different ways. It is hoped that these individual differences will lead to syntactically different picture descriptions and therefore distinctly different passwords.

3.2.2 Strength

Passwords are generally broken in one of two ways if there is no cue: brute strength or dictionary attack. The former only works if the password file can be obtained and is very time consuming. Most would-be attackers will try the dictionary attack and only resort to brute force if that fails. The dictionary attack exploits the fact that most people will use a recognisable word in their own language and works its way through dictionaries until a match is found.

To make it harder for a dictionary attack to succeed we need to make the password less susceptible to this kind of attack. There are two ways of doing this — either by making the password longer

by using more than one word or by making it more complex by including numeric and other special characters.

Since we're asking people to describe non-representational images, we would expect to see longer passwords, which will contribute towards strength. Furthermore, there is evidence that previously seen pictures are named faster than new pictures [10]. Hence by timing responses a system may be able to infer that a possible intrusion attempt is underway. Since abstract images may well initiate the same semantic association in the legitimate user and the intruder, but a slightly different syntactical conversion is produced, the best way to prevent an intruder from trying different possible descriptions until he or she succeeds is by judicious use of the "three tries lockout" policy.

3.2.3 Memorability & Durability

The picture superiority effect states that humans remember pictures better and for longer than words. Psychologists have demonstrated this with a number of experiments [11, 12, 13]. Humans, having seen an image once, will readily be able to attest to the fact, and this effect is stronger than word-related memory effects.

That said, it must be borne in mind that all these experiments have tested *recognition memory* whereas the use of cueing images requires the use of *recall memory*. Recognition relies on the person identifying a previously-seen picture, usually from a group of pictures. Recall requires the person to re-generate the name of a previously-seen picture. There is some evidence that people recall picture names for a long time. Cave [14] found that a single exposure to a picture could be detected even after 48 weeks by examining naming response times at subsequent exposure to the image.

3.3 Summary

Two characteristics images need to exhibit in order to use them for cueing have been identified: *ambiguity* and *efficacy*. In order to satisfy the first requirement a number of abstract image classes were tested. A number of images from each of of these image classes were used in order to determine efficacy of the class, by analysing and testing the following:

1. *Descriptiveness* — to what extent is it possible for people to assign a name to the image?

2. *Strength* — measured in terms of length of the description, the character distribution of the responses, and the entropy of the description. This is reflected by *Low name agreement* and *high image variability*, which tests whether different people provide the same names for the image or whether descriptions differ.

3. *Memorability & Durability* — How durable are the image text associations? Memorability is directly related to *high image agreement* — a stronger single mental image will lead to higher likelihood of the user remembering the image description.

4. TESTING DIFFERENT IMAGE TYPES

The most suitable images for testing are those that exhibit the required level of vagueness in terms of the Gestalt laws [15] discussed in Section 3.1.

As explained in Section 3.3, we require images that have *low name agreement*, *high image agreement*, are *visually complex* and those for which it is possible to come up with a memorable textual description. Our images are shown in the Appendix. The relationship between the image classes and the Gestalt laws is shown in Table 1.

Faces — Humans are famously good at remembering faces. [16] Whereas memorability is clearly not an issue, durability might well be. Chance and Goldstein [17] conducted an experiment to determine whether previously assigned verbal labels would be recalled after a time lapse. They found performance in recalling verbal labels to be weak and unreliable with only 35% of verbal labels being recalled correctly. However, despite this faces were included to see whether their finding was replicated.
Fractals — Singh [18] quotes Works as saying that fractals are appealing to humans due to their innate aestheticism.
Inkblots — Stubblefield and Simon [1] used Rorschach-type inkblots [19], and gained good preliminary results.
Snowflakes — Snowflakes were used by Goldstein and Chance [20] as part of a larger experiment measuring recognition ability but no work has been performed to study users' descriptions of these images.
Textures — The Texture image type was chosen because of their intrinsic variety: smooth or rough, coarse or fine as well as having regular or irregular patterns.

A web-based application presented participants with the images shown in the Appendix and elicited a textual association for each. Users could skip images. The results of the subsequent analysis is reported in the following section.

5. RESULTS

The user could choose not to provide an association for an image and this was taken as an indication that the image was too difficult to describe. This serves as an implicit subjective measure related to the ease of forming an association. In this section we refer to *image n* where *n* is the image presented in the Appendix.

5.1 Descriptiveness

Each of the 49 users was presented with 30 images, 6 of each image class, and prompted to enter a description. We gathered 1355 non-null responses (Faces: 278, Fractals: 272, Inkblots: 270, Snowflakes: 257, Textures: 278). The textual descriptions assigned to image 14 give a good example of the range of responses we obtained: *butteroad*, *demented frog*, *mangled butterfly* and *angry clown*, among others.

We found that there was a statistical difference in the number of responses we received from users based on the image class, $F(1.83, 535.4)=15.53$, $p < 0.05$. The snowflake class had a significantly lower rate of responses. The face and texture classes had higher response rates. There were *no* statistical differences within the image classes for any particular images within their class, $p > 0.05$. The face and texture images were the easiest to form associations with and snowflakes the most difficult. The analysis considered only non-null responses.

5.2 Strength

A long description will not necessarily act as a strong password; one needs to consider the entropy of the description and the variability of the responses.

This section therefore considers the responses in terms of *strength from length* (response length), *image variability* (character distribution and informational entropy) and *name agreement* (predictability).

5.2.1 Response Length

This is a useful simple indicator of security. The results of this analysis are presented in Figure 1 and can be summarised as Faces (M=16.6, SE=0.83), Fractals (M=18.3, SE=1.1), Inkblots (M=18.3, SE=1.0), Snowflakes (M=15.3, SE=0.7) and Textures (M=12.1, SE=0.5)

Image Type	Closure	Continuity	Proximity	Similarity	Symmetry
Faces	√	√			√
Fractals		√		√	
Inkblot	√	√	√	√	√
Snowflakes	√	√	√	√	√
Textures	√		√	√	√

Table 1: Image Classes & the Gestalt Laws

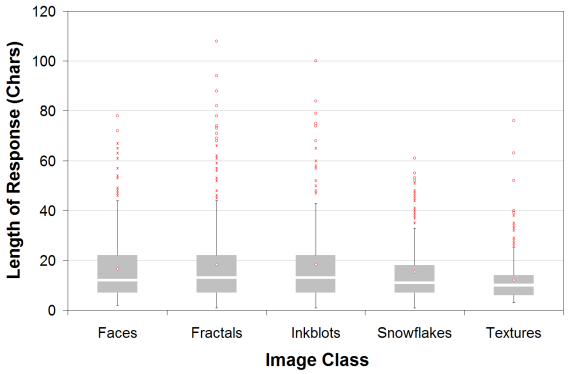

Figure 1: Response Length By Image Class

The response length is significantly affected by the image class, F(3.6, 1000.5)=12.414, $p < 0.05$. The length of the response is significantly shorter for textures than for all other image classes. An extremely simple snowflake with few "rays" had significantly lower response character length (M=13.02, SE=1.83) than a comparable snowflake with many rays (M=19.1, SE=1.93) indicating that overly simple images may result in simple responses, F=(4.16,99.96)=2.85 $p < 0.05$. The type of image shown to the user is important as different image classes can encourage users to provide longer responses.

5.2.2 Image Variability

Character Distribution

The character distribution of the response gives an idea of how predictable the responses are. We discovered that the responses closely parallel that of English, unsurprising as all the participants were English speakers.

Informational Entropy

The informational entropy of responses gives an indication of the image variability of the image. The entropy of the information in a signal, as defined by Shannon[21], specifies how much uncertainty or "randomness" exists within the signal. Specifically

$$H(X) = -\sum_{i=1}^{n} p(X_i) log_2 p(X_i)$$

where $H(X)$ is the entropy of the signal X in bits, X_i is a token in the alphabet of X represented by $1..n$ and $p(X_i)$ is the probability function representing the probability that the token will appear in the signal. The probability function used in this case is a simple weight based on the character frequencies within the textual association. As entropy within the signal increases it becomes less predictable and, as such, the more difficult it becomes to guess the content. Here we represent the entropy of a textual response by the average number of bits required to encode each character using an encoded string. For comparison; a standard ASCII keyboard has 95 printable characters (including the space character), this results in an upper bound on textual entropy of 6.57 bits per character. The lower bound for entropy is clearly 0 bits per character for a string composed entirely of a single character; since the next character in the string is always predictable. This entropic view of textual passwords essentially measures the extent of the usage of the available character set.

The results of entropic analysis of the responses are summarised as follows; Faces (M=3.1,SE=0.01), Fractals (M=3.1,SE=0.02), Inkblots (M=3.1,SE=0.02), Snowflakes (M=3,SE=0.01) and Textures (M=2.9,SE=0.01).

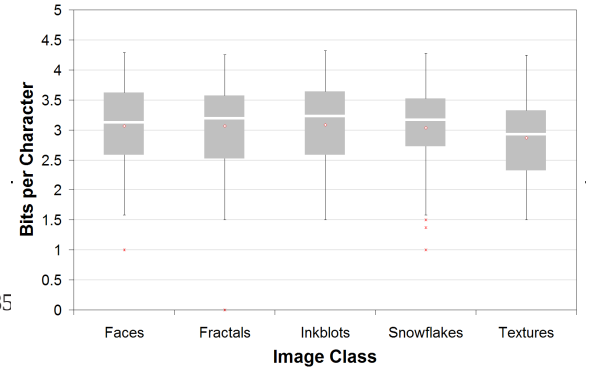

Figure 2: Bits per Character

Figure 2 shows that the texture image class is the only class with a significant difference in the number of bits per character, F=(4,1108)=5.49, $p < 0.05$. The snowflake with the highest number of rays had a significantly higher number of bits per character (Image 22: M = 3.28, SE = 0.04) than three other snowflakes (Image 20: M = 2.93, SE = 0.002),(Image 21: M = 2.89, SE = 0.05),(Image 24: M=2.93, SE=0.05), F(4.11,197.285)=4.083, $p < 0.05$. There were significant effects between individual images within the textures class which was caused by a single image, of a leaf (image 28), which had a significantly lower number of bits per character (M = 2.62, SE = 0.07). There were two textures with high amounts of repetition which had higher than average bits per character for the textures image class (Image 29: M=3.00, SE=0.06),(Image 30: M=3.08, SE=0.04), F(3.753,180.12)=5.508, $p < 0.05$.

The number of bits per character is essentially the same for most image classes (except textures) and the length of the response is the largest contributor to the number of bits per response (ie. total entropy) and therefore the overall security of a particular textual association.

5.2.3 Name Agreement

We can measure name agreement using the Smith-Waterman[22]

algorithm to measure local optimal alignments between strings. These alignments correspond to local similarities between strings and are a useful measure to locate instances where the strings have similar sections — thus measuring the similarity between two strings. A heuristic approach was used to determine a normalised score for each class of images normalised by response length.

The analysis shows that the Smith-Waterman score is significantly affected by the image class, $F(1.79,496)=1487$, $p < 0.05$. The results show that inkblots have the lowest average Smith-Waterman score (least similarity) followed by fractals, snowflakes, faces and finally textures.

Analysis of the Smith-Waterman scores for individual images reveals that within the face class any images with distinctive features (such as images 1 and 4) score higher (more similar) Smith-Waterman scores as these features are more readily commented upon within the user's response, $F(1.905,91.45)=36.567$, $p < 0.05$.

Two highly symmetrical fractals (images 7 and 11) had significantly higher Smith-Waterman scores than the other fractals, $F(1.725,85.62)=23.66$, $p < 0.05$. There was a third symmetrical image within the fractal class (image 12) that did not score similarly so the reasons for these particularly high scores is unknown. There was a significant decrease in Smith-Waterman scores for the inkblots with high density distributions of blots (images 13, 14 and 17) as compared to the more evenly distributed inkblots (images 13, 14 and 17), $F(1.74,83.52)=42.196$, $p < 0.05$. The snowflake class exhibited significant differences in the Smith-Waterman scores, $F(1.53,73.565)=66.757$, $p < 0.05$. The two least complex snowflakes (images 19 and 21) had significantly higher (more similar) scores than all other snowflakes followed by the two most complex images (images 20 and 24). Interestingly the lowest Smith-Waterman scores were for images 22 and 23 which were generated using either the maximum number of rays or the maximum complexity but not both. The images within the texture class were also found to have significantly higher Smith-Waterman scores for easily identifiable textures (images 28 and 29), $F(1.268,60.878)=193.984$, $p < 0.05$.

In conclusion, the inkblot images scored best in terms of having a low name agreement, followed by snowflakes while textures had the highest level of name agreement.

5.3 Discussion

The Inkblot and Fractal classes are particularly good performers for all metrics while Texture and Snowflake classes perform poorly (except for name agreement for the latter)

The bits per character for each image class was essentially the same — indicating that response length was the primary factor when determining the security of the image description. Hence for the majority of our experiments there was no appreciable difference between *individual* images within the image classes, whereas there were many differences across class boundaries. The experiment which assessed durability of descriptions is reported in the next section.

Our experiment indicated that the elicited responses are sufficiently secure to provide a viable cueing mechanism. Inkblots and fractals have the potential to serve as password cues. This answers our first question: inkblots are indeed suitable cueing images. The next question: "will users use them?" can only be answered by means of a longitudinal experiment.

6. INKBLOT AUTHENTICATION

A website was developed for an elective module which gave students access to lecture notes, grades and other resources. A total of 53 undergraduate students used the website. Users were randomly assigned to the password or inkblot conditions. The inkblot-assisted authentication process had the following phases:

1. *Registration*: users were given a user name and registration code, by email, to facilitate the registration process. The system either required them to choose a password or displayed a inkblot, and allowed the user to customise and tailor the inkblot, as illustrated in Figure 3. The user was then instructed to give an inkblot description as a password. The inkblots were comprised of 5 elements: (i) a randomly selected seed, (ii) the maximum diameter of blots on the canvas, (iii) the number of blots on the canvas, (iv) the distance between blots and (v) the number of colours in the inkblot. When the user is happy with their choice of inkblot the system simply saves these 5 parameters which can be used during authentication to regenerate the inkblot. The users were permitted to tailor the inkblot so as to ensure that they were not presented with an inkblot that they found it impossible to create a textual association for. If they were presented with a inkblot they considered to be obscure, they could either request a brand new one or tailor that one until they felt they were able to create an association.

2. *Authentication*: The users entered a user name and were directed to the authentication page. In the case of password users a simple password text entry area was supplied. In the case of inkblot users "their" inkblot was displayed and the user could re-enter the original inkblot description.

3. *Replacement*: users could request a re-registration from the website administrator by email if the password had been forgotten.

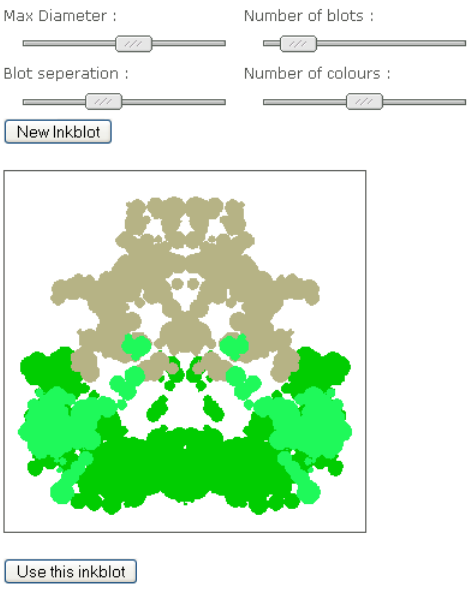

Figure 3: ***Choosing a Inkblot at Enrolment***

The experiment ran for 9 weeks and all accesses were logged.

6.1 Results

A total of 53 users used the site. Of these, 24 were allocated to the password condition and 29 to the inkblots. One user from the password condition needed a password reset during the course

of the experiment and both the original and replacement passwords are included in the analysis; no inkblot users requested a replacement password.

We encountered six instances of people who deviated from the instructions provided for their condition. Two password users used the registration code as their password. Four people chose to ignore their inkblot, instead providing a password or pass-phrase of their own choosing. These passwords/descriptions were retained throughout our analysis. Examples of descriptive passwords given by inkblot users are: *scarypumpkin*, *bunnysplat*, *blob*, *somethin* and *mask*.

The next two sections will consider the findings related to inkblot-assisted authentication in terms security and ease of use.

6.2 Security

When discussing the security of an authentication scheme based on textual input the first measure considered is typically the *length* of the password and its *character complexity*. That is to say: longer passwords with larger choices of available characters (i.e. lowercase and uppercase letters, numbers and special characters instead of just lowercase letters) will result in much more secure passwords.

Password Length

There was no significant difference between passwords (M: 7.52, SE: 0.332) and inkblots (M: 8.31, SE: 0.632), $p > 0.05$. Similarly when we evaluate the mean number of bits required per character for passwords (M: 2.49, SE: 0.09) and inkblot descriptions (M: 2.64, SE: 0.11) we find that this, too, is not significant, $p > 0.05$.

Password Guessability

We also have to consider how similar descriptions are to each other and to what extent they have similar substrings.

We used the Smith-Waterman algorithm[22], which is designed to do local sequence alignment. This allows us to measure the longest common sequences between strings (i.e. common uses of words such as " the "), in this case a higher score indicates a longer sequence and thus a *lower* score is desirable. We found that inkblots (M: 0.08, SE: 0.005) had a significantly higher Smith-Waterman score than passwords (M: 0.05, SE: 0.006), t(48.85) = -4.088, $p < 0.05$, which indicates that users often include a larger subset of common words within their inkblot descriptions than with traditional passwords.

The next section analyses the users' performance at using inkblot-assisted authentication in the context of time and effort required as well as login success rates.

6.3 Ease of Use

In this section the results gathered from the experiment are reported to give an indication of the user's experience of using the inkblot system as compared to the traditional password system.

Registration

Users were were sent a registration code by email, which allowed them to access the site. In our experimental system the password condition was a simple password entry prompt in the traditional style (users were asked to enter the password twice to confirm its correctness). By comparison, since users were allowed to design their own inkblots an inkblot designer was implemented as part of the registration process. This resulted in users spending considerable time designing their inkblot, inflating the registration time (seconds) for the inkblot condition (M: 256.03, SE: 71.364) so that it was much more time-consuming than password registration time (M: 44.88, SE: 8.68), t(52) = -2.729, $p < 0.05$. This can be viewed as a positive or negative effect depending on the reader's point of view. It clearly makes the registration more interactive, which is a good thing, and is likely to lead to more memorable passwords, but it does significantly increase registration time.

Authentication

The mean time (seconds) required to login for successful sessions was measured from user name entry until the login session was completed and may also include more than one login attempt if they were unsuccessful at first. We found that there was a significant difference between inkblots (M: 13.08, SE:0.532) and passwords (M: 11.15, SE:0.469), t(774) = -2.724, $p < 0.05$. This value includes any additional time it would have taken for the user's browser to download and display the image representing the inkblot (generally less than 3KB in PNG format).

During the course of the experiment there were a total of 388 login sessions for inkblot users and 412 login sessions for password users. Of these there were a significantly lower number of sessions with a login failure for the inkblot condition (23) than for the password condition (44), $p < 0.05$. This puts the mean number of failed sessions for passwords at 11% and inkblots at 6%.

However, when we look at the failed sessions in more detail we discover that within a login attempt session the average number of attempts at getting the password correct per session is somewhat different. We found that there was an average of 1.18 attempts (SE: 0.118) per session for a password while inkblots required an average of 1.96 attempts (SE: 0.493). Our results indicate that there was borderline significance (t(65) = -1.985, $p = 0.051$) which may warrant further investigation. Thus inkblot users are more likely to get it right first time but may make more attempts to login if they fail the first time.

We also considered the number of sessions which were regarded as "total failures" ie. sessions within which there was a failed login attempt (or a sequence of failed login attempts) but no eventual success indicating that the user gave up. We found that there was no significant difference in this respect (3 failed inkblot sessions, 2 failed password sessions, $p > 0.05$).

6.4 Are Inkblots Efficacious Password Cues?

Efficacy metrics, as outlined in Section 3.3, are *descriptiveness*, *strength* and *durability*. In terms of descriptiveness and strength, these results appear to conflict with the results of our previous experiment. The length of response decreased significantly once users were asked to perform this task within a live authentication system, and this impacts the strength of the password. Furthermore, provided textual descriptions were shorter and less descriptive, and, indeed, some appeared to have nothing to do with the provided cue, so the inkblot fails the descriptiveness test as well. This result strengthens findings by Brostoff *et al.* [23] during evaluation of the Passfaces[1] authentication mechanism where usage of an authentication system in real life differed significantly from lab-based experimentation.

This experiment shows that when the user knows that the inkblot description is going to be used frequently as a password, he or she provides a much shorter description than would be provided if the description was only going to be provided once or twice in a lab-based experiment. This is perfectly reasonable, because users emphasise convenience over security. Hence the length of response and bits per char are basically the same as passwords. This is rather disappointing since it was hoped that the presence of the inkblot would allay users' fears of forgetting their passwords and therefore encourage them to choose longer (and stronger) passwords, the point of the whole exercise.

Finally, as regards durability, the inkblot users *did* appear to have

[1] `http://www.realuser.com`

less trouble remembering their textual descriptions, although this obviously does not apply to the four users who did not provide a inkblot description.

This confirms the findings of Dhamija and Perrig [24] that people are only willing to expend the minimum effort in managing their passwords. Our results indicate that the descriptions offered by users are of comparable length and complexity to traditional passwords but with the problem that they will tend to include common stop-words in their description which weakens the password. This answers the second question posed in the introduction: users *do not* utilise the cues, preferring to rely on their own memory capabilities.

7. CONCLUSION

We investigated Stubblefield and Simon's proposal that passwords could be cued by using inkblot-like images. Inkblot-type abstract images did indeed elicit the longest and strongest textual descriptions and appeared to be suitable.

However, the final experiment shows that the inkblots were *not* used by the users. It did not appear to encourage them to strengthen their passwords and they did not exploit the true potential of their inkblot in coming up with a textual description thereof, probably because users anticipate the extra effort involved in continuously entering the long description at each authentication attempt with little enthusiasm.

We have to conclude that, whereas the inkblots appear theoretically viable in terms of cueing passwords, the end-user's desire for convenience and speed of access led them not to exploit the potential for cueing provided by the inkblot. Perhaps the only conclusion is that the combination of convenience-seeking users and passwords is doomed to failure. If this is the case then any auxiliary efforts to strengthen the mechanism, such as the one proposed by Stubblefield and Simon, are futile.

Acknowledgement

Tony McBryan carried out this research as part of his Masters project.

8. REFERENCES

[1] A. Stubblefield, D. Simon, Inkblot authentication, Tech. Rep. MSR-TR-2004-85, Microsoft Research (August 2004).

[2] R. J. Witty, K. Brittain, Automated password reset can cut IT service desk costs, Gartner Report (2004).

[3] M. Hertzum, Minimal-feedback hints for remembering passwords, Interactions (2006) 38–40.

[4] A. Paivio, Mental representations: A dual coding approach, Oxford University Press, Oxford, UK, 1986.

[5] P. Jacob, M. Jeannerod, Ways of Seeing. The scope and limits of visual cognition, Oxford University Press, Oxford, UK, 2003.

[6] V. Bruce, P. R. Green, Visual Perception. Physiology, Pschology and Ecology, Lawrence Erlbaum, Hove and London, 1990.

[7] X. Alario, L. Ferrand, A set of 400 pictures standardised for french: Norms for name agreement, image agreement, familiarity, visual complexity, image variability and age of acquisition, Behavior Research Methods, Instruments and Computers 31 (1999) 531–552.

[8] D. Rapaport, Diagnostic Psychological Testing, Vol. 2, Year Book, Chicago, 1946.

[9] J. M. Gold, The role of verbalization in the rorschach response process: A review, Journal of Personality Assessment 51 (4) (1987) 489–505.

[10] D. B. Mitchell, A. S. Brown, Persistent repetition priming in picture naming and its dissociation from recognition memory, Journal of Experimental Psychology 14 (2) (1988) 213–222.

[11] S. Madigan, Picture memory, in: J. Yuille (Ed.), Imagery, memory, and cognition: essays in honor of Allan Paivio, Lawrence Erlbaum Associates, Hillsdale, NJ, 1983, pp. 65–89.

[12] A. Paivio, Imagery and verbal processes, Holt, Rinheart & Winston, New York, 1971.

[13] A. Paivio, T. Rogers, P. Smythe, Why are pictures easier to recall than words?, Psychonomic Science 11 (4) (1968) 137–138.

[14] C. B. Cave, Very long-lasting priming in picture naming, Psychological Science 8 (4) (1997) 322–325.

[15] M. Wertheimer, D. Beardslee, Principles of perceptual organization, Readings in Perception (1958) 115–135.

[16] N. Kanwisher, J. McDermott, M. M. Chun, The fusiform face area: A module in the human extrastriate cortex specialized for face perception, Journal of Neuroscience 17 (11) (1997) 4302–4311.

[17] J. Chance, A. G. Goldstein, Recognition of faces and verbal labels, Bulletin of the Psychonomic Society 7 (4) (1976) 384–6.

[18] G. Singh, Stringing the fractals, IEEE Computer Graphics & Applications 25 (5) (2005) 4–5.

[19] H. Rorschach, Psychodiagnostics, Grune & Stratton, New York, 1942, eds: P Lemkau and B Kronenberg, Original published 1921.

[20] A. G. Goldstein, J. E. Chance, Visual recognition memory for complex configurations, Perception and Psychophysics 9 (2-B) (1970) 237–241.

[21] C. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (1948) 379–423 and 623–656.

[22] T. F. Smith, M. S. Waterman, Identification of Common Molecular Subsequences, Journal of Molecular Biology 147 (1981) 195–197.

[23] S. Brostoff, M. A. Sasse, Are passfaces more usable than passwords?: A field trial investigation, in: Proceedings of HCI 2000, 2000, pp. 405–424.

[24] R. Dhamija, A. Perrig, Déjà vu: A user study using images for authentication, in: Proceedings of USENIX Security Symposium, Denver, Colorado, 2000, pp. 45–58.

[25] L. Spacek, Computer Vision Research Face Database (2006). URL `http://cswww.essex.ac.uk/mv/allfaces/index.html`

[26] F. Slijkerman, Ultra Fractal 4.02 (Standard Edition) (2005). URL `http://www.ultrafractal.com/`

[27] A.I.Studio, Fractal Snowflake Generator (2004). URL `http://a-i-studio.com/snowflake/`

[28] K. J. Dana, B. V. Ginneken, S. K. Nayar, J. J. Koenderink, Columbia-Utrecht Reflectance and Texture Database (2005). URL `http://www1.cs.columbia.edu/CAVE/curet/`

APPENDIX

Faces: Collected from the Essex University Computer Vision Facial Databases[25] "face94" and "face95" and were chosen to represent an equal mix of male and female faces with a range of physical features. Only images that were clearly visible with similar scale and without distracting backgrounds were considered.
Fractals: generated using a commercial program Ultra Fractal[26].

Variations within the image class were obtained by changing the algorithm used to generate the fractal in addition to varying the viewing position and colouring algorithms.

Inkblots The inkblots were generated by a custom PHP script. The inkblots were built by dropping "blots" onto a canvas and ensuring the next blot landed within a fixed area of the previous blot. The canvas was then mirrored to create the final inkblot. The images were varied by changing the values of variables which control the number of blots, blot diameter, colour and distance between the blots.

Snowflakes: generated using A.I. Studio Snowflake Generator[27] and variations within the images were achieved primarily by varying the number and complexity of the rays along with scaling and position details.

Textures: obtained from the CUReT[28] texture database and were chosen to represent a range of different textures including both man-made and natural textures.

Figure 4: Image 1

Figure 5: Image 2

Figure 6: Image 3

Figure 7: Image 4

Figure 8: Image 5

Figure 9: Image 6

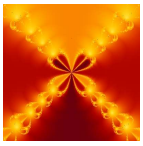

Figure 10: Image 7

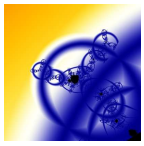

Figure 11: Image 8

Figure 12: Image 9

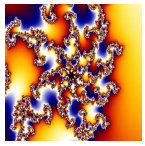

Figure 13: Image 10

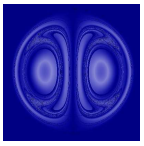

Figure 14: Image 11

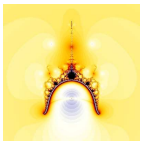

Figure 15: Image 12

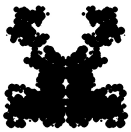

Figure 16: Image 13

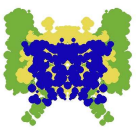

Figure 17: Image 14

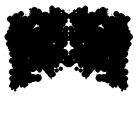

Figure 18: Image 15

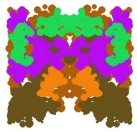

Figure 19: Image 16

Figure 20: Image 17

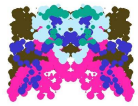

Figure 21: Image 18

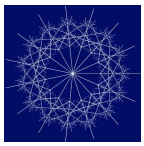

Figure 22: Image 19

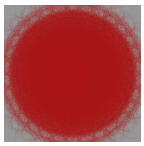

Figure 23: Image 20

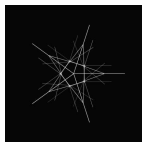

Figure 24: Image 21

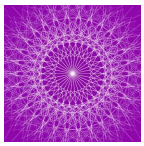

Figure 25: Image 22

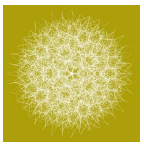

Figure 26: Image 23

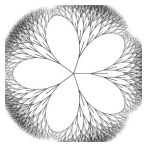

Figure 27: Image 24

Figure 28: Image 25

Figure 29: Image 26

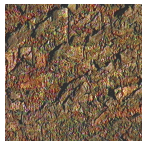

Figure 30: Image 27

Figure 31: Image 28

Figure 32: Image 29

Figure 33: Image 30