Pictures or Questions? Examining User Responses to Association-Based Authentication

Karen Renaud
Department of Computing Science
University of Glasgow
karen@dcs.gla.ac.uk

Mike Just
School of Engineering & Computing
Glasgow Caledonian University
justmikejust@hotmail.com

Challenge questions are commonly used as a backup should users forget their “main” authentication secret. Such questions are notoriously difficult to design properly, and have sometimes allowed intruders to access the system via a back door simply by engaging in some online research about the victim [33]. Most challenge questions rely on a user’s knowledge of their early life, something which tends not to deteriorate over time [15]. Unfortunately, this kind of information can also be discovered by a determined attacker. We developed a challenge protocol in which a set of pictorial cues are used to prompt answers, rather than using the standard mechanism based on textual questions. The prompts solicit associative memories that need not represent factual information (information that aids an attacker in mounting targeted observation attacks) and serve as a stronger cue to aid the recall. Our results reveal that the solution has comparable security with that of traditional challenge questions (when considering external attackers), and suggests additional benefits from posing three or more questions serially. Furthermore, we obtained a 13% increase in the memorability of our (name-based) answers, while our results suggest enhancements could help improve the recall of place-based answers. We conclude by discussing how further modifications could achieve gains on the usability front.

1. INTRODUCTION

Secure and usable authentication is critical to the success of numerous applications. A secure solution ensures that only an authorized user is able successfully to authenticate, thereby reducing the risk to the application and the user, as well as the potential costs associated with authentication compromises. A usable solution ensures that authorized users are able successfully to access their application as desired. An unusable solution can lead to insecure behaviour, increased authentication time, increased help-desk calls, and reduced enrolment, especially if application use is optional.

The most common form of authentication in use today is the password. The usability and security of password authentication has been widely studied [1], and the negative results indicate that alternative solutions are desirable. Alternative information-based solutions, such as graphical passwords [11, 12, 5, 27, 29, 28], have provided encouraging results, though they have not, as yet, seen significant deployment. Biometrics and smartcards have promised an as yet undelivered revolution in how we authenticate. For now they remain niche solutions. Alternative information-based authentication solutions offer the potential for retaining a similar interface, and could provide benefits if shown to have an advantage over passwords. In the past few years, there has been significant research into

the usability and security of challenge questions — a form of authentication that seemingly rose to prominence as a form of account recovery, without significant research or analysis. Rabkin [26] revealed the insecurity of the challenge questions used by 20 financial institutions. Schechter et al. [30] led user experiments to uncover the lack of usability and insecurity of the questions used by several prominent Internet web sites. Just and Aspinall [19, 20] discovered that when given the option of choosing their own questions, both usability and security levels remain low. Bonneau et al. [4] extend the security challenges further by demonstrating that answers to challenge questions lack significant variety on their own, so that a trawling attacker¹ can be quite successful in guessing answers without knowing the user personally.

Given the weight of such evidence, is it reasonable to hope that a secure and usable authentication solution can indeed be built from challenge questions? Alternatively, do we need to conclude that the entire concept is irretrievably flawed? We feel that it is too early to come to condemn challenge questions out of hand. It is possible that the reported security flaws may apply only to existing solutions and protocols, and that novel alternatives might well fare better. For example, notice that the security results of Bonneau et al. [4] apply only to single-question solutions. Also, current solutions are certainly extremely susceptible

¹An attacker who targets users indiscriminately, rather than targeting a specific user

to targeted observation attacks, in which answer data is mined from public sources. A solution that avoids such factually-based answers could well be more secure and therefore merits investigation.

Here we propose a novel alternative which alters the current model in three key ways:

1. Pictures are used to help cue the answers for improved usability.
2. We rely on associative memories that need not be factual, and hence, may not be so easily discovered by a targeted observation attack.
3. We use three questions serially, strengthening the system against lucky guesses and requiring correct answers of all three in order to authenticate.

We conducted an experiment with 184 users to test this proposal. Here we report on our findings. Section 2 provides a more detailed overview and justification of our protocol. Section 3 describes our methods for collecting and analyzing our data, including details regarding our experiments. In Section 4, we present our security and usability results. We provide some additional analysis and suggest opportunities for future work in Section 5. Section 6 concludes.

2. OUR SOLUTION: PICTURES –TEXT

Our proposal relies on, and exploits the strength of, associative and attribution memory. Other associative passwords have been trialled: for sound clips [23] and for other words [32, 25]. Sound associations, which tested the association between a sound and an image, were not particularly successful. Word association works reasonably well, but is very time consuming, both at enrolment and authentication, and is not that dissimilar to traditional challenge questions.

Our proposal is to use three pictorial cues. We shall consider the animal cue first. Users will relate an animal picture to a person of their acquaintance. This requires them to match attributes they associate with that animal with some attributes of a person (known or unknown). It turns out we all engage in this kind of attribution, albeit unconsciously, in many cases. People have cognitive biases which lead them to attribute characteristics to people [2, 14]. People also anthropomorphise animals [24, 7], as evidenced by their use in popular idioms eg. Sly as a fox and Proud as a peacock.

At enrolment, the participant will be offered a choice of a number of animals. He/she will choose one and name a person who is considered to share attributes of the animal. At authentication, participants will be presented with the same animal picture and requested to provide the same name. There is reason to believe that, having provided a name in response to the picture cue at enrolment, the person

will be able to retrieve that name after a time lapse [8] as our experiment will examine.

We also wanted to exploit the strength of other associative memories. Psychologists argue that conscious memory is nothing more than a rich network of associations [18, 3, 16] (cited by [13]). Hebb [16] argued that two items could be associated even if they have not occurred together in a particular person's own history. For our second cue, therefore, we used well known world cities. We wanted to determine whether people would easily associate places to people, either because they have met someone from that city, have vacationed there, or have some other association with the location. We therefore chose a number of pictures of well known world locations, and asked for the name of someone they associated with that location.

The third cue exploits flashbulb memory [10] (e.g., recalling where you were when you heard that President Kennedy was assassinated) which has been shown empirically to have naturally high recall [6, 21]. It is thought that the surprise elicited by such events, as well as the accompanying emotions, make the memories more durable.

The pictures all serve as an additional recall aid, while the use of an indirect question helps to reduce the exposure of the user to targeted observation attacks [19]. The questions are indirect in the sense that we wouldn't expect to find the direct answer to such a question in a database or public source, such as a social network. In relation to a targeted observation attack, a traditional challenge question (such as "What is your mother's maiden name?") is used by an attacker to search for the answer to the question for the targeted user. Such information can often be retrieved externally to the authentication system, which offsets the benefits gained from the much vaunted durability of these facts in user memory. If we further consider the steps an attacker takes to discover the answers, we observe two key components required by the attacker: The first is the identity of the user, and the second is knowledge of the answer to the question.

It should be assumed that an attacker can easily determine the identity of the user, e.g., often inferred by the username for the account (especially if it's an email address). As to the answer to the question, notice that the earlier example of Mother's maiden name reveals significantly more information than if the question were more ambiguous or indirect, such as "Name of Person?". While both questions could solicit the same answer from a user, they would seem to vary in both their usability and security. Asking directly for a Mother's maiden name is vulnerable to a targeted attack against a user, whereas the second question might equally be answered by any name. Bonneau et al. [4] have shown that questions asking for such names can be guessed easily by a

trawling attacker concerned with compromising at least some, unspecified accounts, but the results do not extend to the use of multiple questions. Hence the more ambiguous question would seem to offer the potential for improved security. However, with regard to usability, while the ambiguous question is more challenging to an attacker, it may similarly be challenging for our user to recall later. Our solution attempts to more carefully construct similar, indirect questions, and to use a corresponding picture to aid with memory recall.

In summary, our challenge protocol incorporates the following serial stages:

- **Question 1. Animal – Name:** The user is asked to choose the picture of one animal from a small set of 45 common animals. The user is then prompted: “Who does this animal remind you of?”, and a person’s name is provided. The animals were chosen by consulting lists of common idioms and proverbs which utilised animals and also ensuring that widely-known animals were included. We did not use caricatures or cartoon animals, but rather pictures of real life animals. Examples are: a hyena, peacock, lion and badger.
- **Question 2. Place – Name:** The user is asked to choose the picture of one well known city from a small set of 30 locations. The user is then prompted: “Who does this place remind you of?”, and a person’s name is provided. A list of most famous world cities was consulted in order to populate this list.
- **Question 3. Event – Place:** The user is asked to choose the picture of one famous event from a set of 23 flashbulb events. The user is prompted: “Where were you when this event happened?”, and he or she provides a description of his or her location at the time. A number of sources for flashbulb memories were consulted, and we made an effort to include some events from each decade from the 1930s onwards so that people of all ages would be able to identify with at least one of the flashbulb events.

In summary, our solution involves a level of indirection, using pictures to aid the recall of the corresponding answers. In this paper, we use two types of questions, though we expect that additional variants are possible. The first type asks the user to associate a person name with one of a small set of pictures. The first and second stages are based upon this variant. The second type asks the user to

associate a place name with one of a small set of pictures. The final stage is based upon this variant. See Appendix for a selection of the photos from each of these stages

3. METHODOLOGY

We designed an experiment in which users would register their questions and answers, and later return to provide their answers again. We also asked participants to nominate two additional users who would attempt to guess their answers so that we could test the security of the systems.

As a first step, we conducted a pilot study: a paper-based trial of the animal-based prompt with 10 participants in order to refine our methodology, and to test the effectiveness of our proposal. We showed participants a piece of paper with a number of pictures of animals on it. We then asked them to choose an animal that reminded them of someone, and returned in a week to show them the picture of the animal again and asked them to give us the name of the person they had provided previously. We found that most people remembered the name but that it was often not given in the exactly the same format. People would give both first name and surname the first time, but only provide the first name when they returned or visa versa. Only one participant forgot the name of his person.

We subsequently developed an online version of our authentication solution to conduct an experiment with the following characteristics:

- We distributed a web link for our experiment to various colleagues, and also solicited participation from a class of Year 1 students.
- We assigned participants randomly to two experiment groups: (i) A Control Group which posed traditional challenge questions, and (ii) A Picture Group which used our proposed solution. (see Appendix for a list of the control questions used)
- Each participant was asked to provide the email addresses of two people who were nominated to try to guess their answers. We asked for the email address of close friend and of an arms-length acquaintance. This would help us test the strength of the mechanism against people who knew the participant very well as opposed to someone who only knew the person slightly.
- Participants were emailed a week after registration to request them to return to

supply their answers again. Two participant names were randomly selected to receive a £25 gift certificate each, as a reward for participating.

A total of 184 participants visited our web site, with 172 registering 3 questions (90 Control Group and 82 Picture Group). There were 117 females and 67 males. The age and domicile country distribution is:

Table 1: Demographics of Participants

Age Group	Number	Country	Number
< 20	50	UK	134
20-29	50	South Africa	19
30-39	23	Canada	4
40-49	15	US	3
50-59	29	Denmark	2
60 and over	17	Other	10

153 participants returned for the second stage (77 Control Group and 76 Picture Group). Thus, for the Control Group, there are $77 \cdot 3 = 231$ choices of traditional challenge questions, and $76 \cdot 3 = 228$ choices of picture-based questions from the Picture Group.

We asked each participant to provide us with one email address for a close friend or family member who knows them well and an arm's length acquaintance whom they felt did not know them very well. 142 participants provided the email address of a close friend, and 134 provided email addresses of arms-length acquaintances. The participants were then asked to choose three questions or pictures as prompts and to provide the answers.

We emailed each nominated person (without telling them whether the participant viewed them as "close" or not) and asked them to provide three guesses at an answer to correspond with each prompt (either picture or question). Individuals were not informed whether their guesses were correct or not. 117 people nominated as close friends attempted to guess their participant's answers and 93 arms-length acquaintances visited the website to guess the answers.

All participants and nominated persons were asked to complete a questionnaire upon conclusion of the experiment; the results of which will be referred to below.

4. ASSESSMENTS

We assessed both the usability and security of our solution and report on our findings here.

4.1 Usability Assessment

Usability is traditionally assessed in terms of efficiency, effectiveness and satisfaction. The efficiency of our

solution can be expected to be similar to traditional challenge questions at authentication, but enrolment is likely to take a little longer. We chose not to focus on this aspect since the main problem with challenge questions, which we were attempting to address with this experiment, is their effectiveness in authenticating users. Users often struggle to convince the system of their identity, either because they have forgotten the answer, or because they enter it incorrectly. To assess satisfaction, we asked participants to complete a questionnaire, both after enrolment, and return authentication, to assess general satisfaction.

As noted earlier, 153 participants returned to authenticate at least one week after their initial registration in order to test their memory of their registered information and their ability to enter the answer exactly as entered at enrolment a week earlier.

Exactly Correct Answers

Here we consider answers that were entered exactly as provided at enrolment. A correct answer here indicates two things: that the answer was remembered and that the participant was able to reproduce it exactly. When considering the Control Group, with traditional challenge questions, of the 231 questions and answers across 77 participants, 146 of the 231 questions were answered with the exact same answer (63%). If we consider the first two questions exclusively (since they both ask for a person name), we see that 104 of 152 (68%) answers were exactly correct. For the last question (location-based) 43 of the 77 questions were answered exactly correctly (56%).

When considering the Picture Group, with picture-based questions, of the 226 questions and answers across 76 participants (two participants only guessed their first answer), 174 of the 226 questions were answered with the exact same answer (77%). When considering all three questions, this is a 14% improvement over the Control Group. If we consider the first two questions only (since they both ask for a person name), we find that 136 of 151 (90%) answers were exactly correct. When considering only the first two (person-namebased) questions, this is an 22% increase when compared with the Control Group. For the last question, only 38 out of the 75 questions were answered exactly correctly (51%).

Therefore, when considering a requirement for an exact answer match, we observe that questions requiring a person- name were recalled 70% of the time for traditional challenge questions, and 90% of the time for our picture-based alternative. In both cases though, questions soliciting a place name fared much worse: 56% success for the traditional challenge questions, and 51% for our picture-based alternative. Requiring a two-part place name undoubtedly played a role in this low success rate. For example, a person would enter "Glasgow, UK"

at enrolment, and then entered “Glasgow, Scotland” when they returned. The decision to require a two-part name was made in order to provide a stronger test, but ended up impacting the usability of the system to an undesirable extent.

Similar Answers

Here we consider answers that were entered approximately as provided at enrolment. A similar answer indicates that the answer was remembered correctly but that the participant was unable to reproduce it exactly. To assess this, we manually compared the original answers to the returned answer for Question 3 (which had a place name answer). We specifically focused on this question due to its low recall rates (and, as noted earlier, its low guessing rates). We rated the comparison as either

- exactly correct,
- partially correct (which included spelling mistakes, use of punctuation and reversal of words),
- semantically similar (which included the use of acronyms, nicknames and alternate names), and
- incorrect.

It should be noted that during this manual comparison we discounted punctuation marks and spacing. The previous section reported on exact correctness. If this comparison were done programmatically, however, one would easily be able to remove spaces and punctuation so as to avoid false rejections.

Table 2: Categorisation of Return Answers

Category	Control	Pictures
correct	43 (56%)	38 (51%)
similar	22 (29%)	18 (24%)
semantically similar	5 (7%)	16 (21%)
incorrect	5 (6%)	3 (4%)
Total	77	75

It would appear that participants had difficulty describing the location identically when they returned to the site. With judicious use of an algorithm it would undoubtedly be possible to substantially reduce the number of errors, especially for partially correct answers. What these numbers do indicate is that location memory is very stable and enduring.

Satisfaction

Participants were asked to complete a questionnaire twice — once after registration and another when they returned. Here we report on some of the responses (see tables below). The registration questionnaire was completed by 91 Control Group participants and by 82 Picture Group participants. For the registration questionnaire, most responses appear quite similar across the Control and Picture Groups. Only 2 of the participants in each group had some difficulty understanding the process. Acceptability of authentication mechanisms is a large part of satisfaction so we asked them about their difficulty in choosing questions/pictures. While about a quarter indicated a difficulty with the traditional challenge questions (Control Group), there was more variance across the three question types for the Picture Group. Although we didn’t further determine the reasons, the greater difficulty in choosing an animal association may be due to the increased variance of associations that are possible for someone to make, while with the location and event questions (Questions 2 and 3), most participants might have limited associations/memories upon which to base their choices.

77 Control Group participants completed the questionnaire when they returned and 71 of the Picture Group completed the questionnaire.

The participants appeared ambivalent about the role the pictures had to play in helping them to remember their answers, which is surprising since the memorability was so much higher for the Picture group. Perhaps the cueing is so subtle as to be subconscious. As for the reasons provided for incorrect answers, the most interesting category is “Couldn’t remember which answer provided”, which was particularly high in the control group. This suggests that the online population is becoming increasingly weary and wary of challenge questions. This could be a direct result of newspaper stories directly related to accounts being breached based

Table 3: Registration Questionnaire Responses

Question	Control	Pictures
How secure is this system (1=Not,5=Very)	3.27	3.57
Anticipation of difficulty remembering answers (1=Very,5=None)	3.74	3.95
Difficulty choosing questions/pictures? (Y/N)	27%	41% (animal); 32% (location); 17% (event)
(1=Hard,5=Easy)		
How easy for stranger to guess answers	1.57	1.57
How easy for acquaintance to guess answers	2.2	2.13
How easy for friend to guess answers	3.26	3.27

Table 4: Return Questionnaire Responses

Question	Control	Pictures
Did pictures help re-call?(1=Lots,5=Not at all)		2.31
Reasons for Errors:		
Completely Forgot	8%	3%
Spelling	4%	11%
Couldn't remember which answer	25%	14%
No Idea	7%	7%

on these questions [33, 9] or merely due to the ubiquity of systems requiring users to provide the answers to these questions at enrolment. People appear to provide (possibly false) answers at enrolment which they hope to remember later, but do not. It is an unfortunate fact that humans tend to overestimate their ability to remember things [22]. Interestingly, one of the participants provided her own name as the answer to her questions at enrolment. Unfortunately she had forgotten this when she returned a week later, and provided valid (but incorrect) answers.

We asked for additional comments and these made for interesting reading. Some comments from the control group are: Someone who knows me well might be able to guess one or two answers -but I don not think they could guess all three. ; I think the only reason I remembered the answer to my "least favourite teacher" and "favourite fictional character" questions was because, in this experiment, the time between choosing the answers and answering them was short... In a year it is very likely I would have completely forgotten.; Interested that, favourite singer might well be subjective, depending on my mood. Not as easy as a fact.

Comments from the picture group include: Very good method. I was not surprise because the image/name links were so compelling and strong and personal; I hope this is used world wide when finished.; This scheme is much harder for others to guess

4.2 Security Assessment

Our security assessment consisted of multiple strands. While our analysis considers each strand separately, we assume an attacker will choose the path of least resistance, i.e., the easiest attack strand. Attackers were prompted with the same prompt as that given to the legitimate user — either text or picture. They were allowed to submit three guesses for each prompt.

Blind Guess Attack

A Blind Guess Attack involves an attacker exhausting all possible words in an attempt to guess the answer to a challenge question [19]. The attacker's information is limited, excluding even knowledge of the corresponding challenge question.² In this case, we had no difference between our Control Group and Picture Group answers, as both had average

and median answers lengths of approximately 13 characters. (The median was 13 in both cases, with a mode of 12 for pictures and 13 for questions) Following previous approaches [19, 20] we can compute the Shannon entropy of such an answer by assigning 2.3 bits of entropy to the first 8 characters, and 1.5 bits to other characters. For 13 character answer this gives 25.9 bits of entropy.³

With three questions, either traditional challenge questions, or our picture-based variant would thus provide more than 75 bits of entropy. However, a determined attacker would be likely to accumulate further information in order to mount the subsequent attacks below (so that it is unreasonable for us to consider only a simple blind guess attack).

Statistical (Focused) Attack

Since our solution relies upon answers that are person and place names, we can utilize the analysis of Bonneau et al. [4]. With regard to person names, a full name offers 14.6bits of security versus an online attacker trawling to attack any account with 10 tries per account, and 23.3bits of security (for 50% success) against an offline attacker targeting a particular account. And while this level of security is certainly insufficient, it only applies directly to single question solutions. Our assumption is that three question-answer pairs are used, and must all be answered correctly for successful authentication.

In addition, we also considered the possibility of whether participants were selecting highly likely answers, e.g., the names of actors or actresses. As a partial indication of the 'popularity' of specific answers, we note that there was very little agreement on provided names from our participants. In the control condition, a handful of answers were chosen by more than one person: four people named Michael Jackson, and six named Jesus. Six answers were chosen by two people each. All the other answers (163) were provided by only one person. There was far less agreement than we had anticipated given the limited number of questions. In the picture condition, there was no agreement of answers provided by participants. This was expected since the pictures are intended to tap into personal memories of well-known people, unlike traditional challenge questions which appear to lead to more famous people names. While a cursory review of the answers does reveal a number of "famous names", we note that our picture-based solution increases the challenge for an attacker since the question does not narrow the space further. For example, while a traditional

²This is possible in at least two situations: (i) The attack is automated, and the question cannot be easily read (e.g., if it is CAPTCHAed), or simply not understood by the automatic reader, or (ii) The attack is manual, but what is being asked by the question is not understood by the attacker.

³Our entropy results are greater than those of Just and Aspinall [19, 20] since we encouraged participants to include a first and last name for Questions 1 and 2, and a two-part place description for Question 3, as opposed to allowing them to choose a single-part person moniker or place name.

challenge question might solicit an actor, or writer, the pictures leave such specifics to the user. Such indirectness should, again, increase the work factor for a statistical attacker.

Targeted Observation Attack

Unlike the previous two attack scenarios, a targeted observation attack takes into account personal knowledge known about a user. To assess such an attack, we used existing techniques [30] to determine how successful a close friend or an arms-length acquaintance would be at guessing a participant's answer. The following tables present the percentages of correct guesses. Note that in these tables a guesser is considered to have succeeded if any three of their nine guesses (three for each question) matched the participant's answers exactly. The number of successes is expressed as successes/total (percentage) for each of the three questions chosen by the participants, for a particular attacker type.

Table 5: Control Group -Success with Any Guess

	Quest. 1	Quest. 2	Quest. 3
Close	16/57 (28%)	14/57 (25%)	4/56 (7%)
Arms-Length	7/53 (13%)	4/51 (8%)	2/50 (4%)

Table 6: Picture Group -Success with Any Guess

	Quest. 1	Quest. 2	Quest. 3
Close	23/60 (38%)	20/60 (33%)	8/60 (13%)
Arms-Length	6/43 (14%)	1/42 (2%)	2/42 (5%)

Three people managed to guess all the answers correctly in the Pictures group but no one did this for the Control group. From these results, we observe that more than a quarter of the close attackers managed to guess one of the name-based answers (Questions 1 and 2) from the Control Group, while more than a third were able to guess one of the name-based answers from the Picture Group. However, the results are more comparable with the arms-length attackers in both cases (suggesting that the pictures offer no discernible advantage to external attackers). Though considering that we are relying upon three questions and answers to be answered correctly, so that all would have to be answered correctly at once, a more suitable assessment might consider on the first guess only. Certainly, a more determined attacker might try more possible combinations, though most systems will limit the number of failed attempts. Thus, if one considers only the first guess in each case, the figures are as follows:

In this case, the results between the Control Group and Picture Group are more comparable, across both a close attacker and arms-length attacker.

Table 7: Control Group -Success with First Guess

	Quest. 1	Quest. 2	Quest. 3
Close	12/57 (21%)	14/57 (25%)	3/56 (5%)
Arms-Length	6/53 (11%)	4/51 (8%)	2/50 (4%)

Table 8: Picture Group -Success with First Guess

	Quest. 1	Quest. 2	Quest. 3
Close	15/60 (25%)	15/60 (25%)	6/60 (10%)
Arms-Length	3/43 (7%)	1/42 (2%)	1/42 (2%)

In all the tables above, even though Question 3 demonstrates a noticeably lower guessing rate, this correlates directly with the low usability results for this question, as we discuss in the next section.

One of our three design decisions was the use of three questions serially, requiring correct answers for all. A comparison across the two groups is shown in the Table 9.

Table 9: Number of Correct First Guesses

# Correctly Guessed	Picture	Control
One Correct	22	17
Two Correct	7	6
Three Correct	0	0

Whereas the picture group appears much weaker when one considers how many "attackers" guessed only one answer correctly, these differences disappear once we look at how many people guessed two and three answers correctly. No one guessed all the answers correctly in either group and the number who guessed two is almost equal across the two groups.

Questionnaire Responses

110 "attackers" who guessed traditional challenge question answers completed the questionnaire while 88 guessers from the picture group completed the questionnaire. Below, we summarize their answers as related to their perceived difficulty in guessing the correct answers (reminder: they were not informed as to whether or not their guesses were correct).

From these results, we see that if one were to consider the perception of the attacker, the perceived levels of difficulty and success were the same between the Control and Picture Groups, and also across close or arms-length attackers (Acquaintance in table above). It is interesting to note that in the first row of the table above, the attackers did not perceive the pictures to be a significant aid in their answer guessing.

In addition to the specific questions, a comment box was made available. A selection of responses included: "This is very hard", "found it near impossible to come up with more than one guess.", and "I will think long and hard before I create my own challenge questions from now on!"

Table 10: Attacker Questionnaire Responses

Question	Control	Pictures
Did pictures help them to guess answers? (1=Lots,5=Not at all)	—	3.85
How hard to guess the answers (1=Very,5=Not at all) (Close/Acquaintance)	2.8 (C); 1.8 (A)	2.9 (C); 1.9 (A)
How successful do you think you were? (1=not at all; 5=Very)	2.12	2.17

5. DISCUSSION

We carried out both usability and security evaluations for the picture • text concept, as compared to a control group using traditional questions. What does the data tell us?

In terms of security, the picture-based answers were guessed by close friends 38% of the time, much higher than Jakobsson et al.'s [17] proposed 10.5% false accept rate. This is unsurprising. The participants' picture based responses are related to their day to day lives. The participant is asked to attribute characteristics to a person and link a person to a well known location. Clearly someone who knows the person well has a pretty good chance of guessing the picture-based answers, and our data proves this point.

Traditional challenge questions, on the other hand, tend to relate to facts about the person's early life, under the assumption that these will not easily be forgotten and are hard to discover. (The latter assumption is increasingly invalid with the explosion of personal information now available on Facebook. [26])

On the positive side, the picture-based system appears to be at least as strong as traditional challenge questions in resisting outsider/external attacks. If a single question were used, then the picture-based system resists attacks from close friends and family members far too poorly. At first glance this could lead us to conclude that the picture-based scheme is even weaker than traditional challenge questions.

One could argue that many users happily share their passwords with friends and family members [31] and that, in this respect the picture-based system is no weaker than the traditional authentication systems. Unfortunately one cannot be selective with authentication systems — they should resist intrusion from all comers, even those who know the person best.

However, looking only at the number of people who guessed one question correctly gives us only part of the true picture. One of our design decisions was to use multiple questions, and it is the wisdom of this decision that is confirmed by the results. The results presented in Table 9 show that when more than two questions are used in the analysis no one managed

to breach the system, no matter how close they were to the user.

The usability assessment produces some compelling evidence that the picture-based scheme is significantly more memorable than the traditional questions, especially for the first two stages. Although participants appear to have had some difficulty choosing their picture cues, they certainly recalled the answers readily enough a week later.

The most interesting finding, in terms of future work possibilities, is the location-based question usability and security results. The percentages of incorrect answers is very low which means that people remember very well where they were when a flashbulb event takes place. What they had difficulties with was describing it in exactly the same way when they returned. As future research we intend investigating ways of ameliorating this difficulty without compromising on security. However, one could not consider such an intervention until one considers the impact on intruders as well. The category impacted by such an algorithm is the similar category. In the control condition, the close friends offered similar answers in 34% of cases. It would be damaging, therefore, to use an algorithm to mitigate against the legitimate users' errors which would also make it easier for an intruder to gain access to the system. However, close friends offered similar answers in only 3% of cases in the picture group, which means that the use of the algorithm to mitigate against errors for the picture group users would have a negligible effect on the security of the scheme.

6. CONCLUDING REMARKS

We set out to trial an alternative to traditional challenge questions which would, we hoped, be more usable than traditional challenge questions, while making the online discovery of the answers more difficult. We explicitly tested the security of the proposed system by requesting the participants to provide email addresses of two people — one of whom would be likely to have extensive personal knowledge of the participant.

We found that the picture-based system was far more usable than traditional challenge questions. It also maintained the same level of security as traditional questions as long as multiple questions are used in serial order. There is clearly work to be done in refining this solution but our results suggest that it is worth investigating further.

Beyond these potential improvements, we also plan on investigating the potential for interference where our solution might be used at multiple accounts. Our initial hypothesis is that the use of pictures at each account will be beneficial, and reduce interference, at least when compared to a traditional, textual challenge question solution.

APPENDIX

Challenge Questions Used

Person Questions

The name of your first boss?
 The name of the person who officiated at your wedding?
 Who is the person that has changed the course of your life?
 Your favourite biblical character?
 Who is someone, now deceased, that you would like to dine with?
 Who would you like to be stranded on a desert island with?
 Who was the boss that you disliked the most?
 Who is the worst actor/actress?
 Who is your favourite author?
 Who is the author you dislike the most?
 Who is your favourite singer?
 Who is the singer you dislike the most?
 Who is your favourite fictional character?
 Who is the fictional character you dislike the most?
 Who was your favourite teacher?
 Who was the teacher that you disliked the most?
 Who is your favourite actor/actress?

Location Questions

Where was your maternal grandmother born (mother's mother)?
 Where would you most like to live?
 Where would you least like to live?
 To where did you first fly on an airplane?
 Where is your favourite restaurant or pub?
 Where is the first school you attended?
 Where did you first get engaged?
 Where was your maternal grandmother born (mother's mother)?
 Where was your maternal grandfather born (mother's father)?
 Where was your paternal grandmother born (father's mother)?
 Where did you experience your worst holiday experience?
 Where is your favourite holiday destination?

Selection of Animal Images Used

Selection of Place Images Used

Selection of Flashbulb Images Used

A. REFERENCES

- [1] A. Adams, M. A. Sasse, and P. Lunt. Making passwords secure and usable. In H. Thimbleby, B. O'Conaill, and P. Thomas, editors, *People & Computers XII, Proceedings of HCI'97*, pages 1–19, Bristol, August 12-15 1997. Springer.
- [2] S. T. Allison, D. M. Mackie, and D. M. Messick. Outcomes biases in social perception: Implications for dispositional interference, attitude change, stereotyping and social behavior. In M. P. Zanna, editor, *Advances in Experimental Social Psychology*, volume 28. Academic Press, San Diego, 1996.
- [3] F. C. Bartlett. *Remembering*. London: Cambridge University Press, 1932.

- [4] J. Bonneau, M. Just, and G. Matthews. What's in a name? Evaluating statistical attacks on personal knowledge questions. In Fourteenth International Conference on Financial Cryptography and Data Security. Springer Verlag, 2010.
- [5] S. Brostoff and A. Sasse. Are passfaces more usable than passwords? A field trial investigation. In S. McDonald, editor, *People and Computers XIV Usability or Else! Proceedings of HCI 2000*, pages 405–424. Springer, 2000.
- [6] R. Brown and J. Kulik. Flashbulb memories. *Cognition*, 5(73-99), 1977.
- [7] L. R. Caporael and C. M. Hayes. Why anthropomorphize? folk psychology and other stories. In R. W. Mitchell, N. S. Thompson, and H. L. Miles, editors, *Anthropomorphism, Anecdotes and Animals*, chapter 6, pages 59–74. State University of New York Press, 1997.
- [8] C. B. Cave. Very long-lasting priming in picture naming. *Psychological Science*, 8(4):322–325, 1997.
- [9] R. Cellan-Jones. Embarrassment on Twitter, 26 February 2010. http://www.bbc.co.uk/blogs/thereporters/rorycellanjones/2010/02/embarrassment_on_twitter.html.
- [10] M. A. Conway, S. J. Anderson, S. F. Larsen, C. M. Donnelly, M. A. McDaniel, A. G. R. McClelland, R. E. Rawls, and R. H. Logie. The formation of flashbulb memories. *Memory and Cognition*, 22:326–343, 1994.
- [11] A. De Angeli, M. Coutts, L. Coventry, and G. I. Johnson. VIP: A visual approach to user authentication. In *Proceedings of the Working Conference on Advanced Visual Interfaces AVI. 2002*, pages 316–323. ACM Press, 2002.
- [12] R. Dhamija and A. Perrig. D'ej`a vu: A user study using images for authentication. In *Proceedings of USENIX Security Symposium*, pages 45–58, Denver, Colorado, August 2000.
- [13] H. Eichenbaum. Declarativememory: Insights from cognitive neurobiology. *Annual Review of Psychology*, 48:547–72, 1997.
- [14] H. Garfinkel. *Studies in Ethnomethodology*. Polity Press, 1984.
- [15] P. Graf. *Lifespan development of human memory*. MIT Press, 2002.
- [16] D. O. Hebb. *The Organization of Behavior*. New York: Wiley, 1949.
- [17] M. Jakobsson, E. Stolterman, S. Wetzel, and L. Yang. Love and authentication. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 197–200, New York, NY, USA, 2008. ACM.
- [18] W. James. *The Principles of Psychology*. New York: Holt, 1890.
- [19] M. Just and D. Aspinall. Personal choice and challenge questions: A security and usability assessment. In L. Cranor, editor, *SOUPS, ACM International Conference Proceeding Series*. ACM, 2009.
- [20] M. Just and D. Aspinall. Challenge challenge questions: An experimental analysis of authentication technologies and user behaviour. *Policy & Internet*, 2010.
- [21] N. Kock, R. Chatelain-Jard'ou, and J. Carmona. An experimental study of simulated web-based threats and their impact on knowledge communication effectiveness. *IEEE Transactions on Professional Communication*, 51(2):183–197, 2008.
- [22] N. Kornell and R. A. Bjork. A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, 138(4):449–463, 2009.
- [23] J. Liddell, K. V. Renaud, and A. De Angeli. Authenticating users using a combination of sound and images. In *HCI 2003, Bath, UK, September 2003. Short Paper*.
- [24] O. D. Oerlemans. The meanest thing that feels: Anthropomorphizing animals in romanticism. *Mosaic*, 27, 1994.
- [25] J. Podd, J. Bunnell, and R. Henderson. Cost-effective computer security: Cognitive and associative passwords. In *OZCHI '96: Proceedings of the 6th Australian Conference on Computer-Human Interaction (OZCHI '96)*, page 304, Washington, DC, USA, 1996. IEEE Computer Society.
- [26] A. Rabkin. Personal knowledge questions for fallback authentication: security questions in the era of facebook. In L. F. Cranor, editor, *SOUPS, ACM International Conference Proceeding Series*, pages 13–23. ACM, 2008.
- [27] K. Renaud. A visuo-biometric authentication mechanism for older users. In *Proc British HCI 2005. Sept 5-9, Edinburgh*, pages 167–182, 2005.
- [28] K. Renaud. Web authentication using Mikon images. In *World Congress on Privacy, Security, Trust and the Management of e-Business*, August 25 -27, 2009 Saint John, New Brunswick, 2009.
- [29] K. Renaud and E. Olsen. Dynahand: Observation-resistant recognition-based web authentication. *IEEE Technology and Society. Special Issue on Usable Security and Privacy*, 26(2):22–31, 2007.
- [30] S. Schechter, A. J. B. Brush, and S. Egelman. It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In *IEEE Security and Privacy*. IEEE, 2009.
- [31] S. Singh, A. Cabraal, C. Demosthenous, G. Astbrink, and M. Furlong. Password sharing: implications for security design based on social practice. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 895–904, New York, NY, USA, 2007. ACM.
- [32] S. L. Smith. Authenticating users by word association. In G. Papp and R. Posch, editors, *Proceedings of the Human Factors Society 31st Annual Meeting*, pages 135–138, Wien, 1987.
- [33] M. J. Stephey. Sarah Palin's E-Mail Hacked, 2009. <http://www.time.com/time/politics/article/0,8599,1842097,00.html>.