

Securing Passfaces for Description

Paul Dunphy
School of Computing Science
Newcastle University
Newcastle UK
p.m.dunphy@ncl.ac.uk

James Nicholson
School of Computing Science
Newcastle University
Newcastle, UK

Patrick Olivier
School of Computing Science
Newcastle University
Newcastle, UK
p.l.olivier@ncl.ac.uk

ABSTRACT

One common practice in relation to alphanumeric passwords is to write them down or share them with a trusted friend or colleague. Graphical password schemes often claim the advantage that they are significantly more secure with respect to both verbal disclosure and writing down. We investigated the reality of this claim in relation to the *Passfaces* graphical password scheme. By collecting a corpus of naturalistic descriptions of a set of 45 faces, we explored participants' ability to associate descriptions with faces across three conditions in which the decoy faces were selected: (1) at random; (2) on the basis of their visual similarity to the target face; and (3) on the basis of the similarity of the verbal descriptions of the decoy faces to the target face. Participants were found to perform significantly worse when presented with visual and verbally grouped decoys, suggesting that Passfaces can be further secured for description. Subtle differences in both the nature of male and female descriptions, and male and female performance were also observed.

Categories and Subject Descriptors

H.1.2 [User/Machine Systems]; D.4.6 [Security and Protection]

General Terms

Security, Human Factors

Keywords

Graphical Passwords, Passfaces, Description

1. INTRODUCTION

The fundamental requirement of knowledge-based authentication is that knowledge of the authentication secret must stay private. Conventional advice aiming to reduce the likelihood of unauthorised access urges users to select distinct and hard to guess passwords for each separate account that they

own. User compliance, requires a feat of human memory that flies in the face of the limits of human cognition. Techniques users can adopt for memorising these seemingly unmemorable passwords include mnemonics [29] and passphrases [16], while alternative approaches to passwords include cognitive [30] and associative passwords [23]. Technical solutions to the problem also exist in the form of *Single Sign-on (SSO)* password managers [4]. Despite this, in many security-sensitive contexts it has been found that for reasons of both perceived convenience and necessity, users often write down their alphanumeric passwords and share them with others.

1.1 Password practices

The fact that a typical user has a significant number of passwords to remember, and copes with this by recording passwords externally, has been demonstrated in a number of studies. Adams and Sasse [1] suggest that 4-5 regularly used passwords per user should be considered a maximum; one study [22] estimates the number of passwords per user to be 6 while a survey of BT employees [20] suggests this number is as high as 16.

Adams and Sasse also report that at least 50% of participants in a survey of 139 wrote down passwords as a result of the difficulty of remembering multiple passwords and responding to password expiration policies. More resoundingly two-thirds of respondents in a survey of 300 corporate users wrote down work-related passwords [21]. The same study also uncovered that 66% of workers admitting giving their password to a colleague, and 74% knew one or more passwords belonging to a co-worker. To a much lesser extent Zviran and Haga [31] report 7% of users successfully recalling self-generated passwords in their user study wrote them down to aid memorability, despite being instructed to the contrary. This figure increased to 9% for recalling passphrases.

Although communicating personal passwords to others in a corporate setting can seldom be considered secure, recent discussions as to the risks of writing down and sharing passwords have been moderated by an awareness of the need to take into account the realities of different work contexts (e.g. that different settings pose different levels of threat and risk). In an appropriate setting, and when carried out responsibly, writing down and even sharing passwords can be a practical and effective solution to the password management problem (e.g. providing a readily accessible back-up if a user forgets a password).

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.

Symposium on Usable Privacy and Security (SOUPS) 2008, July 23–25, 2008, Pittsburgh, PA USA

1.2 Password description

Graphical passwords are thought to be particularly resistant to being written down or verbally communicated. That users will find it difficult to share graphical secrets is a widely held assumption, and an implicit (though untested) design feature of most graphical password schemes. That graphical passwords are more memorable than their alphanumeric counterparts means that users are less likely to need to record their secrets as a memory aid. However, were graphical passwords to become widely used, users would be presented with exactly the same password management problem as for alphanumeric passwords. In such situations, users would be likely to attempt to adopt the same strategies as for alphanumeric schemes, that is, record the password externally or communicate it to a trusted friend or colleague.

In relation to graphical passwords, the nature of users' descriptions of graphical secrets, and the vulnerability of password schemes to *description*, have not been previously explored. In this paper we evaluate description in relation to the Passfaces graphical password scheme. The work comprises three steps, firstly we collect sets of naturalistic descriptions of faces. For these faces we analyse both the similarity of the verbal descriptions, and users' judgements as to their visual similarity. We then evaluate people's ability to authenticate themselves using a verbal description of a Passfaces password and demonstrate that by systematic organisation of the decoy faces, Passfaces can be made considerably less vulnerable to description.

Unlike alphanumeric passwords, most secrets for graphical password schemes can neither be precisely written down using a static media (e.g. pen and paper) nor verbally communicated. Users therefore have to revert to the production of a written or spoken *description* of their secret. Though our interest here is in graphical passwords, we define *password description* as any non-digital attempt to record or communicate a password, using either an external representation or verbal/non-verbal means. This encompasses sketches, written and spoken descriptions and instructions, and even accompanying physical gestures.

Technologies to facilitate the sharing of graphical passwords clearly exist; given the ubiquity of camera phones, these provide an obvious and familiar way to record images in recognition-based schemes. However we believe that despite the onset of digital devices providing connectivity anywhere, anytime; there will still be occasions when description is the only means of sharing a graphical password. Description itself is a phenomenon of some interest given the widely held assumptions related to Passfaces and other graphical schemes.

One claimed advantage of the Passfaces scheme over conventional alphanumeric passwords is that Passfaces "can't be written down or copied" and "can't be given to another person" ([18], p. 3). A number of subtle configurations have been adopted to mitigate against the risk of easy description, such as that "the grids of faces in Passfaces are grouped by sex and are selected to be equally distinctive so that Passfaces cannot be described by gender or obvious characteristics. None of the faces stand out from the others" ([18], p. 5) and that 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" ([18], p. 4).

The rest of the paper is organised as follows: firstly we discuss *password description* then give a short review of related

work in the field of graphical passwords. Then we introduce the design of our user study and the two phases involved: (1) description collection and (2) login. We conclude with a discussion of the results of the study and some closing remarks.

2. GRAPHICAL PASSWORDS

The term *graphical password* was introduced by Blonder [2] in a click-based scheme in which the user was required to choose a sequence of pre-defined click regions in an image. Passpoints [28] extended Blonder's original proposal by allowing users to select any location in the image as a click-point bounded by a tolerance region. Having user defined click-points significantly increases the password space. Passpoints continues to receive considerable attention from researchers, most recently in relation to vulnerabilities arising from structural aspects of visual perception. For example, that user selection of click-points is biased by so-called *hotspots*. Incorporating such insights within a graphical attack dictionary has been performed using image processing techniques to acquire salient points [10], and using corpus-based approaches by harvesting click-points from a small population of users [25].

Passfaces [19] is a commercial graphical password scheme, for which the vulnerability and usability has been relatively under explored. Brostoff and Sasse [3] conducted an evaluation of the system in a field study that found that participants using Passfaces experienced one third of the login errors of alphanumeric password users, despite accessing the system less frequently. Other unpublished evaluations conducted by Valentine [27, 26] discovered that users had low expectations that they would remember passfaces over a long period of time. Amongst other issues, participants felt that due to the apparent difficulty of writing down their passfaces they had no strategy in relation to forgetting (i.e. they had a commonsense awareness of the difficulty of writing down their passfaces). Valentine found that users' memory for passfaces was significantly better than for conventional passwords.

A number of interesting trends as to users' choices of images emerged in a study of the Face system [6] in which users are allowed to choose the faces that made up their graphical password. Users were found to consistently choose attractive faces of a different sex, but of the same racial group, as themselves. Male users exhibited this behavior to a greater degree, although it should be noted that such sex differences are notoriously subject to large individual differences.

Other applications of faces in security and privacy systems have emerged in recent years. The creator of Passfaces has patented a method of using faces to facilitate PIN recall by creating an intuitive mapping between faces and PIN digits [5]. A small study has suggested that this FacePIN has potential to be a usable solution to the PIN memorability problem [12]. Similarly, Facemail [17] employs faces to visualize recipients of email for the sender, aiming to avoid embarrassing mistakes of sending personal communications to mailing lists or the wrong recipient. Xu, Zhu and Owen [24] present a recent survey of graphical password schemes.

3. USER STUDY

We designed a user study that simulated a Passfaces login, comprising 5 face grids in a 3×3 formation. A total of

Passfaces Experiment

Click the face to which you think the audio description refers



Figure 1: The Passfaces challenge for the description study. Participants were required to select the face to which an audio description refers. Participants were free to use the on-screen controls to replay audio descriptions.

45 faces were collected from the Passfaces online demo (27 female and 18 male), enough for 3 grids of female faces and 2 grids of male faces. Rather than assigning Passfaces to participants in the usual way, at the presentation of each of the 5 grids an audio description of one of the faces (the *target face*) was played. Participants were required to click the face which they believed was being described. So to ‘login’ successfully a participant had to match each of the 5 spoken descriptions with the corresponding face in the grid. In the experiment our goal was to explore the scenario of the participant receiving a description of a sequence of passfaces to login to a resource on behalf of a colleague. Our objective was to study three aspects of Passfaces in relation to *description*:

1. to quantify Passfaces’ vulnerability to description;
2. to evaluate approaches to reduce its vulnerability to description through the choice of decoy faces;
3. to explore whether there are any significant sex differences in relation to creating and interpreting descriptions of Passfaces.

Our attempts to decrease the vulnerability of Passfaces to description centred around the procedure by which decoys were selected for a given target face. Though the exact procedure by which decoy selection is undertaken in the commercial Passfaces system is not known to us, we explored three conditions, each corresponding to a different decoy selection procedure: (1) random selection (of faces of the same sex); (2) selection based on independent judgments of the visual similarity of the decoys to the target face; and (3) selection based on the similarity of the verbal descriptions of the faces to the target face.

In the description collection phase (section 3.1) we discovered a significant difference in the duration of the descriptions between males and females, an effect suggested by the

psycholinguistic literature. Discussions of cognitive gender differences are the subject of ongoing controversy, however, there are three widely accepted differences: linguistic ability, visuo-spatial ability, and perceptual-motor ability [13]. The term linguistic ability encompasses but is not limited to: verbal fluency, grammar, writing, and vocabulary. In relation to password description, differences in terms of linguistic ability are particularly relevant. These appear early (developmentally) though it is not clear if these differences are maintained into old age [13].

A study by Huttenlocher et al. [14] revealed that on average there is a 16 word difference in vocabulary between females and males at 13 months, increasing to 51 words at 20 months and 115 at 24 months. Horgan [11] discovered that females between 2 and 4 years old used longer utterances at a younger age than males, also showing more linguistic maturity by demonstrating ability with the passive voice, participles and adjectives. Halpern [13] presents a summary of the complex yet compelling topic of gender differences in cognitive ability.

Such results from gender difference research suggest that females could perform better at articulating descriptions of their Passfaces to a friend. As well as this expectation, it is interesting to consider the effectiveness of females when using descriptions from male counterparts and vice-versa. One of the key ingredients of effective communication (and thereby description) is a shared and subtle understanding of any vocabulary used, yet potentially males and females are subject to quite different cultural influences.

3.1 Description collection procedure

The first stage of the user study was to collect descriptions for a bank of 45 faces. We recruited 18 contributors, 9 male and 9 female, each of whom recorded verbal descriptions of 15 random faces, providing us with 6 descriptions per face, 3 male and 3 female. The average age of contributors was 26 (standard deviation of 8) and as shown in figure 2 there was some variation in their background. Although computer scientists dominated, there was an approximately even split between computing science and mathematics on one hand, and arts and humanities on the other.

The recordings took place in a recording studio at Newcastle University; a quiet environment free of distractions. Contributors were given a digital recorder and a set of faces to describe. Faces were presented on a sheet of paper with each face sized 3cm × 4cm. Contributors were asked to examine each face and record a description in their own time, that is, using as much time as they required both in preparation for and during the recording. No advice was given on how to approach the descriptions, other than that contributors were asked to imagine the scenario that they were describing faces to a friend.

Contributors were left alone in the studio and afterwards the recordings were assessed in terms of the sound quality. In two cases it was deemed that the quality of the recordings were not high enough to be used in the subsequent study (due to not holding the microphone close enough to the mouth) and two new contributors were recruited. The classification in figure 2 includes only contributors whose descriptions were actually used. We allowed elements of hesitation in the recording, although no examples of this were found to be detrimental to the quality of the description.

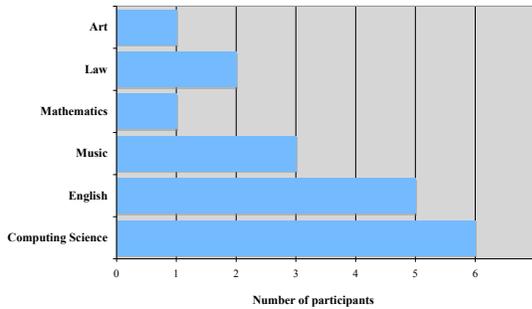


Figure 2: The classification of backgrounds of contributors to the description phase. Contributors were recruited opportunistically and comprised university undergraduate and postgraduate students or graduate level employees. Each contributor was classified according the subject of their highest level qualification.

3.2 Approaches to description

The most readily apparent strategy used by contributors was to first describe a distinguishing feature, either relating to the person (e.g. race) or their face (e.g. hair colour or length, nose or ear size and shape). Where contributors did not detect any such features they would mechanically describe the face feature by feature. Where there was a sense that contributors were frustrated with their own inadequate descriptions they often used even the slight presence of clothes in addition to their description of the face itself.

Female contributors were much more likely to adopt a more holistic approach to description. By contrast male contributors often resorted to mechanical list-like feature-by-feature descriptions of faces. Females often attempted to present a richer sense of “the person behind the face” and included judgements as to whether the person was happy, likely personality traits, and even social class.

“A white Caucasian female. Uhm, mousy brown blonde highlights. Quite big ears. Big mouth. Uhm, wide jaw, big eyes.”

List-like feature-by-feature description.

“This girl looks young and slightly embarrassed, again kind of, uhm, highlighted blonde hair. She’s kind of, uhm, I don’t know, uhm, her hair is again kind of half falling out, it’s kind of up in a kind of tied back way so, uhm, she looks friendly enough again, looks like she’s probably well bred and, uhm, does lots of sport, particularly with horses.”

Holistic description incorporating facial features, personality traits and social class.

The average length of the recorded descriptions was 23 seconds although there was significant variation between contributors (standard deviation of 11 seconds). Interestingly the average recording length for females was 27 seconds

and for males was 20 seconds ($t = -5.37 p < 0.01$). Furthermore, the male contributors used 567 distinct words (from a total of 4329), and the female contributors 654 distinct words (from a total of 5560). This is broadly consistent with known results from the gender difference literature where females are known to excel at linguistic style tasks [13] and use a wider and more fine grained vocabulary. However, when describing the faces of people from different races some contributors (particularly females) were reluctant to use explicit racial descriptors.

3.3 Experimental Conditions

The three experimental conditions introduced corresponded to three different approaches towards association of a target face with its decoy faces. In this way we could evaluate any effect on the resistance of Passfaces against description:

Random groups the decoys for a target face were selected randomly for each grid (base case).

Visual groups for each face decoys are selected on the basis of visual similarities between the target face and the decoys. Visual grouping seeks to maximise the number of decoys that might match a description of the target face based on visual similarity judgments alone.

Verbal groups for each face decoys are selected on the basis of the similarity of the verbal descriptions of the faces of the target face and the decoys. Verbal grouping seeks to maximise the ambiguity relating to verbal descriptions without necessarily requiring visual similarity.



Figure 3: Random group: decoys are selected at random within the set of faces of the same sex (target face highlighted in red).

3.3.1 Random groups

The base condition was the random groups. When a face was randomly selected to be the target face, the system selected 8 random faces of the same sex to be decoys. This was repeated for each of the 5 grids and within a trial a particular face was only used once as either a target or decoy. Thus each trial used all of the 45 images in the bank of faces. Figure 3 shows an example of a randomly generated grid where the target face is highlighted in red.

In this case, examples of descriptions generated by the contributors for the target face are as follows:

“Long, red brown hair, parting in the middle. Uhm <break> happy girl.”

“Okay, female, smiling. Uhm, Caucasian I would guess. Slightly longer than shoulder length reddy brown hair. Dark reddy brown hair. Dark eyes. Uhm, not think not fat, average sort of, uhm, chubbiness. Uhm, slightly pointed chin. Slightly odd looking smile. Uhm, quite big hair, straight again, or slightly curly.”

“Sort of late 20’s, uhm, white girl with long brown hair. Looks a bit like Kate Winslet.”

Though these are all descriptions of the target face, aspects of each do describe other faces.

3.3.2 Visual groups

Decoys in the visual groups condition were selected based on their visual similarity to the target face. To gain a consensus on which faces looked similar, we recruited a second group of contributors (similarity judges), whose similarity rankings were used in the selection of 8 visually similar decoys for each target face. These volunteers were recruited in computer labs around Newcastle University campus and each was assigned 5 faces out of our bank of 45. For of these 5 faces they were asked to select as many other faces as possible that were lookalikes.

Typically contributors selected 6 or 7 though some were assigned more faces, depending on their perceived difficulty of the task and contributor responses were recorded by an experiment moderator. Upon gathering enough data we manually looked through the results for any agreement between participant choices of lookalikes so as to create a set of 8 decoys for each face. Where it was not possible to select 8 decoys for a particular face based on the similarity judgements alone (i.e. an insufficient number of faces had been identified as similar) we manually selected the short-fall based on our own judgement of similarity. At most only 2 decoys were selected at any one time using this ad hoc technique.

Figure 4 shows the visually similar grid assembled for the same target face highlighted in figure 3. Many of the elements of the sample descriptions presented in section 3.3.1 apply to these faces. As expected the visual similarity of the members of the visual group is readily apparent.

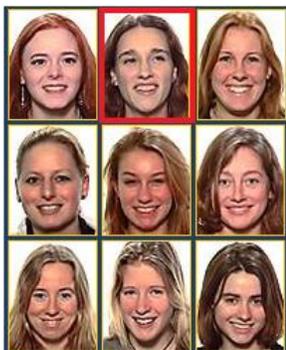


Figure 4: Visual groups: a grid of faces grouped by visual similarity (the target face is highlighted).

3.3.3 Verbal groups

The decoys in verbal groups were selected based on the similarity of the verbal descriptions to the target face. Two descriptions were deemed similar if the same features were emphasised and similar values used to describe those features. We developed a systematic way to construct these groupings based on the set of 250 descriptions. The first step was to manually transcribe each collected description, and use the *TextSTAT* concordance software to collate key terms used about each face. By looking at the common descriptive terms arising in the descriptions we identified the following features the contributors were most likely to include in their descriptions (and the set of values used):

Hair used 338 times in descriptions including the values: blond, black, dark/brown, long, tied-back, short, red, curly, fringe.

Face shape used 162 times in descriptions including the values: round, oval, small, long, pointed.

Eyebrows used 91 times in descriptions including the values: heavy/bushy, shaped, dark, groomed, thin.

Nose used 68 times in descriptions including the values: big, crooked, button/small, long, wide, thin/pointy.

Figure 6 illustrates the verbal description grouping process. In selecting the verbal group decoys for face 2, we would first make a note of all faces described as *long*. This is repeated for hair, eyebrows and nose descriptors and scores are incremented for each matching feature value (and decremented for explicit conflicts). The eight faces with the greatest tally are used as decoys for the target face. As already described in a small number of instances we had to use our own judgement to select decoys where this verbal similarity ranking did not yield 8 decoys.

Face Shape		Nose	
Value	Faces	Value	Faces
Round	10,11,12,13	Long	2,3,4,5,6,7,8
Long	2,7,8,9	Flat	22,23,24,25
Short	34,37,39	Pointed	2,23,24,27
Pointed	14,15	Wide	14,15,16

Hair Colour	
Value	Faces
Blonde	2,3,4,5,6,7,8
Brown	14,16,17
Black	33,36,39,42
Red	41,44

Figure 5: Simplified verbal grouping procedure: Example tabulation of raw data for a cross-section of possible facial features. For face 2, all face numbers occurring on the same row as face 2, for any feature, are candidate decoys. Faces that occur with face 2 most often are prioritised as decoys. In this example faces 7 and 8 are the best match.

As a continuation of figures 3.3.1 (random grouping) and 4 (visual grouping); figure 6 shows the decoys for the same

target face in the verbal grouping condition. Though a number of faces are shared between the visual and verbal groups there are a notable differences. Indeed, the verbal groups exhibits significantly more variation in appearance than the visual group.



Figure 6: Verbal groups: a grid of faces grouped by verbal similarity (the target face is highlighted).

3.4 Procedure

We recruited 56 participants for controlled trials that took place during undergraduate computing science practical sessions. Of the 56 participants, 31 were male and 25 female; none had taken part as contributors in the previous phases. The average age of participants was 22 (standard deviation = 7). For each grid, one face was chosen at random as the target face, with the decoys generated depending on the experimental condition. A particular face could not be the target more than once in the same condition.

We also ensured that each description was delivered by a randomly chosen speaker, reducing dependence on a particular speaker. In order to study sex differences, we imposed the additional constraint that participants heard descriptions from speakers that were either all male, or all female (decided at random upon starting each condition).

Participants were asked to perform all 3 conditions in an order determined randomly by our software. To provide a small incentive we offered £5 to participants who were able to match all five descriptions in any condition.

4. RESULTS

Out of 158 attempts made by participants collectively across 3 conditions only 13 (8%) of logins were successful, that is, where the participant correctly associated all 5 verbal descriptions with the target face. Table 2 shows the number of attempts made in each condition, the male/female participant split in that group, along with the number of successful login attempts. Login success was greatest for the random groups and lowest for the verbal groups. The average score (out of 5) in the random condition was 3.57 with a standard deviation of 0.91 compared to 2.87 and a standard deviation of 1.07 in the visually grouped condition $t=3.63$ $p<0.01$. The average score in the verbally grouped condition was 2.81 with a standard deviation of 1.14. The difference between the verbal groups condition and the random condition was also statistically significant: $t=3.64$ $p<0.01$

Figure 7 shows the distribution of scores across the three conditions. Scores in the random condition are very much concentrated about 3 and 4 with 40 of the 54 participants scoring in this range. This theme of scoring highly was also reflected in 8 participants achieving 5/5, higher than both the visual and verbal conditions combined.

The very close result between the visual and verbal conditions warrants further analysis. The shape of the distribution of the visual groups scores is bell shaped with scores concentrated mainly around 2/5 and 3/5. It is surprising that 5 more participants scored 4/5 in the verbal condition than the visual condition, while so few in the verbal condition scored 5/5. A possible contributing factor being that as the number of participants in each condition was not overly large, 8 fewer participants in the verbal condition over the visual condition did not allow any extremes to even out.

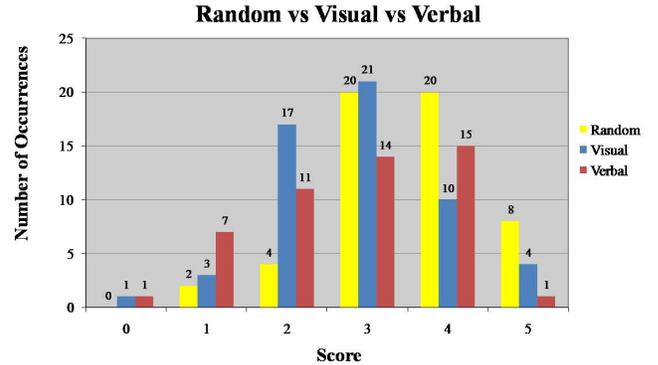


Figure 7: Random vs Visual vs Verbal groups: A breakdown of the scores achieved in each condition.

Although participants were not placed under time pressure, we measured the time taken to complete each condition. Our assumption was that participants would take the least time to complete the random condition due to the audio descriptions appearing less ambiguous. In fact the mean timings differed little between the three conditions, 155 seconds (standard deviation of 72) for the random groups, 152 seconds (standard deviation of 51) for the visual groups, and 149 seconds (standard deviation 55) for the verbal groups – there is no statistical significance in the difference.

Table 1 shows the breakdown of successful logins to display the sex of the participant and the sex of the speakers in the audio descriptions. Females using female audio descriptions were the most successful, with male listeners using female descriptions a close second. Despite the small numbers this suggests that female descriptions were most useful when completing this exercise.

Table 1: Sex combinations of describer and listener for successful login (e.g. M:M = male listener using male description).

	M:M	M:F	F:M	F:F
Random groups	1	0	1	5
Visual groups	1	3	0	1
Verbal groups	0	1	0	0

Table 2: Number of successful logins in the different experimental conditions.

Condition	Attempts	Male/Female	Successes
Random	54	29/25	8
Visual	56	31/25	4
Verbal	49	26/23	1

4.1 Descriptions

Overall participants associated audio descriptions with the correct face 62% of the time (482/780). Using male descriptions participants were correct 60% of the time (238/395) and using female descriptions this figure was 63% (244/385).

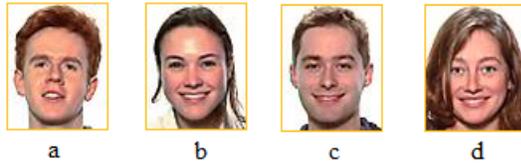


Figure 8: Example faces: users performance for different faces varied between extremes.

Some descriptions were clearly more effective than others. For example, consider the faces shown in figure 8 (actual examples from our face bank). We expected face (a) to be easily distinguished as only two faces out of the 18 male faces had red hair. Indeed, examples of descriptions for face (a) include:

Description 1: “Big red hair, large forehead.”

Description 2: “Pale skin, ginger hair, almost smirking rather than smiling.”

Description 3: “Male, ginger hair. Ginger eyebrows. Uhm, slightly curved eyebrows. Uhm, broadish nose, head slightly tilted back. Not smiling so much. Ears showing around the haircut, quite fluffy hair. Can’t see any clothing, uhm, the way the shot’s taken.”

Though relatively short these description were highly effective in supporting the identification of the correct face. Description 1 was used correctly 8 times out of the 8 it was played to participants. Description 2 and description 3 were both used correctly 3 times out of the 3 they were heard. Unlike face (a), face (b) does not have such distinctive features and as a result the task of constructing an effective verbal description is made significantly more difficult. Example descriptions for face (b) include:

Description 4: “Female with, straggly < break> dark < break> brown with slight lighter blondish < break> highlights or areas within her hair. Tied back presumably. < break> Looking up in the corner, slightly chubby face.”

Description 5: “A white female, with her hair up but some of it coming across her face like two bits down the side of her face. < break> Uhm < break> uhm < break> like a brown jumper on with a bit of a white collar.”

Description 4 was used twice with no correct responses while description 5 was used three times and also received no correct responses. Describers of faces (c) and (d) experienced similar difficulties and the worst performing descriptions in the experiment were associated with these faces. Descriptions of each of these faces were heard 4 times and resulted in no correct responses:

Description 6: (face c) “Uh, he’s got blond short hair, and his eyebrows are quite prominent, quite thick eyebrows, uhm, but also quite big eyes. Uhm he’s got quite a long face.”

Description 7 (face d): “Uhm, large eyes. Uhm, long blondy hair. Happy looking.”

Two particularly interesting descriptions were collected that were very short and relied on participants having a common understanding of judgements of sexual attractiveness:

Description 8: “She’s quite fit” (*fit* is a British slang term for attractive)

Description 9: “She’s very average looking”

Unfortunately due to the random selection of descriptions in the experiment these two descriptions were never heard by a participant. However, in a small pilot test outside of the study where a grid was generated using the random decoy technique, 4/5 males guessed correctly which female was “quite fit” whereas 1/5 females were able to guess correctly. Matching the second description was more difficult as both male and females matched the correct twice.

4.2 Sex differences

On average females outperformed males with an overall score of 3.21 vs 2.99 (not significant). However, female performance was more consistent than that of males as displayed (see figure 9).

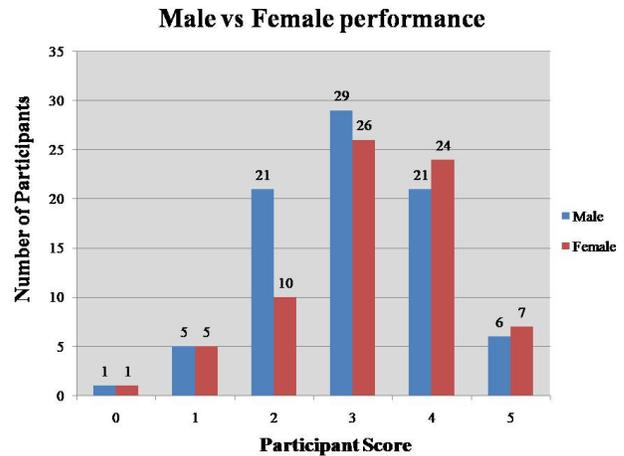


Figure 9: Comparison of male and female performance.

The upper tail ratio for female participants either achieving 4 or 5 is 1.14. This means that for every 100 males achieving 4 or 5 you would expect 114 females to do the

same. We also noticed trends that females scored better using female descriptions. Aside from females scoring the most successful logins using female descriptions, they also performed better overall using female descriptions rather than male descriptions (3.39 vs. 2.97, $t=1.71$ $p=0.09$). Interestingly males performed better using male descriptions rather than female descriptions 3.04 vs 2.92 ($t=0.51$ $p=0.6$) although this was not statistically significant. Females marginally outperformed males in every condition.

5. DISCUSSION

The results highlighted that in our set-up participants did indeed find it difficult to use descriptions to login to our system, with only 8% doing so successfully. However we believe this could be a factor of our experimental set-up rather than being a completely accurate reflection of the phenomenon. In the descriptions we collected there was such variation in participants' ability that a participant could have been hindered in scoring 5/5 by being assigned a description which to them was not so useful. In this situation a participant is almost reduced to a random guess. In a set-up incorporating a two way dialogue between describer and listener we feel the results would have been more resounding.

Our main contribution has been to demonstrate that description could be made more difficult by strategic selection of decoy faces. Participants in the random groups condition achieved an overall average score of 3.57, compared with 2.87 in the visual groups condition and 2.81 in the verbal groups condition. As these figures indicate participants performed best in the random condition, where it was most likely faces were distinctive enough for descriptions to have maximum impact. Interestingly there was no significant difference between performance in the visual group and the verbal group which suggests each has a similar detrimental effect on the sharing of Passfaces. Evaluating the impact of these techniques on memorability is the next step.

Real-world implications of future research confirming and extending these results could involve the similarity of Passfaces decoys being adjusted depending on the security context of system deployment. In corporate environments where password sharing is not desirable, the similarity of decoys to the target face according to some metric could be increased, whereas for other contexts this could be relaxed.

While our results did not reveal any significant performance differences between male and female participants, it was interesting to note that females did perform better on average across all conditions over males. In addition, descriptions created by females tended to include more detail and be significantly longer in duration than those of male participants. As such, females performed significantly better when using female descriptions over those of males. In fact, the behaviour of both sexes did conform to establish gender differences, for example, male performance showed a larger

variance than female performance. In the gender difference literature [13] it is often noted that due to the greater individual differences in male performance, more males than females should be recruited in experiments to gain a better impression of the extremes of male performance.

One simple technique that could be adopted to make description of recognition based secrets difficult is for the system to vary the order of the challenge images presented to the user. In this situation the best a describer could do is to describe all their faces in the hope the listener could identify with one of the descriptions. We speculate this technique would be highly prone to error on the part of the listener.

In terms of improvements in experimental design, we are interested in gaining access to a larger bank of images and as mentioned previously making future experiments more socially embedded affairs. For example, by grouping participants in pairs, with one describer and one listener. The describer would remotely communicate the (for example by telephone) graphical password and the listener would need to interpret the description to perform the correct action. This change would permit a two way dialogue to emerge between each participant to debug ambiguity in descriptions. Such an experimental setting would give a better indication of the likely manifestation of graphical password sharing in the real world.

6. CLOSING REMARKS

Our study has in part demonstrated the degree to which Passfaces can be verbally described, but also how through judicious choice of decoys we can reduce the vulnerability of Passfaces to description. Our empirical study has highlighted the reality that contrary to common wisdom users can share Passfaces graphical passwords. This observation can very likely be extended to other graphical password schemes and we would contend that the topic of description warrants levels of investigation on a par with accepted issues such as interference and shoulder surfing.

We also anticipate that the vulnerability of graphical password schemes to description could have impact on issues of both memorability and shoulder surfing. In a related ongoing unpublished study using Passfaces, a large number of participants have commented on how their memory strategy for their Passfaces comprised memorising verbal descriptions. If decoy faces in Passfaces employed a grouping policy that means that they were in some way similar to the target face, it is likely to impact on the ability of a user to select the target face. Likewise, in a shoulder-surfing scenario where attackers have a limited amount of time to form a quick memory association with components of an authentication secret.

We conducted an informal pilot study using our own implementation of the passpoints [28] system. 5 users ability to select click points in response to verbal descriptions were investigated. Figure 10 shows the sample image we used along with the 5 chosen click points and the order of these. The experiment involved the listener sitting in front of a computer screen with our passpoints software loaded. A male experiment moderator stood behind the participant and described the location of each click point in turn. No gesticulation was allowed and descriptions were not permitted to include reference to the current position of the mouse pointer, only points on the image. The listener was allowed to request more information or clarification to which the de-

Table 3: Female vs male performance in each experimental condition

	Female	Male
Random groups	3.72	3.42
Visual groups	2.92	2.83
Verbal groups	2.96	2.68

scriber could respond. The results of this pilot study showed 4 out of 5 participants were able to correctly interpret the descriptions into the correct sequence of click points.

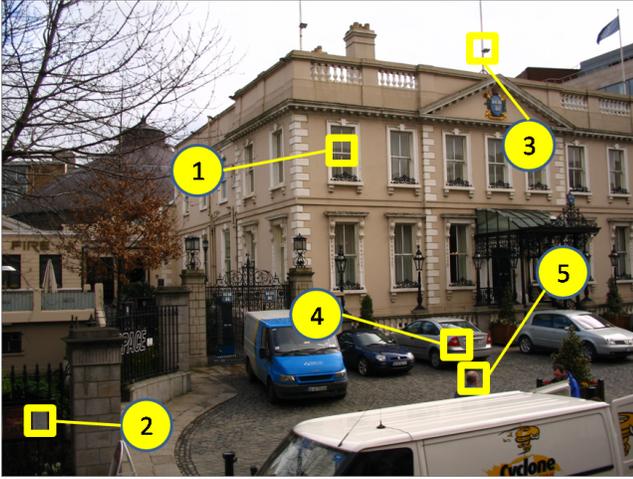


Figure 10: Passpoints description study.

In ongoing work we are investigating the vulnerability of other graphical password schemes to description, and the implications of this on users' authentication experience. Image selection schemes such as Story [7] and VIP [8] use images representing very different themes and are likely to be easily described based on the content alone. Schemes such as Déjà Vu [9] in part aim to explicitly address description, yet the computer generated random art yields unique images. Thus Déjà Vu may admit description via the very distinctive features that are intended to guarantee its immunity. Similarly, Draw-a-Secret (DAS) [15] drawings are likely to be easy to record on paper, using sketches that can be readily supplemented with pointers indicating where to start each stroke. Furthermore, in relation to verbal description it seems likely that techniques could be developed to communicate DAS drawings.

In an alphanumeric password setting users are typically forbidden to write down and share passwords. However, conventional passwords are so well suited to this type of distribution that users are able to share them with anyone they please. Though this is not necessarily undesirable, if system administrators do not wish users to write down and share passwords they might use a scheme which by its very nature mitigates against the sharing and external recording of authentication secrets. Indeed, as we better understand the nature of password description, the ability to write down and describe secrets might in future be incorporated as an explicit selection criteria for authentication schemes.

7. ACKNOWLEDGMENTS

This work was supported in part by Microsoft Research through its PhD Scholarship Programme, and all faces were sourced from the Passfaces online demo.

8. REFERENCES

- [1] A. Adams and M. A. Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, 1999.

- [2] G. Blonder. United States Patent 5559961, Graphical Passwords, 1996.
- [3] S. Brostoff and A. Sasse. Are Passfaces more usable than passwords? A field trial investigation. In *HCI 2000: Proceedings of People and Computers XIV - Usability or Else*, pages 405–424. Springer, 2000.
- [4] S. Chiasson, P. C. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *USENIX-SS'06: Proceedings of the 15th conference on USENIX Security Symposium*, pages 1–1, Berkeley, CA, USA, 2006. USENIX Association.
- [5] J. Davies. Visual Code Recordal and Communication Thereof International Patent PCT/GB1999/001688, 1999.
- [6] D. Davis, F. Monroe, and M. K. Reiter. On user choice in graphical password schemes. In *SSYM'04: Proceedings of the 13th conference on USENIX Security Symposium*, pages 11–11, Berkeley, CA, USA, 2004. USENIX Association.
- [7] D. Davis, F. Monroe, and M. K. Reiter. On user choice in graphical password schemes. In *SSYM'04: Proceedings of the 13th conference on USENIX Security Symposium*, pages 11–11, Berkeley, CA, USA, 2004. USENIX Association.
- [8] A. De Angeli, L. Coventry, G. Johnson, and M. Coutts. Usability and user authentication: Pictorial passwords vs. pin. In *McCabe, P.T. (Ed.), Contemporary Ergonomics 2003.*, pages 253–258. Taylor & Francis. London, 2003.
- [9] R. Dhamija and A. Perrig. Déjà Vu: a user study using images for authentication. In *SSYM'00: Proceedings of the 9th conference on USENIX Security Symposium*, pages 4–4, Berkeley, CA, USA, 2000. USENIX Association.
- [10] A. E. Dirik, N. Memon, and J.-C. Birget. Modeling user choice in the PassPoints graphical password scheme. In *SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security*, pages 20–28, New York, NY, USA, 2007. ACM.
- [11] D.M.Horgan. Language development. *University of Michigan doctoral dissertation*, 1975.
- [12] P. Dunphy and J. Yan. Is FacePIN secure and usable? In *SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security*, pages 165–166, New York, NY, USA, 2007. ACM.
- [13] D. F. Halpern. *Sex Differences in Cognitive Abilities*. Lawrence Erlbaum, 3 edition, 2000.
- [14] J. Huttenlocher, W. Haight, A. Bryk, M. Seltzer, and T. Lyons. Early Vocabulary growth:Relation to Language Input and Gender. In *Developmental Psychology*, volume 27, pages 236–248.
- [15] Ian Jermyn and Alain Mayer and Fabian Monroe and Michael K. Reiter and Aviel D. Rubin. The design and analysis of graphical passwords. In *SSYM'99: Proceedings of the 8th Conference on USENIX Security Symposium*, pages 1–1, Berkeley, CA, USA, 1999. USENIX Association.
- [16] M. Keith, B. Shao, and P. J. Steinbart. The usability of passphrases for authentication: An empirical field study. *Int. J. Hum.-Comput. Stud.*, 65(1):17–28, 2007.
- [17] E. Lieberman and R. C. Miller. Facemail: showing faces of recipients to prevent misdirected email. In

SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security, pages 122–131, New York, NY, USA, 2007. ACM.

- [18] Passfaces Corporation. The Science Behind Passfaces.
- [19] Passfaces Corporation: <http://www.passfaces.com>.
- [20] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the 'weakest link' a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19(3):122–131, 2001.
- [21] SearchSecurity.com - http://searchsecurity.techtarget.com/news/article/0,289142,sid14_gci895483,00.html. Employees willing to share passwords with strangers (2003) last accessed 23/02/2008.
- [22] Searchsecurity.com - http://searchsecurity.techtarget.com/news/article/0,289142,sid14_gci902867,00.html. Most workers must remember 6 passwords or more (2003) last accessed 23/02/2008.
- [23] S. L. Smith. Authenticating users by word association. *Comput. Secur.*, 6:464–470, 1987.
- [24] X. Suo, Y. Zhu, and G. S. Owen. Graphical Passwords: A Survey. In *ACSAC '05: Proceedings of the 21st Annual Computer Security Applications Conference*, pages 463–472, Washington, DC, USA, 2005. IEEE Computer Society.
- [25] J. Thorpe and P. V. Oorschott. Human-Seeded Attacks and Exploiting Hot-Spots in Graphical Passwords. In *16th USENIX Security Symposium, Aug. 6-10, 2007*.
- [26] T. Valentine. An Evaluation of the Passface Personal Authentication System. Technical Report. London: Goldsmiths College University of London., 1998.
- [27] T. Valentine. Memory for Passfaces after a long delay. Technical Report. London: Goldsmiths College University of London., 1998.
- [28] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon. PassPoints: design and longitudinal evaluation of a graphical password system. *Int. J. Hum.-Comput. Stud.*, 63(1-2):102–127, 2005.
- [29] J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password Memorability and Security: Empirical Results. *IEEE Security and Privacy*, 2(5):25–31, 2004.
- [30] M. Zviran and W. J. Haga. Cognitive passwords: the key to easy access control. *Comput. Secur.*, 9(9):723–736, 1990.
- [31] M. Zviran and W. J. Haga. A comparison of password techniques for multilevel authentication mechanisms. *The Computer Journal*, 3(3), 1993.

9. APPENDIX

A number of faces from the facebank and descriptions:



Female: This girl looks quite young, she's got

uhm dark hair that looks as though she's had highlights in but it's on the whole very dark. Uhm, she's a uhm, got dark eyes and sort of quite heavy eyebrows again. Sort of pronounced chin this girl. She looks quite friendly, her forehead's quite big and her cheeks quite pink.

Female: uhm <break> smiling a little bit shyly. Uhm, she's got dark brown hair hanging loosely probably in a long bob along her face, centre parting again, no fringe. Probably quite slim. <break> Winona Ryder type if I had to describe her like anybody.

Male: White female, sort of light brown to reddish hair. Long. Uhm, wideish mouth, quite tanned, dark eyes <break> square head. Prettyish.

Male: Fairly attractive woman with long brown hair.



Female: Uhm, he's got short, dark hair that's quite closely cropped and very prominent dark eyebrows, uhm, they're quite long as well. Uhm <break> when he smiles he's got quite a big smile that kind of transforms the rest of his face. Uhm <break> he also has, uhm, either quite dark jaw line, uhm, when he doesn't shave, oh, he's got stubble as well. Uhm, and he looks, uhm, he looks quite surprised.

Female: White male with short <break> brown black hair. Very thick eyebrows. Uhm, quite smily, quite cheeky smile. Uhm, got a black jumper that kinda collars up which is, I think, black or brown. <break> and he's got quite a small smile. And <break> he's got quite a heart shaped face.

Male: Male with slightly eccentric smile on his face. Has rather dark, busy eyebrows. Has, some, slight, moustache and beard stubble going on. And has a rather pointed chin.

Male: Uhm, big chin, bit of facial hair going on. Wide open eyes. Big bushy eyebrows. Uhm, sort of long thinish face.



Female: Here's another cheery looking girl with slightly reddish face and slightly shiny brow. She's got dark blonde hair which is hanging loose from her centre parting. <break> Uhm, dark brown eyes.

Female: Uhm, this girl look, uhm, she has expensive hair, actually, sort of quite blond but sort of looks like it's been done nicely. She has a nice smile.

Male: Fairly attractive woman with long brown hair.

Male: Uhm, <break> girl, brownish hair, quite tanned. Nice smile. Uhm, normalish eyes, normal eyebrows.



Female: He's got very neat, short dark, very dark hair. Uhm, quite a low, wears his hair, style, uhm, which is quite low over his forehead. Uhm, perhaps quite a low hairline. And he's got very prominent dark eyebrows that do very long, is actually one side of his face to the other, really. Uhm, he's got, uhm, very slight mustache. Uhm, and when he smiles he doesn't, his teeth don't show. Uhm, and he looks quite young as well like he doesn't, uhm, the fact that he, uhm, he has quite a youthful complexion also she's got quite slightly tanned skin.

Female: Okay, this one's, uhm, a man. He's got short, very dark brown hair that looks like its got some curl to it. Huge bushy eyebrows. Uhm, dark brown eyes and a bit of stubble where a mustache could grow but not on the rest of his chin. He's got quite full lips. Uhm <break> and he looks he's got an European kind of Italian or Spanish colouring to his skin, or a sun tan. Uhm, he's got quite a wide nose at the bottom of his nose but the bridge of his nose is quite narrow. Uhm.

Male: Boy. Laughing without showing his teeth. Looking straight ahead, thick eyebrows.

Male: Male with, very thick dark black hair. Very strong hairline. Has dark bushy eyebrows and slight moustache stubble.



Female: Uhm, <break> a blonde young woman, with a quite hesitant look about her. She's got her hair tied back and there's <break> little bits coming down the sides. <break> Scarf tied around her neck.

Female: Okay, she's a girl and she's got kind of longish blonde hair, looks as though it's been highlighted, so. Uhm, quite straggly wearing a scarf.

Uhm, looks quite friendly. Uhm, got her hair tied back. That's about it really.

Female: She's got blonde hair. Uhm, quite a wide open face, quite pretty. <break> Uhm

Male: Okay, female, uhm, sort of longish, uhm, darkish blonde hair with a some sort of yellow top and some sort of whiteish collar. Erm. <break> I don't know, slightly long chin <laughs>.

Male: Female, blonde, dark eyes, thick red lips. <break> Round face, good skin

Male: Er, blonde woman, quite quite sticky out chin. Big lips, quite attractive



Female: He's got, uhm, hair that come down to his jaw line, uhm, it's quite thick hair almost like a girl would wear in a bob. Uhm, he's got, he's not smiling a great deal but even he's got quite a small mouth as well, he's got dark eyes which are quite prominent, uhm, and quite, uhm, a reddish pink complexion as well.

Female: Okay, the next one's a man. Uhm <break> he's got quite <break> an oval face, quite a pointy chin <break> he's got brown hair cut down to his, uhm, chin. It's quite, a bit wavy a bit unkept. <break> Uhm, he's got a small mouth <break>. He's got brown eyes and quite bushy eyebrows. Uhm <break> he's got quite a wide face, and a bit of a vacant look about him in this picture.

Female: White male. Uhm, chin length brown brown hair that's quite wavy. Uhm <break> he's got a brown jumper on. Uhm.

Male: A boy with long brown hair. A small smile on his face. Wide coloured face.

Male: Male with <break> chin length, slightly unkept, dark brown hair, long dark brown hair. Ovalish square type face.

Male: Uhm, first thing to notice is the hair. Uhm <break> quite big open eyes, uhm, <break> again face isn't symmetrical, uhm, <break> first thing to notice is the hair. The mouth, not much of a defined jaw line, cheek bones or anything like that



Female: Blond woman, smiling cheerfully straight at the camera. Her hair's tied back as well. She's got quite an open look to her, face. Rosy cheeks. Nice smile.

Female: She's got, her hair scraped back, quite a large chin, quite pretty, quite a large nose, looks a lot like Jo from S Club 7.

Female: Okay she looks quite friendly. She's got her hair tied back. Blondish by the look of things. Nice, uhm, showing her teeth so she looks though she's very friendly with sort of pale white skin and kinda wearing a scarf, looks refined.

Male: Female, round face, uhm, wide chin, dark eyes, brown hair, blonde highlights.

Male: Erm, female, uhm, combed back sort of hair, wearing some sort of dark scarf. <break>Err <break> Sort of a slightly broader bottom of the face, with showing her teeth grinning.



Female: He's got dark hair, sorry brown hair, it's not really very dark. It's quite spiky at the top, it's short. Uhm, he's got kinda a ready complexion, very nice smile, laughing eyes, uhm, he's got a pointed chin but his features are quite close together.

Female: Happy looking guy. <break> Short brown hair. <break> A bit messed up at the top of his hair.

Male: Uhm, white, uhm, kinda pinkish, probably British. Uhm, kinda spike hair. Uhm, short at the sides. Uhm, quite prominent smile. Uhm, pointed nose. Ears slightly sticking out. And rounded eyebrows, sort of like brown eyes.

Male: It's a, man, with, brown, quite short hair. It's shorter at the sides than at the top. Uhm, quite small eyes, and <break> clean shaven. Quite broad mouth. <break> Possibly a pointy nose. I think I said quite small eyes.

Male: Male. Fluffy brown hair on the top cut short at the top. Brown eyebrows, curved. Smiling face, uhm, although not particularly smiling. Slightly flushed looking cheeks. Uhm, pointed chin. Dark eyes. And, uhm.



Female: Chinese or Taiwanese girl. Smiling, slightly. Erm, looking full face sort of onto to me with her dark hair tied back, middle, uhm, side-parting, sorry. <break> Just looking down slightly but still looking right at me.

Female: She look quite young. She's got short, dark hair. Nice eyes, nice eyeshadow. Uhm, she looks young. I think she's, sort of Asian. Round face, and hair is short and dark, with a severe parting.

Female: Okay, this girl has either got, uhm, quite short hair or it is tied back. Oriental with, uhm, brown eyes and very dark hair. Smiling, doesn't look desperately friendly, but I think it's just the camera in front of her.

Male: Uhm, female, uhm, slightly Oriental looking, dark hair, comb parting on the left hand side, uhm, wearing some sort of white top.

Male: Uhm, female dark eyes, dark hair, Asian origin, uhm <break>, round face, round chin.

Male: Er, female, Asian origin, dark hair.



Female: She's got long blonde hair, uhm, it's not really light blonde, it's sort of dark blonde. She's got a middle parting. She's again smiling. She's got quite a reddy complexion, sort of healthy looking. She's got, prominent cheek bones. <break> Uhm <break> probably dark eyes. A longish straight nose. Uhm <break> quite a big chin, perhaps. And that's about it, I think.

Female: Girl with a parting slightly to the side. Her hair is long <break> but she's got <break> dark eyes. Colour in her face and <break> her hair grows past her shoulders.

Male: Uhm, pink skin. Uhm <break> blonde fair hair. Looks very long. Uhm, straight, centre parting. Uhm, quite prominent cheek bones.

Male: Uhm, it's a woman with, quite a, quite a square, look at the top of her hair, sorry head. There are tapers at the bottom. Uhm, hair is sort of brown but it's almost yellow, uhm, not blonde but yellow. Uhm, and quite a pronounced forehead, uhm, with a parting in the middle. Uhm, blue eyes, I think. Uhm, <break> and <break> quite pronounced cheeks.

Male: Blonde female. Uhm, wavy hair. Going down past the shoulders. Uhm, slightly curved eyebrows, smiling. Uhm, tanned complexion, maybe. And, uhm, maybe highlights in the blonde hair. Uhm, quite big dark eyes and a, uhm, curved face with a high forehead.