Memorability of Alphanumeric and Composite Scene Authentication (CSA)
Passcodes Over Extended Retention Intervals

Korey Johnson  
University of Idaho

Steffen Werner

Current authentication strategies seek to increase security by requiring users to create more secure alphanumeric passwords. Unfortunately, the inverse relationship between alphanumeric password security and memorability prevents users from being able to create a password that is both secure and memorable. Graphical user authentication mechanisms have been explored as a means to maintain security while enhancing memorability of passcodes. Current approaches often use unrelated picture sets from which participants have to remember a subset, with mixed results. The study outlined in this paper seeks to further validate the Composite Scene Authentication (CSA) graphical passcode mechanism (Johnson & Werner, 2006). Extending retention intervals and increasing the variability of stimuli clearly demonstrated the superiority of CSA over alphanumeric passwords. In addition, we manipulated the mode of presentation (serial vs. composite) to assess the memorability of stimuli presented in different temporal formats. In the current study CSA passcodes consisting of nine categorical dimensions were compared to nine character alphanumeric passwords. Participants showed a strong advantage in passcode retention of graphical passcodes for both modes of presentation. This effect grew larger with increasing retention intervals. At the longest retention interval (6 weeks), only 10 (12%) participants were able to produce their alphanumeric password vs. 50 (60%) participants who were still able to produce the correct graphical passcode.

The majority of current user authentication mechanisms are based on alphanumeric passwords (Renaud & De Angeli, 2004; Jermy, et al., 1999). While high security demands long arbitrary sequences of elements, humans are good at remembering short and meaningful sequences of elements – thus pitting security and memorability of passwords against each other (Adams, et al., 1997; Adams & Sasse, 1999; De Angeli, et al., 2005).

The inverse relationship between memorability and security has led researchers to look for alternatives to alphanumeric passwords. Graphical authentication mechanisms, which use pictorial material as elements of the passcode, have been proposed to improve passcode memorability. This goal is achieved in two ways. First, meaningful visual information is more easily remembered than textual material (e.g., Standing, 1973). Second, graphical passcodes can use recognition memory for pictorial elements rather than the free recall required for alphanumeric passwords. In a recent paper (Johnson & Werner, 2006) we outlined an alternative graphical passcode system, Composite Scene Authentication (CSA), as a viable alternative to traditional alphanumeric password systems and other graphical passcodes, such as Passfaces™, by Real User Corporation, Déjà vu (Dhamija & Perrig, 2000), and the Visual Identification Protocol (VIP) system by De Angeli, et al. (2005, see Johnson & Werner, 2006, for a summary.)

Surprisingly, few studies on graphical passcodes show a substantial improvement in memorability over traditional alphanumeric passwords. One possible problem might lie in a misinterpretation of research on the picture superiority effect – memory for detailed visual information is not boundless, as often thought, but severely limited (e.g., O’Regan, 1992). Pictures do, however, allow rapid access to meaning and the gist of pictorial information can be processed very quickly. This gist is then combined with distinct visual elements to provide a rich memory representation (Potter, et al., 2002). Focusing on abstract visual material or unknown faces thus fails to take full advantage of the superior memorability of pictorial information.

In previous studies, when pictorial information was used it was presented in the form of separate and unrelated visual images. The isolated presentation of pictorial material might negate one of the main benefits of picture memory – the interaction of multiple objects in a scene. As research on mental imagery and picture memory has shown (Biederman, et al., 1974), interactivity and meaningful context often leads to an increase in processing and memorability of visual information (also see Hollingworth & Henderson, 1998, in press). In addition, previous studies on graphical authentication have mainly focused on short passcodes of 10-13 bits. While this passcode length is similar to the current 4-digit authentication codes of ATM machines, it is usually not considered a secure password length and might not tap the full potential of graphical passcodes.

The present study is a continuation of a program of research centered on CSA as an alternative graphical authentication system that seeks to increase memorability by presenting passcode elements as part of a meaningful, composite scene. The composite scenes currently in use consist of nine categorical dimensions, each of which can take any of 16 different values, resulting in a 36 bit passcode. This triples the bit-length of the De Angeli et al. VIP system. Because the passcode will be perceived as a scene, with the critical elements interacting with each other, each element of the passcode serves as a retrieval cue for the other elements. By separating the passcode into nine independent dimensions, the system is made scalable and the passcode length can be extended. The authentication procedure also takes advantage of the dimensionality of the passcode, relieving the user from having to remember sequential...
information about the order of visual elements. A preceding study (Johnson & Werner, 2006) established an advantage in memory for a prototypical CSA mechanism compared to an alphanumeric password of equal bit-length over relatively short retention intervals (up to 14 days). The study was somewhat limited in that only one combination of elements (passcode scene) was used with a small sample. In the current study, the maximum retention interval was extended to 6 weeks to investigate the persistence of the graphical passcode in human memory. We employed 10 different, randomly selected passcodes to eliminate any possible explanations of the effect based on the particular images used in the study. Finally, mode of presentation was manipulated to determine if part of the advantage in memorability for the CSA passcodes could be attributed to the composite nature of the pictorial material, as opposed to being attributed solely to the pictorial nature of the material itself.

METHOD

Participants
The participants in this study (71 females, 31 males, M<sub>age</sub>=22 years old) were 102 undergraduates and graduates drawn from the University of Idaho subject pool. All participants reported normal or corrected to normal vision and were compensated for their participation with course credit.

Materials
The images that were used as stimuli for this study are an extension of the locally created and compiled database of images used in previous CSA research (e.g., Figure 1). The database of images consists of 16 images in each of 9 categorical dimensions. The 9 categories of image dimensions are backgrounds, males, pets, other animals, objects, females, children, female pose, and child pose. For the encoding phase in the composite presentation condition, ten composite scene passcodes were created by randomly selecting one item from each category and combining them in such a way that there were meaningful interactions present between the individual elements of the scene. Participants in the serial presentation condition viewed the individual scene elements (categorical dimensions) one at a time rather than as a part of a composite scene. The alphanumeric stimuli in this study were 10 randomly selected strings of 9 characters from the hexadecimal character space. Hexadecimal characters were used in order to hold the bit-length of both graphical and alphanumeric passcodes equal at 36 bits. The alphanumeric strings were also presented either in composite format (all 9 characters in the string at once), or serially (one character at a time). A short story (Henry, n.d.) was used as a filler task between the encoding phase and first test phase.

The stimuli for the graphical recognition phase were 9 grids containing all 16 possible images in each category. For the alphanumeric recall phase, participants were prompted to enter the string of characters in a text box. The images, short story, and alphanumeric material were presented to participants on a computer screen via the experiment’s designated website.

![Figure 1](image-url)  
**Figure 1.** (a) composite image presented during the encoding phase, consisting of 9 critical elements: background, male, object, pet, other animal, female, female pose, child, and child pose (b) object element and distractors. (c) female element and distractors. (d) female pose element, consisting of the correct pose for the female in (a), and 15 distractor poses.

Procedure

*Enrollment and first test phase*

Participants were seated at a computer workstation and given a set of written instructions that outlined the four-part experiment. Session one was completed in groups ranging from 1 to 9 participants per group. After providing electronic consent, participants were presented with their respective alphanumeric and graphical passcodes. Half of the participants were presented the material serially, e.g., one character or image element at a time, and half were presented the material in composite format, e.g. the whole character string or composite scene at once. A wider range of composite images was used compared to Johnson & Werner (2006), but the basic level categories of elements used remained the same. In the serial condition each of the 9 elements was presented for 4 seconds each, whereas in the composite condition the composite was presented for 36 seconds. The total exposure time was thus held constant across conditions. Order of presentation across all conditions was counterbalanced to eliminate any order effects.

After presentation of both the alphanumeric and graphical stimuli, participants spent the next 30 minutes reading a short story. Upon initiation of the reading task, each of the 20 pages of the story was presented to the participant for 90 seconds, resulting in a total presentation time of 30 minutes.

At the conclusion of the reading task, the first test phase prompted participants to enter the alphanumeric password and to identify the graphical material they had been presented with during the first part of the experiment. To identify their CSA graphical passcode, participants had to select the target item out of a set of 16 items (15 distractors) from 9 successive image sets. Each set contained all 16 items of a category. Items were selected by clicking a radio
button directly beneath the image. Participants also answered 9 questions about the short story they read. The questions were presented in multiple-choice format with 16 possible answers. Participants selected their answer by clicking a radio button corresponding to an answer.

A small group of participants (N = 18) participated in the experiment online and completed the enrollment and first test phase from a remote location.

All participants completed test phases 2, 3, and 4 from remote locations online. The purpose of using two methods in administering this experiment was to establish whether future experiments in this line of research could be conducted purely online.

**Test phases 2, 3, and 4.**

The remaining test phases were designed to measure participants’ passcode retention after 1, 3, and 6 weeks. Participants were reminded to return to the experiment via email. For the final analyses data for participants that did not complete the appropriate test phase within 7 days after having received the reminder email were excluded. If a participant did not complete a test phase, but later returned and completed subsequent phases, the data was excluded for the analysis of the missed phase and included in the analysis of subsequent phases only if the subsequent phases were completed within 7 days of the target date. However, since participants were free to return over the course of multiple days, the exact retention intervals differ slightly between participants. Descriptive statistics regarding the number of participants who participated in each phase of the experiment and the ranges for each retention interval are included in Table 1. For these additional test phases, participants were simply prompted to identify the alphanumeric and graphical material they had been presented with during the first part of the experiment, as well as answer questions about the story that they had read. The procedure for each of these test phases was identical to the procedure described above for test phase 1.

![Table 1](image)

When computing the percent of alphanumeric information retained, it was necessary to compute both a strict and lenient interpretation because participants had the freedom to manipulate both number and position of characters in a text string. The strict interpretation of percent information retained $(\text{AS})$ was simply a binary decision of correct or incorrect for each of the 9 positions in the text string. The Damerau-Levenshtein edit distance (Damerau, 1964) was used to obtain a lenient interpretation of percent information retained. The Damerau-Levenshtein edit distance represents the minimal number of deletions, insertions, substitutions, and transpositions required to transform one string of text into another (i.e., an incorrect alphanumeric password into a correct alphanumeric password). Participants did not have the freedom to manipulate the number or position of images in the graphical recognition test. For this reason, errors by dimension can be unambiguously computed for this data.

Two three-factorial analyses of variance were conducted. The first analysis compared percent graphical information $(\text{G})$ retained to the strict interpretation of percent alphanumeric information retained $(\text{AS})$ across Modes of Presentation $(\text{MOP})$ and Retention Interval $(\text{RI})$. The second analysis was identical, with the exception that the Damerau-Levenshtein edit distance was used to calculate percent alphanumeric information retained $(\text{AS})$. The first analysis showed significant main effects for Type of Passcode $(\text{TOP})$, Retention Interval, and Mode of Presentation. The average percent of graphical information retained across retention intervals $(M_{\text{G}} = 93\%)$ was significantly greater than the average percent of alphanumeric information retained across retention intervals $(M_{\text{AS}} = 44\%)$. $F_{\text{G,AS}}(1, 75) = 252.338, p < .001$. There was also a significant difference between the percent of information retained at each retention interval $(M_{\text{G,AS}} = 78\%, 68\%, 66\%, \text{and} 64\%$ for retention intervals 1-4 respectively), $F_{\text{RI}}(2.105, 157.867) = 33.082, p < .001$. Finally, the analysis revealed a significant difference between percent retention of information presented in composite format $(\text{C})$ $(M_{\text{C}} = 74\%)$ compared to percent retention of information presented serially $(\text{S})$ $(M_{\text{S}} = 63\%), F_{\text{AS}}(1, 75) = 10.179, p = .002$.

These main effects were moderated by two 2-way interactions. Type of Passcode interacted with Mode of Presentation such that the presentation of information in serial format had a selectively negative impact on the percent of alphanumeric information retained $(M_{\text{AS,C}} = 56\%; M_{\text{AS,S}} = 32\%)$ compared to the percent of graphical information retained $(M_{\text{G,C}} = 92\%; M_{\text{G,S}} = 94\%), F_{\text{TOP,MOP}}(1, 75) = 17.957, p < .001$. Type of Passcode also interacted with Retention Interval such that an increase in the retention interval had a selectively negative impact on the percent of alphanumeric information retained $(M_{\text{AS}} = 63\%, 43\%, 38\%, \text{and} 34\%$ for retention intervals 1-4 respectively) compared to the percent of graphical information retained $(M_{\text{G}} = 92\%, 93\%, 94\%, \text{and} 94\%$ for retention intervals 1-4 respectively), $F_{\text{TOP,RI}}(2.202, 165.174) = 45.989, p < .001$.

Table 1. Number of participants that successfully completed each of the test phases and minimum, maximum, and average return times for each retention interval (RI). Data was excluded for participants failing to complete the appropriate phase of the experiment within 7 days of the target date.

**RESULTS**

When computing the percent of alphanumeric information retained, it was necessary to compute both a strict and lenient interpretation because participants had the freedom to manipulate both number and position of characters in a text string. The strict interpretation of percent information retained $(\text{AS})$ was simply a binary decision of correct or incorrect for each of the 9 positions in the text string. The Damerau-Levenshtein edit distance (Damerau, 1964) was used to obtain a lenient interpretation of percent information retained. The Damerau-Levenshtein edit distance represents the minimal number of deletions, insertions, substitutions, and transpositions required to transform one string of text into another (i.e., an incorrect alphanumeric password into a correct alphanumeric password). Participants did not have the freedom to manipulate the number or position of images in the graphical recognition test. For this reason, errors by dimension can be unambiguously computed for this data.

Two three-factorial analyses of variance were conducted. The first analysis compared percent graphical information $(\text{G})$ retained to the strict interpretation of percent alphanumeric information retained $(\text{AS})$ across Modes of Presentation $(\text{MOP})$ and Retention Interval $(\text{RI})$. The second analysis was identical, with the exception that the Damerau-Levenshtein edit distance was used to calculate percent alphanumeric information retained $(\text{AS})$. The first analysis showed significant main effects for Type of Passcode $(\text{TOP})$, Retention Interval, and Mode of Presentation. The average percent of graphical information retained across retention intervals $(M_{\text{G}} = 93\%)$ was significantly greater than the average percent of alphanumeric information retained across retention intervals $(M_{\text{AS}} = 44\%)$. $F_{\text{G,AS}}(1, 75) = 252.338, p < .001$. There was also a significant difference between the percent of information retained at each retention interval $(M_{\text{G,AS}} = 78\%, 68\%, 66\%, \text{and} 64\%$ for retention intervals 1-4 respectively), $F_{\text{RI}}(2.105, 157.867) = 33.082, p < .001$. Finally, the analysis revealed a significant difference between percent retention of information presented in composite format $(\text{C})$ $(M_{\text{C}} = 74\%)$ compared to percent retention of information presented serially $(\text{S})$ $(M_{\text{S}} = 63\%), F_{\text{AS}}(1, 75) = 10.179, p = .002$.

These main effects were moderated by two 2-way interactions. Type of Passcode interacted with Mode of Presentation such that the presentation of information in serial format had a selectively negative impact on the percent of alphanumeric information retained $(M_{\text{AS,C}} = 56\%; M_{\text{AS,S}} = 32\%)$ compared to the percent of graphical information retained $(M_{\text{G,C}} = 92\%; M_{\text{G,S}} = 94\%), F_{\text{TOP,MOP}}(1, 75) = 17.957, p < .001$. Type of Passcode also interacted with Retention Interval such that an increase in the retention interval had a selectively negative impact on the percent of alphanumeric information retained $(M_{\text{AS}} = 63\%, 43\%, 38\%, \text{and} 34\%$ for retention intervals 1-4 respectively) compared to the percent of graphical information retained $(M_{\text{G}} = 92\%, 93\%, 94\%, \text{and} 94\%$ for retention intervals 1-4 respectively), $F_{\text{TOP,RI}}(2.202, 165.174) = 45.989, p < .001$.

The second three-factorial analysis that was conducted using the edit distance to calculate percent of alphanumeric information retained showed very similar results. The effect sizes varied slightly, but all of the effects remained, as well as the direction of those effects.

In addition, a 2x2 ANOVA was conducted at each retention interval to compare the amount of graphical information retained to the alphanumeric information retained as well as percent information retained for serial versus composite presentation. For the first retention interval we found main effects for Type of Passcode as well as Mode of Presentation. The percent of information retained for the graphical stimuli $(M_{\text{G}} = 92\%)$ was significantly higher than the percent of information retained for the alphanumeric stimuli $(M_{\text{AS}} = 65\%, M_{\text{ASU}} = 73\%), F_{\text{AS}}(1, 99) = 80.523, p < .001; F_{\text{ASU}}(1, 99) = 44.714, p < .001$. The percent of
information retained for stimuli presented in composite format \( (M_G = 91\%, M_{AS} = 76\%, M_{AM} = 82\%) \) was also significantly higher than the percent of information retained for stimuli presented in serial format \( (M_G = 93\%, M_{AS} = 53\%, M_{AM} = 65\%) \), \( F_{G, AS}(1, 99) = 9.288, p = .003; F_{G, AM}(1, 99) = 7.084, p = .009 \). This main effect was moderated by an interaction between Type of Passcode and Mode of Presentation, \( F_{G, AS}(1, 99) = 16.170, p < .001; F_{G, AM}(1, 99) = 11.907, p = .001 \). Presenting stimuli in a serial format had a selectively negative impact on percent of information retained for the alphanumeric material (see Figure 2). The number of successful logins (completely correct responses) after the first retention interval was 53 for the graphical passcodes and 34 for the alphanumeric passcodes (see Table 2).

The same ANOVA was conducted for the second, third, and fourth retention intervals, again showing very similar results. Effect sizes varied, but the direction of relationships remained constant and significant.

For the first three retention intervals, a cursory analysis of the frequency of errors per dimension for the graphical passcode revealed that there were two specific dimensions of the graphical stimuli in which there were consistently more errors. 27\%, 21\%, and 17\% of participants failed to recognize the correct female in the first, second, and third retention intervals respectively. 31\%, 21\%, and 17\% of participants failed to recognize the correct child in the same retention intervals. The pet dimension was also causing a disproportionate number of errors over the longer retention intervals. By the fourth retention interval 13\% of participants could not recognize the correct pet.

**DISCUSSION**

The results clearly show that there is a strong memorability advantage for the CSA passcodes compared to the alphanumeric passcodes. Longer retention intervals, such as the maximum 6 week retention interval in this study, lead to a large decrease in the amount of information retained for the alphanumerical passwords (65\% to 30\%) but don’t seem to have much of an impact for graphical passcodes (92\% vs. 93\%). Similarly, the number of successful logins was five times as high in the graphical passcode condition (60\%) than the alphanumerical password condition (12\%). The interaction between Type of Passcode and Mode of Presentation illustrates that there is a distinct advantage for memorability of alphanumeric information presented in composite format. However, that advantage was not observed for the graphical information in this study. Further investigation is needed to determine whether the potential benefit for the composite graphical scenes might be more pronounced if the scenes are more meaningful than the ones designed for this study. In addition, the serial presentation of the items might also have increased the amount of attention devoted for each item.

The error analysis by image dimension clearly highlights a few potential sources of the majority of errors in recognition of the graphical stimuli. The manipulation of the pose of a target stimulus led to a dramatic increase in the number of errors associated with the recognition of that target in a different (neutral) pose. As one goal of the present study, we wanted to investigate whether a conditional dimension, such as the pose of a particular image element, might be used independently of the item itself as an additional passcode dimension. From these results it appears that the costs incurred by incorrect identifications of target items was too high for the chosen image dimensions. Future studies of CSA will therefore try to maximize the similarity of a presented object to its test item, instead of changing the pose or other elements about it. However, other image elements might still take advantage of conditional manipulations of particular properties to increase the bit-length of a graphical passcode without taking up more space on the display.

**REFERENCES**


