Passwords you'll never forget, but can't recall*

Daphna Weinshall
School of Computer Science and Engineering
Center for Neural Computation
Hebrew University
Jerusalem, Israel

Scott Kirkpatrick
School of Engineering and Computer Science
Hebrew University
Jerusalem, Israel
kirk@cs.huji.ac.il

ABSTRACT
We identify a wide range of human memory phenomena as potential certificates of identity. These "imprinting" behaviors are characterized by vast capacity for complex experiences, which can be recognized without apparent effort and yet cannot be transferred to others. They are suitable for use in near zero-knowledge protocols, which minimize the amount of secret information exposed to prying eyes while identifying an individual. We sketch several examples of such phenomena[1-3], and apply them in secure certification protocols. This provides a novel approach to human-computer interfaces, and raises new questions in several classic areas of psychology.

Author Keywords
Identity, human memory, passwords, security, adaptive interfaces.

ACM Classification Keywords
H5.2. User interfaces – Theory and methods.

INTRODUCTION
Imagine having a password or other method of certifying your identity that doesn't have to be consciously remembered, and can't be stolen or coerced from you. Not biometrics, such as fingerprints or iris patterns, which are observable (with special hardware) and can be copied, but something unobservable. The literature of psychophysics and cognitive psychology has classic studies of imprinting phenomena, which are quickly learned and can be recognized years later[1-9]. With imprinted memories, it is important to distinguish between recognition and recall. We easily recognize them as familiar, but to systematically recall what we have learned and transfer it to another ranges from difficult to impossible.

*ACM conference on Computer Human Interaction (CHI) 2004, Vienna, Austria, April 24-29.

Pictures provide an excellent example. Perhaps we can make a certificate out of imprinted images. Obviously the pictures can't be captured by external inspection – they are inside your head, and, at present, we don't even know where to look. Images are stored with little conscious awareness of what was learned, and are hard to describe. You will not be able to give another person such a stored certificate, even if you wished to do so. The novelty in the present work lies not in the use of pictures (e.g. "pictures replacing PINs")[10,11], but in suggesting that many natural characteristics of human memory can be exploited. Equally important is that these behaviors integrate naturally into cryptographic protocols. These give an evaluation of acceptance error, the likelihood that you are not who you claim to be, and minimize the danger of eavesdropping.

PASSWORDS
Today we certify ourselves to computers using a password or a numeric PIN code. You are completely aware of your PIN or password, so it is easy for you to describe it to others. You can be impersonated by someone who knows your password, and are not very safe from sophisticated eavesdroppers. The protocol used to verify a password is quite simple, and usually involves comparing an encrypted version of the password with a stored encrypted copy. The weakness is the difficulty of remembering all the passwords and PINs that modern life requires without writing them all down (unencrypted) and posting them in an obvious place or using easily-guessed personal information. But the security of passwords, based on the astronomical number of passwords that one can invent is often illusory, given the many "social" means available of obtaining a password by personal knowledge or eavesdropping, and the powerful tools[12] now available for guessing them.

FUTURE CERTIFICATES
In the future, we propose that certificates can be based on either recognition of complex memories or on detection of subtler "priming" effects. Human ability to recognize previous experiences is so effortless that one can learn a great number of things, and might need to demonstrate only a few of them to be identified. As a result, each remembered item need only be used once for certification, minimizing exposure to eavesdropping. Unlike recognition, unaided recall of memories, e.g. providing a password,

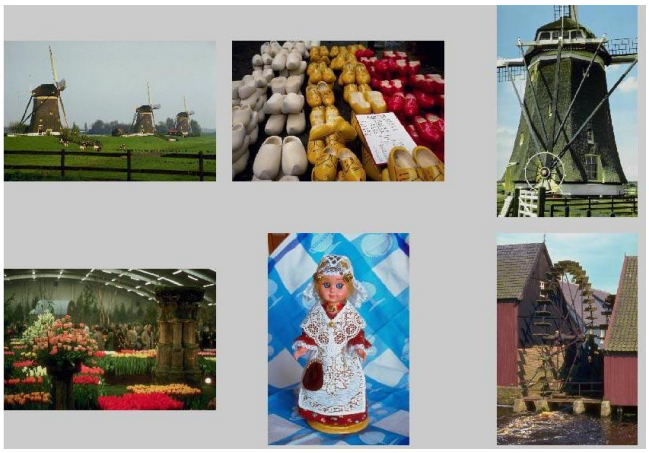

Figure 1. A picture group used in this study. One of these has been previously shown to the subject.

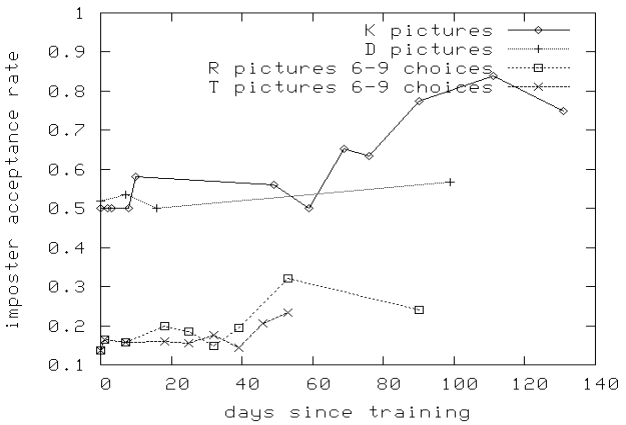

Figure 2. Effective rate per presentation at which a guessing imposter would be accepted in the picture recognition protocol (smaller is better).

requires conscious effort and appears to offer much less capacity. As we consider our third category, learned behaviors of which the user is unaware[8,9], we find another tradeoff. The subtler effects take longer to learn, and more complex protocols are required to detect their presence. In fact, we concluded that "priming" effects, for example learned skills such as very rapid perception of phrases or graphical icons, exhibit so much individual variation and are so sensitive to user fatigue or mood, that we could not create robust certificates with them at present.

We shall describe three cognitive phenomena that are useful for cryptographic certificates: picture recognition[1,4,5], pseudo-word recognition[3,7], and what might be called "language recognition," distinguishing grammatical from ungrammatical "statements" in an artificially-generated language consisting of strings of symbols[2,6]. All three are described in the relevant literature of perception and cognitive psychology. Since the existence of these effects is not controversial, our experiments have focused on understanding the best ways in which to train and test subjects to demonstrate the phenomena, and on quantifying the acceptance error that can be achieved using them.

PICTURE RECOGNITION

Most of our effort has gone into picture recognition. To build your certificate, we conduct a unique training session with you, in which you are shown a relatively large (100-200) set of pictures, randomly selected for you from a database of 20,000 pictures. The database is organized in small groups of 2 to 9 images on a common theme. One image from each group is selected for you. The training is self-paced – you can go back and forth through the randomly selected training set of images at will. The images are presented at the size which will later be used in testing. During authentication (and in our tests) you are shown several groups, and must select the one image in each group which was in the original training set. This is repeated a number of times, to defeat random guessing. To defeat eavesdropping, each group is used only once for certification purposes. Retraining is needed when the training set is exhausted. For an example of a picture group, see Figure 1.

To analyze the effectiveness of picture recognition as a certificate, we compare you with an imposter who has not been trained on the same specific images and can only guess. Let *n* denote the number of images shown side by side in each trial. In our studies, values of *n* from 2 to 9 were employed. Thus at most *log n* bits of information can be securely transmitted by each presentation, and an adversary would guess correctly *1/n* of the time. Your performance might also not be perfect, but can be distinguished from guessing after a few presentations. A certification application can operate by presenting images for recognition and stopping as soon as the chance that guessing would have produced the observed number of correct recognitions is reduced below a preset threshold, such as 0.01. A user making no mistakes reaches this threshold in a binary forced-choice protocol in seven presentations. If each presentation provides 6 images, however, this level is reached with perfect user performance in three presentations, and a fourth reduces the chance of an imposter succeeding to less than 0.001.

To assess the effect of user mistakes in recognition, we have scored our tests in terms of the probability that an imposter would do as well or better in a series of presentations as did the subject. In calculating this, we allow the imposter to make the same number or fewer errors as the subject, but at any step in the *p* presentations. If, in a multiple choice protocol, the subject guessed wrong at first but recognized the familiar picture on a later trial (in almost all cases where subjects made errors, a second try got the right answer) we allow the imposter to take the same number or fewer tries on some presentation in the test series. For a subject tested on *p* presentations, we take the *p*-th root of this probability as a per-presentation measure of the imposter's chances, a quantity lying between *1/n* and unity. The success of a particular protocol is measured by how small this quantity is found to be with actual subjects. Our studies are summarized in Figures 2 and 3.

Figure 2 describes recognition of photographs. Subjects using our final methodology were able to recognize previously seen pictures with better than 90% accuracy for one to three months. Three aspects of the procedure appear to have a visible influence on accuracy and retention: choosing picture groups with a clear theme but individual distinctions, the number of training sessions, and frequency of testing. At first, we used image pairs which were similar in most of their elements (e.g., two pictures of giraffes, one with two and the other with three giraffes). This proved more confusing than helpful to early subjects. Their performance, initially high, deteriorated to 70-80% after two months. When we selected pictures with a clear central subject or action and greater differences within the group, performance improved to that shown in Figure 2. Recognition accuracy when groups of 6-9 pictures were presented was just as good as with binary forced choice presentation, suggesting that increasing the number of choices was a good design decision. We see in Figure 2 that with multiple choice the imposter's acceptance rate is significantly less than was seen in the binary forced choice.

OBJECT RECOGNITION

After finding that pictures with a clear central subject or theme were more easily recognized, we also explored using a standard database of 260 artist-drawn images [13,14] of common objects. Results of having two subjects train and test on these are shown in Figure 3. This was somewhat less successful than using photographs. A possible reason is that the familiarity of the objects made them more confusing as distractors.

PSEUDOWORD RECOGNITION

When the memory and storage required for graphical objects or an adequate display facility are not available, a recognition protocol can still be designed with strings of letters. We studied recognition of previously seen pseudowords, generated by taking the list of common English words given in Wilson[15], and modifying them in one letter position using the program provided by van Heuven[16]. A native English speaker then selected pseudowords which are pronounceable, and do not exist as valid words. Our subjects, as shown in Figure 4, achieved lower accuracy levels on pseudoword recognition than we saw on pictures, varying between 70% to 90% over a three month period. While the picture recognition protocol is easy (and even fun) to use, and more or less universal across cultures, pseudowords are sensitive to the user's native language. Thus they are less comfortable to use and somewhat less reliable, but they can be used when pictures are not an option. As an example of a pseudoword test group, we might ask which of "frong," "polocy," "nevar," "cloar," or "lurther" is a familiar pseudoword. Results with three subjects are shown in Figure 3.

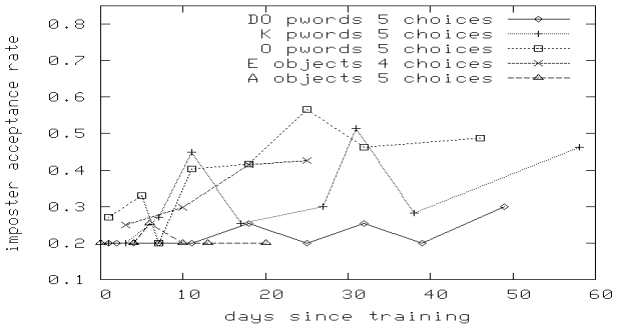

Figure 3. Results for object and pseudoword recognition in multiple choice paradigms.

ARTIFICIAL GRAMMAR LEARNING

The final example which we have implemented is based on the AGL (Artificial Grammar Learning) task first introduced by Reber[2]. In this paradigm, subjects first learn sets of approximately 20 short strings generated by a Finite State Machine (FSM). Reber's subjects could memorize "grammatical strings" which were generated by the FSMs more readily than they could learn truly random strings, yet the subjects were unable to articulate the patterns they had learned. We have implemented a version of this experiment with which to train subjects. We used strings of 3-10 characters, and an FSM with 8-10 internal nodes to generate them. For each user requiring a certificate, we created such an FSM at random. The strings are presented to the subject for identification paired with non-grammatical strings, which are generated from grammatical strings by the exchange of 2 arbitrary letters at interior locations in the string.

The following four strings are examples of our FSM's output: XVJTJVTJTV, XVJTHPHV, PJVTHPJHXJ, and PJVJJJHV. These were paired with ungrammatical strings such as PJXJTHXHV, XJVTPXV, XPHTHXTXPV, PVHJVJTPTJ. One subject who had studied 20-30 of the correctly produced strings was able to distinguish them from the ungrammatical strings with an accuracy of better than 90% for a short period of time but this declined to 75% accuracy after several weeks. Another subject did less well. We believe that tuning the design of distractors to make the discrimination a little easier would improve the results. In this certification protocol, an adversary overhearing the transactions might try to reverse-engineer the FSM. This is very difficult in principle, and is made even more so by the fact that the adversary sees information corrupted by the user's occasional mistakes.

PREVIOUS APPROACHES

Several other groups have exploited picture recognition for access control, generally viewing their techniques as a way of making passwords easier to recall. Thus Dhamija and Perrig[10] had the user select a small, fixed, group of pictures, then pick them out of a larger group. Researchers at Microsoft[11] used cued recognition of artificially generated Rorschach patterns to generate passwords. The

user is shown a set of pictures and asked to assign a word to each, keeping it secret. Letters selected from these words become the password for subsequent certification. The pictures provide cues to recall the chosen words, and thus the passwords. Both methods are vulnerable to eavesdropping.

CONCLUSIONS AND FUTURE WORK

We conclude that the innate human capability to capture effortlessly large amounts of everyday experience can be exploited to create a novel sort of computer-human interface. Identity and perhaps other intentions can be conveyed through a dialogue in which the computer accumulates the probability that it has identified the user and her intent correctly until it can safely act. Pictures prove to be the most effective tokens with which to conduct this dialogue. Pseudo "code" words or strings of letters can also be used, but their proper setting and training protocols will require further development. A "zero-knowledge" approach of never showing a picture group twice gives immunity from eavesdropping, but separate tests showed that when groups were reused, the subjects' accuracy improved. They did not confuse the distractors with the images on which they had been trained, and thus could use our methods for longer times without the need for retraining.

In our recognition-based, probabilistically evaluated "imprinted" certificates, the validity of the basic effects is not in doubt. The literature of cognitive psychology offers many additional human behaviors which may extend gracefully into strong protocols for identification. Our results to date with subjects have demonstrated the protocols' feasibility and identified several important issues for tuning the methods to be efficient and friendly.

This approach exercises a different aspect of human behavior than most highly explicit computer-human interfaces. We can also ask whether there are unique aspects to the computer's side of the interaction, in which the program is skeptical of the user's identity until enough evidence is amassed from actual performance. The calculus of probability which we use should be extendable to "skeptical" interfaces which would be appropriate for safety-critical or financially sensitive applications which must be careful not to respond hastily until the user has proved their identity and competence for the task at hand.

REFERENCES

1. R. N. Shepard (1967). Recognition memory for words, sentences, and pictures. *J Verb Learn Verb Behav* **6,** 156-163.

2. Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* **6**, 855-863.

3. A. Salaso, R. M. Shiffrin, and T. C. Feustel (1985). Building Permanent Memory Codes: Codification and Repetition Effects in Word Identification. *Journal of Experimental Psychology: General* **114(1),** 50-77.

4. L. Standing, J. Conezio, and R. N. Haber (1970). Perception and memory for pictures: single trial learning of 2500 visual stimuli. *Psychol. Sci.* **19**, 73-74.

5. Cave, B. C. (1997). Very long-lasting priming in picture naming. *Psychol. Sci* **8**, 322-325.

6. Perruchet, P. and Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General* 119, 264-275.

7. E. Tulving, D. L. Schacter, H. A. Stark (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory & Cognition* **8(4)**, 336-342.

8. G. Musen and A. Treisman (1990). Implicit and Explicit Memory for Visual Patterns. *J Exp Psychol Learn Mem Cogn* **16(1)**, 127-37.

9. A. Karni and D. Sagi (1993). The time course of learning a visual skill. *Nature* **365**, 250-252.

10. R. Dhamija and A. Perrig (2000). Déjà vu: A user study using images for authentication. In *Proceedings of the 9th USENIX Security Symposium*, 2000.

11. A press report is given at http://research.microsoft.com/displayArticle.aspx?id=417

12. http://www.atstake.com/research/lc

13. B. Rossion and G. Pourtois (2001). Revisiting Snodgrass and Vanderwart's object database: Color and Texture improve Object Recognition. Presented at *VisionScienceS*, Sarasota, FL, May, 2001. Abstract in *Journal of Vision*.

14. J. G. Snodgrass, and M. Vanderwart (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Learning, Memory & Cognition, 6:174-215.

15. M. Wilson (2003). MRC Psycholinguistic Database: Machine Usable Dictionary, Version 2.00. Rutherford Appleton Laboratory, Oxfordshire, England.

16. W. van Heuven (2003). Pseudo: a nonword/pseudoword generator. NICI, Univeristy of Nijmegen, http://www.nici.kun.nl/~heuven/.