Reinforcing System-Assigned Passphrases Through Implicit Learning

Zeinab Joudaki
University of Ontario Institute of Technology
Oshawa, Canada
Zeinab.Joudaki@uoit.ca

Julie Thorpe
University of Ontario Institute of Technology
Oshawa, Canada
Julie.Thorpe@uoit.ca

Miguel Vargas Martin
University of Ontario Institute of Technology
Oshawa, Canada
Miguel.Martin@uoit.ca

ABSTRACT

People tend to choose short and predictable passwords that are vulnerable to guessing attacks. Passphrases are passwords consisting of multiple words, initially introduced as more secure authentication keys that people could recall. Unfortunately, people tend to choose predictable natural language patterns in passphrases, again resulting in vulnerability to guessing attacks. One solution could be system-assigned passphrases, but people have difficulty recalling them. With the goal of improving the usability of system-assigned passphrases, we propose a new approach of reinforcing system-assigned passphrases using implicit learning techniques. We design and test a system that implements this approach using two implicit learning techniques: contextual cueing and semantic priming. In a 780-participant online study, we explored the usability of 4-word system-assigned passphrases using our system compared to a set of control conditions. Our study showed that our system significantly improves usability of system-assigned passphrases, both in terms of recall rates and login time.

CCS CONCEPTS

• **Security and privacy** → *Authentication*;

KEYWORDS

Authentication; Passwords; Implicit Learning; Usable Security; Cued-Recognition

ACM Reference Format:
Zeinab Joudaki, Julie Thorpe, and Miguel Vargas Martin. 2018. Reinforcing System-Assigned Passphrases Through Implicit Learning. In *2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18), October 15–19, 2018, Toronto, ON, Canada.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3243734.3243764

1 INTRODUCTION

The popular use of passwords that people choose is controversial—people tend to choose the same or similar passwords across multiple accounts, many of which have been leaked in password breaches. The wealth of password data that is publicly available has been shown to enable targeted guessing attacks that successfully guess over 32-73% of passwords within 100 attempts [44]. This motivates other approaches to user authentication.

A passphrase is a collection of words used instead of a typical text password, intended to increase password length and therefore security, while retaining memorability [24]. Passphrases are more memorable than passwords when they contain meaningful expressions or follow a sentence/gramatical structure [23, 24, 38, 46, 49]; therein lies the security drawback. Bonneau and Shutova [7] studied the linguistic properties of Amazon Payphrases for 100,000 people. They found that if an adversary were to use lists of popular books and movies, plus natural language bigrams, they could successfully guess many of these phrases [4, 7].

System-assigned passphrases were proposed due to their many security advantages, including resistance to targeted impersonation, throttled guessing, unthrottled guessing, and leaks from other verifiers [5]. Unfortunately, the use of system-assigned passphrases comes at a cost to memorability. It has been suggested that if people paired a system-assigned passphrase with a story, it would improve memorability [30]; however, studies indicate that this is not a successful strategy [36]. This motivates other approaches to improve the memorability of system-assigned passwords. Spaced repetition has been used to improve passphrase memorability, but at the cost of a long training time [6]. Using multiple verbal and graphical cues has also yielded memorability improvements, but the login times remain long [2].

We propose an approach we call *Implicitly Reinforced Passphrases* to improve memorability for system-assigned passphrases using implicit memory techniques. The essence of the idea is to reinforce the passphrase using a short implicit learning phase during enrollment, in order to involve both implicit and explicit memory processes. Our design (CC-SP) employs two implicit learning techniques: contextual cueing and semantic priming. The system design also aims to reduce input errors and long login times associated with other passphrase systems [6, 24, 36].

We design and implement a system using this approach with 4-word passphrases, intended to offer resistance to online guessing attacks. This is of interest as previous studies on system-assigned PINs and passwords with similar security have been found to be ineffective. While user-chosen PINs and passwords have reasonable usability [49], studies show that system-assigned PINs and passwords do not. For instance, in an Amazon Mechanical Turk (MTurk) study, Huh et al. [19] found 6-digit system-assigned PINs to have a 65% success rate 2 days after it was assigned with mean login time of 41.7 seconds. In another MTurk study, Shay et al. [36] found a 56% login success rate for system-assigned five-character passwords. Their results showed a mean login time of 27.5 seconds

CCS '18, October 15–19, 2018, Toronto, ON, Canada

ACM ISBN 978-1-4503-5693-0/18/10.
https://doi.org/10.1145/3243734.3243764

2 days after the password was assigned. We evaluate our proposed system through a 780-participant online MTurk study involving five control conditions. The results demonstrate that our design offers significant memory improvements (88% login success rate one week later), with short mean training and login times (64 seconds and 13.74 seconds, respectively). Our results also suggest that the improvement can be attributed to the implicit learning techniques we employ. Participants reported high levels of satisfaction with this system as 77% of them would like to use our system frequently.

Contributions. The contributions of this paper are as follows:

(1) We present a novel approach for system-assigned passphrases that harnesses implicit memory with the aim of improving memorability.
(2) Through an online user study with 780 participants on Amazon Mechanical Turk, we show that an implementation of this approach outperforms our control conditions in terms of recall rate, login time, and storage behaviour.

Impact. The design we present and study in this paper could potentially be useful in a variety of settings that seek protection against online attacks, such as banking, email, social networking, and e-commerce systems. It could be used as a single factor in some environments, or as a second knowledge-based factor in high security environments. The general approach we propose of using implicit learning to reinforce explicit memory in authentication can be applied to many different designs, intended for systems with different security and usability requirements. We anticipate that our findings will stimulate research into the use of implicit learning in other authentication systems.

2 RELATED WORK

2.1 System-Assigned Secrets

Shay et al. [36] investigated the potential of using random system-assigned passphrases as opposed to randomly assigned characters, by encouraging users to imagine a scene that links each word. Their results were unfortunately not very positive; only 51% could recall the passphrases after 2-5 days. Wright et al. [45] compared usability of three types of system-assigned passwords (Word Recall, Word Recognition, and Letter Recall), whereby users entered their passphrases by selecting their assigned words from series of displays containing lists of candidate words. They did not find any memorability differences between the three groups, which confirms that not all forms of memory cues are effective. Bonneau and Schechter [6] examined the impact of a training period of a few weeks that employed spaced repetition for a random password. The findings were that 88% of users were able to recall their passphrase after 3 days, however the training period was quite long (about 12 minutes over the course of 10 days on average) for memorizing the full 56-bit secret. In another study Haque et al. [18] employed users' spatial and visual memory to improve memorability of system-assigned passwords. They provided users a training session enabled by the method of loci using videos to help them memorize their assigned passwords. They achieved 86% recall success rate for a login session one week after the training. However, it had a long registration time of 160 seconds.

Al-Ameen et al. [2] proposed the use of graphical cues to aid recall of system-assigned passphrases; pilot studies showed this method holds promise as all 11 users recalled their phrase after one week. In another study, Al-Ameen et al. [1] argued how leveraging users' cognitive abilities through different spatial and verbal cues can improve memorability of system-assigned graphical passwords. In an in-lab study with 56 participants, 98% of their participants were able to successfully login one week after the training. The median login time for the proposed scheme was between 35 to 51 seconds.

2.2 Implicit Learning-Based Authentication Schemes

A number of papers have explored employing implicit memory for authentication due to its potential for more usable yet secure authentication keys.

Denning et al. [14] proposed an authentication scenario which employs a priming effect as a mechanism using implicit memory. Their suggested image-based authentication system used pairs of images; that is, complete and degraded counterpart images. They initially showed sets of complete images and for later authentication, degraded images are exposed through a familiarization task. As per their user study results, they found that the median image labeling time is 3.9 seconds which would result in a password recovery time of 8.8 minutes. Their results also showed that 45.8% of users were able to label images they were primed and 38.8% of users were able to label images they were not primed. This showed a non-significant effect for the priming they used. Moreover, the requirement to provide a large set of images makes the system less deployable.

Bojinov et al. [3] offered an authentication scheme with the property that users are unaware of their secret and thus cannot easily communicate it. Their scheme used the Serial Interception Sequence Learning (SISL) task originally introduced by Sanchez et al. [33]. Subjects were trained to implicitly learn a random key sequence using a game similar to the Guitar Hero video game. After a 30 to 45 minute training period, they were tested through a session of playing the same game. Only 71%, 47%, and 62% of participants could successfully authenticate using this method immediately, 1 week, and 2 weeks later respectively.

Castelluccia et al. [10] proposed an implicit learning approach based on Mooney images [29]. A Mooney image is a low-information two-tone representation of an image which is hard to label unless the user was primed with the original image. Castelluccia et al. [10] achieved 0.1% False Acceptance Rate (FAR) and a 97.14% True Acceptance Rate (TAR) at the cost of an average authentication time of 3.5 minutes.

Joudaki et al. [21] explored the feasibility of Tacit Secrets, which are system-assigned passwords that you can remember, but cannot write down or otherwise communicate. Their approach is based on the Contextual Cueing (CC) implicit learning method. The findings indicated that the approach has resistance to offline attacks, online attacks, phishing attacks, some coercion attacks, and targeted impersonation attacks. The approach was found to have an average authentication time of 2.5 minutes, 14.5 minutes training time, 0.8% FAR, and 90% TAR. Their use of CC differs from ours as we are: (i) not using traditional CC displays involving 'T' and 'L' characters

(see Figure 1), but words instead (one target passphrase word surrounded by distractors), (ii) using Semantic Priming (SP) as well as CC. Additionally, our approach is designed to have much shorter training and login times.

3 SYSTEM DESIGN

We design our system using two implicit learning techniques: Contextual Cueing (CC) and Semantic Priming (SP) to train users to implicitly learn a system-assigned passphrase. The goal is to facilitate the memorization process of system-assigned passphrases and make it easier to recall them later by harnessing implicit memory to reinforce memory.

3.1 Explicit vs. Implicit Memory

Implicit learning occurs through the repetition of a specific task. Implicitly learnt skills are acquired unconsciously, unintentionally, and without declarative knowledge about what has been learnt [25, 32, 39]. Implicit learning is associated with complex features or probabilistic patterns, whereas explicit learning is most probable when involving salient stimuli [48]. Implicit learning is used in different areas such as perceptual-motor skills, language acquisition, social intuition, or detecting a target in a complex scene [39].

3.2 Implicit Learning Techniques

3.2.1 Contextual Cueing. CC is a mechanism [12] through which visual attention can be guided by implicitly learnt knowledge [37]. CC was first developed by Chun and Jiang [12] to study implicit learning and memory. To provide insight into the process, consider that objects and events occur in a rich visual context, aiding their recognition. For example, we may need to identify a traffic signal amongst an array of information in a busy street. Such a search might be facilitated by repeatedly seeing that on traffic signals the red light is most often at the top and the green light at the bottom. In CC, a context can be defined as a 2-dimensional spatial configuration of irrelevant objects (aka. *distractors*) in which a target is presented. In effect, CC relies on distractor positions to provide spatial cues for the location of a target. The entire context is shown on a display, for a fixed period of time. For a person to learn the context, he or she must see it repeatedly and locate the target (typically 4-6 times [12, 13]). Implicit learning of the context is then verified by observing that it takes less time to locate the target than for previously unseen contexts. Figure 1 indicates a sample display for a typical CC task.

3.2.2 Semantic Priming. Semantic memory is often described as humans' acquired, structured record of facts, meanings, concepts, word naming, lexical decisions, generic knowledge about the external world, and semantic priming [22]. Priming is an improvement of performance in a cognitive or perceptual task, relative to an appropriate fact, produced by context or previous experience [26]. In semantic priming, a target word (such as dog) is preceded by a semantically-related prime word (such as cat), it is processed more quickly and efficiently than when preceded by an unrelated prime (such as book) [26, 28, 31]. In 1971, Meyer et al. [28] had subjects deciding whether two strings of letters (i.e., word-word) are both words or not. When the words are semantically related the average response time is 85 milliseconds faster compared to unrelated pairs.

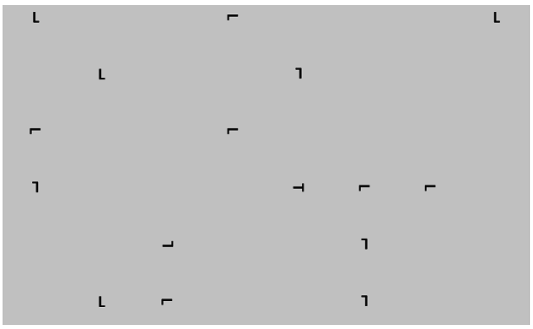


Figure 1: Illustration of a traditional CC display. The 'T' is the target, and the 'L's are distractors.

Semantic priming results in the improvement in speed and accuracy to respond to a stimulus (e.g., word, picture). For many years of research, the semantic priming paradigm has been used as a tool to improve understanding the organization of the mental lexicon and word retrieval from long-term memory [27].

3.3 Design Specifications

The essence of Implicitly Reinforced Passphrases is to provide a short training phase during enrollment to invoke implicit learning, then subsequent logins proceed normally without additional training. In our design, the training is enabled by a combination of two implicit learning based paradigms, CC and SP. Therefore, we call our system design CC-SP.

Display Design - CC. Users are assigned a 4-word passphrase and each word is presented in a display surrounded by 31 semantically-related words. We involved CC in the proposed approach by creating word displays, which have spatial configurations of words preserved on them. These four displays are shown to the users repeatedly. Each display contains a word of the passphrase and the user task is to find a word (i.e., passphrase word) with different font (see e.g., Figure 3a). To decide how many repetitions are required for a stable training, we referred to previous studies, confirming that CC knowledge is accessible after four repetitions [8, 17]. We ran some qualitative pilot studies with 10 participants to primarily evaluate the effectiveness of our scheme and finalize certain design decisions. We tested multiple display configurations in order to find if users are able to process the words effectively given the screen size, number of words, and time limit to look at each display. We also asked users if the provided instructions for performing the task are sufficient and informative. Through these rounds of pilot testing, we found that five repetitions provide sufficient training in order to learn the passphrase. In terms of the number of words shown on each display, 32 words, we also tested different numbers of words on 6×11 displays during pilot studies. We found that increasing the number of words per display prevents users from properly processing the relationship between the words and could cause distraction. We found that 32 words placed in a 6×11 matrix, is a number that can be processed for both the location of the items as well as the semantic relation between the words. These four displays (one for each assigned word) are shown five times for five seconds, each. If the user does not find the target word within five

seconds, the next display appears. Limiting the time to find the target is used in all previous CC study designs [12, 13, 35, 41]. We randomly selected each display's words from a dataset of 923 different words in which we have sets of words with close similarities (between 0.4 to 1) from each other.

Display Design - SP. SP is included by having each display's 32 words (distractors and target) semantically related. Note there is no semantic relationship between each of the 4 displays, but rather between the 32 words on each display. The provided displays are intended to help users to make mental associations (using their semantic memory) for these words through the task of searching for the target word. Studies of SP have observed that a response to a target is faster when it is preceded by semantically-related primes [26, 28, 31]. The priming occurs because the provided primes activate the viewer's mental encoding of related words or concepts, facilitating their later processing or recognition. The goal of this design is to encourage such mental relations to prime them to recall the assigned passphrase later. To generate the semantically related word sets, we used word2vec [16], which provides an efficient implementation of the continuous bag-of-words for computing vector representations of words. The word2vec tool takes a text corpus as input and produces the word vectors as output. It first makes a vocabulary from the training text data and then learns a vector representation of the words. These representations can be subsequently used in many natural language processing applications. By finding the distance of word pairs, using a distance tool [16], we can find the similarities between the words. Using this tool, given the similarities between the words (in a range of -1 to 1), we selected a set of 32 words with equal similarities (between 0.4 to 1) from each other. If we only consider the distance of each word from the target word, then the target word can be easily guessed by the attacker as all words have a relation with that word and not necessarily each other; however, making this relation between all words on the display prevents the target word from being computable. For our study we use the same dataset of 923 distinct words, and generate 40 sets of semantically related words for the purpose of the study. We note however, that with larger word datasets that more sets of semantically related words would be possible. Even with only 40 word sets, it is possible to generate 2^{21} sequences of displays ($P(40, 4)$), and 2^{41} distinct passphrases (as each display has 32 words to choose from).

3.4 Training Phase

The goal of training is to ensure the user implicitly learns a 4-word passphrase; this is accomplished during account registration.

We randomly generated 10 passphrases for the study, which were used in each of the conditions. Within each condition, the passphrases were randomly assigned to the participants. Having a fixed set of passphrases avoids any potential effect on the participants' performance which could be due to different levels of difficulty.

Each word of the passphrase is presented on a display and surrounded with 31 distractor words. The user task is to find a word which has different font than the other words (i.e., the passphrase word) and click on that word. The user is provided with feedback on her response with a border appearing around the display which is either green (when the correct word is clicked) or red (when an incorrect word is clicked). Figure 3a shows an example of a display that participants are exposed to in the training phase. The sequence of 4 displays is shown 5 times for at most 5 seconds per display. Figure 2 depicts the training phase with sample word displays.

3.5 Login Phase

At login, participants are provided with the four 32-word displays, used in the training phase where the target word is no longer in different font (see Figure 3b). The user needs to find the target (passphrase) word and click on it. Once the word is clicked, the next display is presented to the participant to find the next passphrase word.

3.6 Security Analysis

As mentioned above, system-assigned passwords offer many security advantages, including resistance to targeted impersonation, throttled guessing, unthrottled guessing, and leaks from other verifiers [5]. Since CC-SP is a system-assigned authentication system, it provides these benefits. The following discussion related to the key space and phishing resistance can be made without an analysis of user data, as the system is not subject to issues concerning user choice.

3.6.1 Key Space. Given that cues are provided for the login session, for a successful online attack, since there is a sequence of four displays each containing 2^5 possible words, $(2^5)^4 = 2^{20}$, the expected number of guesses is $2^{20}/\ 2 = 2^{19}$. We decided on a 4-word passphrase as it provides a keyspace of 2^{20} ($\gtrapprox 10^6$) which has negligible risk of online attack [15]. This has been the value cited for online attack resistance in many subsequent security solutions (e.g., [42, 44]). Florêncio et al. [15] discuss strength beyond 2^{20}, concluding that keyspaces between 2^{20} and 2^{47} fall in the "don't care region" or the online-offline chasm, whereby little is gained in terms of security, but the cost to usability can be noteworthy. Thus, we designed our system to have a 2^{20} keyspace. In Section 8, we discuss possible extensions for future work that can be studied to increase the keyspace further if desired.

3.6.2 Phishing Resistance. Our design aims to complicate classical phishing attacks, wherein the attacker creates a phishing site that mimics the CC-SP login process to harvest user passphrases. The attacker's goal is to trick the user into responding to the provided challenges and thus find the passphrase words. In Section 8, we discuss an extension to this design that would offer resistance to targeted phishing attacks as well. However, this protection does not extend to man-in-the-middle attacks.

In a classical phishing attack, to lure a user to click on their passphrase words, the attacker would need to present the user with the correct displays. However, the number of possible sequences of displays is quite large. For CC-SP, we have the semantic relation of the words to be considered on each display (SP), and 66 possible locations of the 32 words of each display (CC).

The worst case scenario is that the user does not recognize that the target word is in a different location; even in this case, with our 40 SP word sets, we expect it would take 2^{20} phishing challenges from the attacker on the same target user i to successfully recover

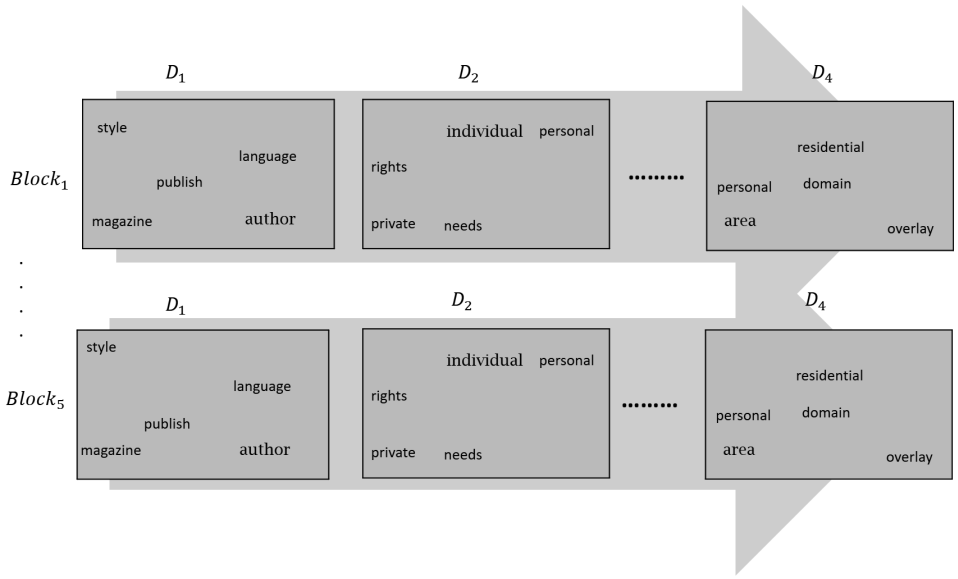
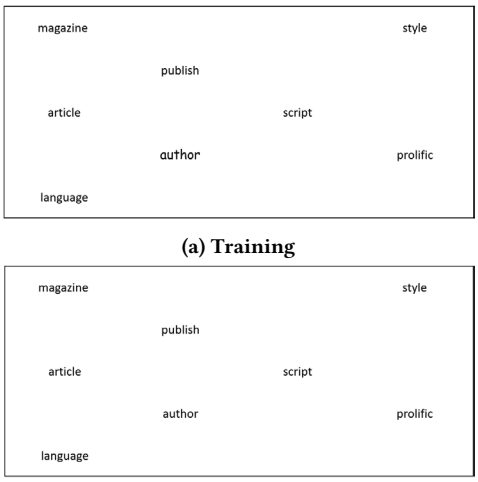


Figure 2: An example of displays arrangement during the training sessions.



(a) Training

(b) Login

Figure 3: Simplified CC-SP condition display for training and login. Notice that with the training phase shown in (a), the target word (author) is shown in different font. Please note that in our actual experiment each display contains a target word (passphrase word) which is surrounded by 31 distractor words (see Appendix A).

i's passphrase. If the user is able to detect the display's configuration is different, then even more are required.

3.7 Server Storage

We propose storing each user's images, and a salted hash, where e.g., JavaScript on the client end computes a word from the normalized click location on the image, and sends the words as though they were typed. The server then salts and hashes the sequence for comparison.

4 STUDY DESIGN

On a high level, our design trains users to implicitly learn a secret word on each of a set of displays which becomes their system-assigned passphrase. We designed an experiment through which users are provided with a training phase, designed to evoke implicit learning, for system-assigned passphrases.

4.1 Hypotheses

Our high-level research question is: *Can implicit learning improve (or reinforce) memorability of system-assigned passphrases?* Using implicit learning mechanisms, we expect the provided training would result in improved memorability compared to the other conditions which do not involve any training. The following statements articulate our hypothesis:

$H_{memorability}$: There will be significantly greater memorability in implicit learning-based trained passphrases compared to control conditions. To test this hypothesis, the following hypotheses should be tested:

- $H_{memorability\_recall}$: There will be a significant improvement in the number of users who correctly recalled their assigned passphrase words.
- $H_{memorability\_record}$: There will be a significant improvement in the mean number of users who recorded their passphrase.

$H_{usability}$: There will be significantly greater usability in implicit learning-based trained passphrases compared to the control conditions. More specifically, we consider the following hypotheses for implicit learning-trained passphrases vs. the appropriate control conditions (see next section for a discussion of which control is used for each implicit learning condition).

- $H_{usability\_logintime}$: There will be a significant improvement in time required to login.
- $H_{usability\_perception}$: There will be an improvement in users' perception regarding the system.

4.2 Study Conditions

Given the high-level research question (i.e., can implicit learning improve memorability of system-assigned passphrases?), our specific approach to answering this question is explained below. In particular, we examine the use of specific processes known to invoke implicit learning. Thus, our research examines the following more specific question:

Can we improve memorability of system-assigned passphrases using a combination of CC and SP for training users on their assigned secrets? To answer this question, we provide the following conditions:

Condition 1 (CC-SP). The first experimental condition provides participants with a training session that presents semantically-related words in repetitive-stable contexts (i.e., similar to the contexts used in CC). This condition aims to determine if the combination of CC and SP can improve passphrase memorability. For the training session, each word of the assigned passphrase is presented on a display containing 31 other semantically-related words. Each user is shown four displays (for a 4-word passphrase). The user task is to find a word with different font than other words (i.e., the passphrase word) and click on that one. During training, the sequence of 4 displays is shown 5 times.

For participants in this group, a login session is set up in a way that they will be provided with the same displays as their training session (to be used as cues). The only difference is that the target word is no longer in different font (see Figure 3b).

Condition 2 (Basic Passphrase Control). Participants in this group are assigned a passphrase with no training involved. They are given unlimited time to memorize their assigned words. This group's participants are later asked to recall their assigned passphrase by typing the four words in four text boxes. Comparing this with Condition 1 (CC-SP) can tell us whether our CC-SP approach has been effective in improving system-assigned passphrase memorability. It is worth noting that the keyspace for system-assigned passphrases is far larger than CC-SP; however, we included this condition to compare our special training without having any training at all. Our goal is to evaluate memorability of the same tokens independent of training and login interface. The remaining conditions are used to identify how effective each implicit learning technique, and other elements of our interface have been in improving memorability.

Condition 3 (CC). This condition is the same as Condition 1 (CC-SP), but there are no semantic relations between the words. In other words, the displays contain unrelated words. Displays shown to the participants in this group are as shown in Figure 3a; however words are random with no semantic relations. We provide the previously seen displays as cues for the login sessions. Comparing this condition with Condition 1 (CC-SP) will indicate whether any improvement offered by CC-SP would also be offered by CC alone.

Condition 4 (SP). This condition is the same as Condition 1 (CC-SP), but there is no repetition of the displays during training, and no fixed locations for the words. Since there is no repetition and no stable locations, there is no CC in the training. The 32 words are randomly placed in a display with 66 possible positions. We provide these words (in shuffled locations within the display) as cues for the login sessions. In training, each display of 32 words is shown once, and the user task is to find the word with different font and click on it. This condition is included to see if users have higher recall rates when semantically-related words are provided for their login sessions. Comparing this condition with Condition 1 (CC-SP) will indicate whether any improvement offered by CC-SP would also be offered by SP alone.

In the event of any memorability improvement for system-assigned passphrases while using our implicit learning-based interfaces, is it due to our special implicit learning-based training or it is just due to repetition and/or recognition?

To answer this question, the following control conditions need to be evaluated and compared with the previous conditions.

Condition 5 (Repetition). The condition is the same as Condition 1 (CC-SP) and 3 (CC) in terms of the displays (each containing one passphrase word) having the same number of repetitions; however, there is neither semantic relation between words, nor stable location of the words. For this condition, users are provided four consecutive displays where each contains a passphrase word that is surrounded by 31 random words (i.e., no semantic relation exists). As in all other training conditions (except Condition 2), users are supposed to find a word which has different font. For the login session, the users are provided with the displays that have no preserved locations for the words. Comparing this condition with the CC condition indicates whether the combination of repetition and recognition can be the source for improved memorability rather than CC. Comparing this condition to CC-SP indicates whether the combination of implicit learning techniques can be the source for improved memorability rather than repetition and recognition.

Condition 6 (Recognition). The condition is the same as Condition 5 (Repetition); however, there is no repetition involved for the training. For this condition, users are provided with four consecutive displays where each contains a passphrase word surrounded by 31 random words (i.e., no semantic relation exists). In the training phase, for each of the four displays, users are tasked with finding a word which has different font than the other words. Each display contains one passphrase word, and is shown only once. For the login, the users are provided with the displays to see if they can recognize their passphrase words. Note that for each login, each display has a random configuration of the same 32 words. Comparing this condition with SP indicates if recognition is the reason for memorability success rather than SP. Table 1 indicates how each condition includes CC and/or SP.

Table 1: Each condition specification. Some of the conditions include fixed location of the words, repetition, exposure time, and/or the words with semantic relation. CC includes fixed locations, exposure time, and repetition.

Condition	Semantic Relation (SP)	Fixed Location	Repetition	Exposure Time	Login Cues
		CC			
CC-SP	✓	✓	✓	✓	✓
Control	-	-	-	-	-
CC	-	✓	✓	✓	✓
SP	✓	-	-	-	✓
Repetition	-	-	✓	✓	✓
Recognition	-	-	-	-	✓

4.3 Study Structure and Instructions

We first tested our approach through a web application pilot study where we asked 10 participants to test our designed web application and provide us with their feedback. Using their feedback, we were able to further improve the design of our system. Their comments helped us to modify the instructions that participants were provided.

We used the MTurk crowdsourcing service to evaluate our conditions. All our user studies were reviewed and approved by our institution's Research Ethics Board. We first recruited 100 participants through MTurk to evaluate the feasibility of our proposed scheme. Of the 100 participants, 50 were assigned the control condition and the remaining 50 CC-SP. As we were using semantic relations of English words, we needed to maximize the chances that our participants knew English well; thus, we limited the participants to be from English-speaking countries. The first phase of our online study confirmed the effectiveness of CC-SP for memorability improvements for the users. Thus, we started the second phase of our study and recruited another 780 participants and randomly assigned them one of our study conditions. We compensated them 50¢ for completing the first session of the study and two additional 25¢ for completing the second and third sessions. The paid amounts for this study are common in MTurk for studies (e.g., [19, 36]) with similar durations. Our payments of 50¢, 25¢, and 25¢ had the average hourly rate of $14.48 which is more than the USA nationwide minimum wage rate ($7.25 in 2018). Our average training time was 64s; login times were 14s, 22s, and 31s for the first, second, and third login sessions respectively; and questionnaire times were 68s each.

Once each participant signs up for the study and consents, he/she is randomly assigned to one of the six study conditions. The participant is then assigned a 4-word passphrase. Depending on the condition, the participant will be provided with either no training or one of our six training sessions (see Section 4.2). In the first session of our study, we provided the participants with the same statement as Shay et al. [36] used in their study. "Imagine that your main email service provider has been attacked and that because of the attack, your email service provider is also changing its password rules. Instead of choosing your own password, you will be assigned a 4-word passphrase."

Participants in the Control condition were provided with the following instructions through three consecutive web pages: (1) "Below you can see a sample that shows the four words of an assigned passphrase.", (2) "Next you will be shown the 4 words of your passphrase. Please take the time you need to memorize your passphrase words (and their order).", (3) "The training session will begin next. After the training, you will be asked to login with your passphrase." For the Control Condition's participants, by training session, we meant the time given to the users to memorize their passphrase and no special training mechanism was involved. For participants in the SP, and Recognition conditions wherein no repetition was involved, the following instructions were provided: (1) "Each word of your passphrase will be presented in a grid of words. This word is shown in different font. Below you can see a sample that has the word with a different font circled in red. Note that in your task, these words will not be circled in red as in this sample.", (2) "When you find the word with different font, click on it. Notice that the table border provides you the feedback based on your response. Practice on the display below.", (3) "The training session will begin next. After the training, you will be asked to login with your passphrase." For participants in the CC-SP, CC, and Repetition the following instructions were provided: (1) "Each word of your passphrase will be presented in a grid of words. This word is shown in different font. There is a time limit of five seconds for each arrangement of words. Below you can see a sample that has the word with different font circled in red. Note that in your task, these words will not be circled in red as in this sample.", (2) "When you find the word with different font, click on it. Notice that the table border provides you the feedback based on your response. Practice on the display below.", (3) "The training session will begin next. After the training, you will be asked to login with your passphrase."

Before the starting training session, a sample display was provided for the participants to practice the task before going through training.

After training, participants were asked to login. For all login sessions in the study, participants needed to recall their passphrase within a maximum five attempts. If they failed remembering after five attempts, their passphrase was shown to them and they were asked to memorize it. For this session, participants in the Control condition wherein no cue was provided for the participants, the following instructions were provided: "You will be provided with 4 text boxes to enter 4 words of your passphrase. You can have up to 5 attempts in order to input your passphrase successfully. Below you can see a sample input page.". The provided instruction

for other conditions was as follows: "You will be provided with 4 displays, each containing 32 words. Your task is to: (1) Find the word of your previously assigned passphrase words. (2) When you find the word, click on it. (3) Once you click on the word, the next display appears. (4) If you don't find the correct words, you are given up to 5 attempts to find correct words. Below you can see a sample display." Participants could take as long as needed to find each target word. After five unsuccessful attempts, participants were shown their passphrase.

We then asked the participants to return after 24-48 hours, and again one week after their first session in order to complete the second and third sessions respectively. They also received an email notification in order to remind them about these follow-up sessions. In the second session, participants were asked to recall their passphrase. The third session was identical to the second with an additional questionnaire which was provided at the end of the login task.

4.4 Statistical Testing

Using a significance level of $\alpha = .05$, for each comparison, we first ran an omnibus test across all conditions. We used Kruskal-Wallis (indicated KW), for omnibus tests on quantitative data (e.g., login times) and χ^2 on categorical (e.g., number of attempts needed for successful login). If the omnibus tests showed significance, we performed selected pairwise tests of interest. We also performed the Holm-Bonferroni correction (indicated HC) for multiple-comparison correction. This correction performs an adjustment to significance levels when several statistical tests are being performed simultaneously on a single data set. It is used to reduce the chances of obtaining false-positive results (Type I errors) when multiple pairwise tests are performed on a single set of data.

5 RESULTS

We start by providing some demographics of the participants. We then provide some information about the participation and drop-out rates across all the experimental conditions. Finally, we describe the data regarding participants in each condition who recorded their passphrases, recalled their passphrase, forgot their passphrase, login times, and exit survey results.

5.1 Participants

893 participants initially signed up for our study, 780 finished the first part. Of the participants who finished the first part, 476 and 430 finished the second and third part of the study. 52% of our participants were male and 48% were female. 5% with high school or equivalent, 72% had some college or university degree; 17% master, and 5% with doctoral degree. 2% of the participants were aged below 20, 18% between 20-25, 41% between 26-35, 31% between 36-50, and 8% above 50. 98% of the participants had English as their first language.

5.2 Study Dropouts

Of 893 participants who started our study, 780 finished the first part; 476 participants returned within 24 to 48 hours of receiving our email invitation and completed the second part of the study, and 430 participants completed the third part of our study. These statistics, broken down by condition, are shown in Table 2.

Condition	Started	S1	S2	S3
CC-SP	155	84%	53%	51%
Control	149	88%	52%	48%
CC	137	93%	58%	50%
SP	139	91%	54%	50%
Repetition	150	88%	54%	42%
Recognition	163	81%	50%	47%

Table 2: The number of participants who signed up for the study in each condition, and the percentage who continued all three sessions (i.e., Session 1, Session 2, and Session 3).

5.3 Storage

Our application captured those participants who either did a copy-paste action (for the control condition) or screenshot while they were performing the task. For those participants who finished all three sessions, through the exit questionnaire we asked them if they have recorded their assigned passphrase. Table 3 indicates the number of participants in each condition who recorded their passphrase. Note that in this table we did not double count those who mentioned in the questionnaire that have stored their passphrase and also our system detected their copy-paste action.

Condition	Copy-Paste	Screenshot	Record	Total %
CC-SP	0	2	9	7%
Control	6	8	34	26%
CC	0	2	12	9%
SP	0	0	12	9%
Repetition	0	1	20	15%
Recognition	0	2	19	14%
Total	6	15	106	

Table 3: The number of participants who recorded their assigned passphrase.

The table indicates the number of the participants who either mentioned in the questionnaire that they recorded their passphrase or the system caught their copy-paste or screenshot actions. The last column of the table indicates the total percentage of the participants in each group who have performed any type of storage. We hypothesized there will be a significant improvement in mean number of users who recorded their passphrase. The null hypothesis that we claim for the purpose of dependency between the condition and storing bahaviour, assumes that there is no association between the condition and storage behaviour. Running χ^2 on the number of users who recorded their passphrase, showed a significant difference in recording behaviour of the users of different conditions

($p < .001$). Running χ^2 showed a significant difference in recording behaviour of the CC-SP condition compared to the Control condition ($\chi^2 = 17.96, p < .001$) which rejects the null hypothesis and confirms the participants in CC-SP needed to record their passphrase less while for the control condition, more participants had recorded their passphrase.

Conditions	*p-value*	HC	Effect Size
CC-SP* and Control	**<.001**	**0.008**	0.25
CC-SP and CC	0.8	0.02	0.04
CC-SP and SP	0.8	0.02	0.04
CC-SP and Repetition	0.05	0.01	0.13
CC and Repetition	0.1	0.012	0.09
SP and Recognition	0.2	0.016	0.08

Table 4: The results of χ^2 for the pairwise comparison for the storage behaviour for the experimental conditions. Holm-Bonferroni Correction (HC) was applied on the set of 6 pairwise tests for the S3 storage behaviour. The HC column shows the updated alpha value for achieving significance. The boldfaced row shows the condition with a statistically significant difference. The asterisk-marked conditions are the ones that performed better than the paired condition.

5.4 Recall Rates

After the training session with an average duration of 64 seconds, we asked our participants to recall their passphrase through three login sessions. Participants who were not able to recall their passphrase after five attempts are considered as having forgotten their passphrase and were shown the passphrase on the screen.

5.4.1 First Session. For an immediate recall test, most participants successfully recalled their passphrase. Table 5 shows the successful recall on both first entry and those who needed more attempts to recall. As per $H_{memorability\_recall}$, we hypothesized there will be significantly greater memorability in implicit learning-based trained passphrases compared to the control condition. The null hypothesis that we claim for the purpose of dependency between the condition and success rate, assumes that there is no association between the condition and login success rates. As shown on the table, the CC-SP condition outperformed the others, having the highest total success rate and lowest average number of attempts in order to successfully login. Running χ^2 on the login success rates, indicated a significant difference across conditions ($\chi^2 = 26.20, p < .001$). This will reject the null hypothesis, indicating an association between the group and login success rates.

5.4.2 Second Session. We sent our participants a notification email 24 hours after the first session and asked them to login to our web interface by recalling their assigned passphrase. The participants who came back within 24-48 hours were able to access our system. Running χ^2 on the login success rates, indicated a significant difference across conditions ($\chi^2 = 14.05, p = .01$). Table 6 indicates the success rates across all the experimental conditions. CC-SP remained the condition with the highest total success rate and lowest login time.

5.4.3 Third Session. The third part of our study was 7-8 days after the first session. The participants who had completed the first two sessions were qualified to perform this task. Running χ^2 on the login success rates, indicated a significant difference across conditions ($\chi^2 = 22.76, p < .001$). CC-SP remained the condition with the highest total success rate and lowest login time.

As shown on Table 7, there were some participants on each condition who they needed to have more attempts in order to successfully recall their passphrases and login. This number ranged between 5 to 9% across all conditions, with CC having the lowest and SP the highest rate.

5.5 Login Time

We hypothesized there will be a significant improvement in time required to login. The null hypothesis assumes the distribution of login time for the conditions are equal. Running the KW test, for the third login session, considering successful logins, there was a significant difference for the login time between all conditions ($p < .001$).

5.6 Pairwise Comparisons

As per our main research questions, we want to know if the combination of two implicit memory techniques could improve memorability of system-assigned passphrases. Our findings confirmed the usability benefits of the proposed approach as there was a significant difference between CC-SP and both Control and Repetition conditions; however, we were also interested in finding out which implicit learning technique is the most effective; that is, CC, SP, or the combination of both. To answer this question, we designed two other experimental conditions, one for CC and one for SP solely. Pairwise comparison of CC-SP and CC conditions total authentication success rate for the third login session did not show significant difference of performance between the two conditions with $\chi^2 = 1.06$, $p = .3$. It is interesting to note, although there was no significant difference for the login success rate, the average login time has a statistically significant difference between CC-SP and CC ($p < .001$). Pairwise comparison of CC-SP and SP conditions authentication success rate for the third login session did not show statistically significant difference $\chi^2 = 4.63$, $p = .03$ ($HC\alpha = .012$).

We also included a Repetition condition to examine if the usability improvements of our approach are due to repetitions in the training phase and recognition in the login phase. Pairwise comparison of CC and Repetition condition's total authentication success rate for the third login session did not show significant differences, $\chi^2 = 5.07$, $p = .03$ ($HC\alpha = .012$). Because the comparison between CC and Repetition was borderline, it might be worth further study. The same analysis was performed to evaluate if the effectiveness of our implicit learning-based approach is due to SP, or the participants just recognize the words without any help from the semantic relation of the words. To evaluate this, we performed a pairwise comparison of the success rates for the third session of the SP and the Recognition conditions, finding that there is no significant difference in performance between the two conditions $\chi^2 = .38$, $p = .5$.

		Success first attempt	More attempts	Avg login	Total success
Conditions	CC-SP	96.92%	0.77%	8.14	97.69%
	Control	78.63%	6.11%	14.33	84.73%
	CC	91.41%	2.34%	9.08	93.75%
	SP	85.83%	3.94%	12.33	89.76%
	Repetition	85.61%	5.3%	19.6	90.91%
	Recognition	74.24%	6.06%	22.09	80.30%

Table 5: First login session total success rate percentages, the percentages of those who needed more attempts to login, and average login duration (in seconds) for each condition.

		Success first attempt	More attempts	Avg login	Total success
Conditions	CC-SP	87.8%	3.66%	10.76	91.46%
	Control	65.38%	8.97%	25.43	74.36%
	CC	81.01%	6.33%	15.32	87.34%
	SP	70.67%	9.33%	21.32	80.00%
	Repetition	72.84%	9.88%	27.07	82.72%
	Recognition	67.90%	4.94%	30.39	72.84%

Table 6: Second login session (one-two days later) total success rate percentages, the percentages of those who needed more attempts to login, and average login duration (in seconds) for each condition.

		Success first attempt	More attempts	Avg login	Total success
Conditions	CC-SP	83.54%	5.06%	13.74	88.61%
	Control	51.39%	5.56%	45.78	56.94%
	CC	76.81%	4.35%	22.08	81.16%
	SP	68.57%	8.57%	25.67	77.14%
	Repetition	57.14%	7.94%	49.89	65.08%
	Recognition	51.95%	7.79%	30.45	59.74%

Table 7: Third login session (one week later) total success rate percentages, the percentages of those who needed more attempts to login, and average login duration (in seconds) for each condition.

We also performed pairwise comparison for the login time to find if there is any statistically significant difference between different pairs. Our analysis confirmed the login time could be affected depending on the training. Running the MWU test, for the third login session, there was a significant difference for the login time of the CC-SP condition compared to the Control condition (p = .008). The comparison between SP and CC-SP also showed a significant difference between the login time for the third session (p = .008). The pairwise comparison of CC-SP and CC also showed a significant difference in the login time (p = .008). We also evaluated if there is any significant difference between CC and Repetition as well as SP and Recognition. Our analysis confirmed, the login time for the third session of CC and Repetition and SP and Recognition also had significant differences for the login time (p = .008).

5.7 Exit Survey

We hypothesized there will be an improvement in users' perceptions for implicit learning-based trained passphrases compared to the control condition. A variety of questionnaires have been used for assessing the perceived usability of interactive systems. To assess subjective reactions that participants in our usability test had to our system, we used SUS (System Usability Scale) [9]. A SUS score above 68 would be considered above average and anything below 68 is below average [34]. As shown on Table 10, the only conditions receiving a score over 68 were CC-SP and CC.

Running the KW test on the SUS scores, there was a significant difference for the scores between all conditions ($p < .001$). Table 11 shows the results of the pairwise comparisons for all the experimental conditions. CC-SP had significantly improved SUS scores over both Control and Repetition conditions, supporting our hypothesis

Conditions	*p-value*	HC	Effect Size
CC-SP* and Control	**<.001**	**0.008**	0.34
CC-SP and CC	0.3	0.016	0.08
CC-SP and SP	0.03	0.012	0.19
CC-SP* and Repetition	**0.004**	**0.01**	0.30
CC and Repetition	0.03	0.012	0.18
SP and Recognition	0.5	0.02	0.18

Table 8: The results of χ^2 for the pairwise comparison for the success rate for the experimental conditions. Holm-Bonferroni Correction (HC) was applied on the set of 6 pairwise tests for the S3 success rates. The HC column shows the updated alpha value for achieving significance. The boldfaced rows show the conditions with a statistically significant difference. The asterisk-marked conditions are the ones that performed better than the paired condition.

Conditions	*p-value*	HC	Effect Size
CC-SP* and Control	**<.001**	**0.008**	0.70
CC-SP* and CC	**<.001**	**0.008**	0.74
CC-SP* and SP	**<.001**	**0.008**	0.66
CC-SP* and Repetition	**<.001**	**0.008**	0.76
CC* and Repetition	**<.001**	**0.008**	0.44
SP* and Recognition	**<.001**	**0.008**	0.24

Table 9: The results of MWU test for the pairwise comparison for the login time for the experimental conditions. Holm-Bonferroni Correction (HC) was applied on the set of 6 pairwise tests for the S3 login times. The HC column shows the updated alpha value for achieving significance. All rows are boldfaced since all of the conditions had a statistically significant difference. The asterisk-marked conditions are the ones that performed better than the paired condition.

that implicit learning-based passphrases will produce an improvement in user perception. We also included some other questions to evaluate users' sentiment about the scheme (see Appendix B). Figure 4 shows an overview of the results. The figure shows the responses of the participants in the CC-SP condition which had the best performance among other conditions. As indicated, most users did not find the system boring, or difficult to use instead of passwords. We also asked them: *"Given that the training session teaches you a system-assigned passphrase, which provides more security, would you use it instead of a regular password?"*, for different types of accounts. The majority of the participants showed interest in using our system for online-banking or email accounts.

Condition	SUS score
CC-SP	75.86
Control	59.62
CC	69.51
SP	67.73
Repetition	64.19
Recognition	67.69

Table 10: The average SUS score for participants of each group.

Conditions	*p-value*	HC	Effect Size
CC-SP* and Control	**<.001**	**0.008**	0.40
CC-SP and CC	0.1	0.01	0.13
CC-SP and SP	0.01	0.01	0.18
CC-SP* and Repetition	**0.002**	**0.01**	0.26
CC and Repetition	0.1	0.01	0.12
SP and Recognition	0.8	0.02	0.02

Table 11: The results of MWU test for the pairwise comparison for the SUS scores for the experimental conditions. Holm-Bonferroni Correction (HC) was applied on the set of 6 pairwise tests. The HC column shows the updated alpha value for achieving significance. The boldfaced row shows the condition with a statistically significant difference. The asterisk-marked conditions are the ones that performed better than the paired condition.

6 DISCUSSION

We proposed an approach enabled by implicit learning in order to facilitate memorability of system-assigned authentication secrets. Through an online study, we evaluated different conditions and found the implicit learning-based training improves memorability of system-assigned passphrases. The results suggest that when the two implicit learning techniques, CC and SP, are combined and used to train users on system-assigned passphrases, we can have the best short-term and long-term memorability. In this section, we discuss the results and summarize the high-level findings.

6.1 Usability Improvements

As per the Usability-Deployability-Security framework [5], there are different measures known to confirm usability of an authentication approach. In terms of these usability measures, the CC-SP system outperforms passwords in some ways. Our design is *Quasi-Physically-Effortless* as users only need to click (or on a touchscreen, touch) on the words after they find the passphrase word. The simplicity of our system makes it *Easy-to-Learn* as per our questionnaire, 96% of the participants did not find a need for learning a lot of things before using the system. The short length of the login phase means that it is *Efficient-to-Use*. Comparing to the required time to input a system assigned 5-character password (mean 27.5s, 2 days

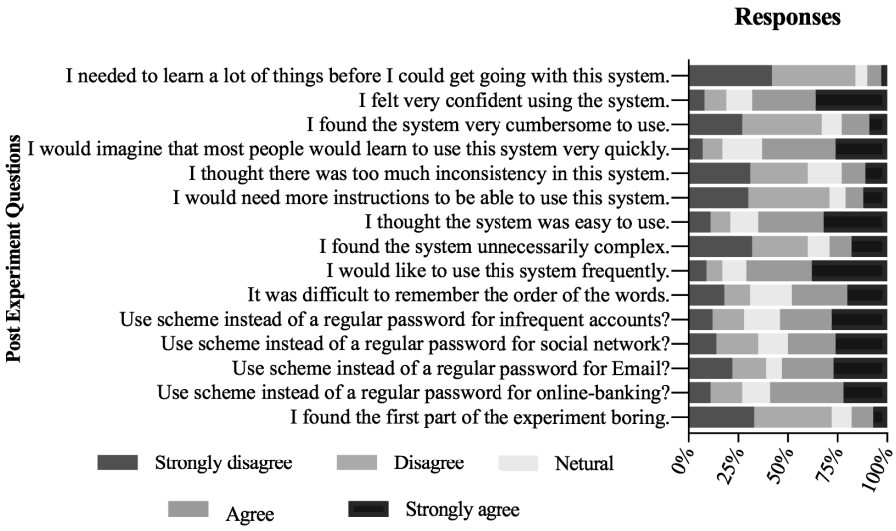


Figure 4: Likert response data on the post experimental questionnaire for the CC-SP condition participants.

later [36]), CC-SP is *Efficient-to-Use* (mean 13s 1 week later). Our authentication success rates were high (88%, one week later), and since the users do not need to type their passphrase, CC-SP has *Infrequent Errors* and performs better than system-assigned passphrases (57%, one week later in our Control group; 44% 2 days later [36]). To offer the *Easy-Recovery-from-Loss* benefit, the approach needs to provide convenience when the credentials are lost or forgotten to regain a new authentication secret. The required training time for assigning a new passphrase is 64 seconds on average, making our approach worse than passwords for offering *Easy-Recovery-from-Loss*. However, since resetting passwords can have consequences in terms of money (e.g., IT helpdesk costs) and time (e.g., in some cases 2 hours to propagate to all systems [20]), this ~1 minute training may be worthwhile given the reduced number of resets. In summary, CC-SP offers an improved balance of usability and security since user-chosen passwords are insecure against online attacks [44], and previous approaches to system-assigned PINs and passwords with comparable security have poor memorability and high login times.

6.2 Deployability

In terms of deployability benefits [5], CC-SP offers the same deployability benefits as system-assigned passphrases, except server-compatibility since it is more than a text string being hashed and stored.

6.3 System Use Case

Random assignment of authentication tokens is good for security as it makes it difficult for attackers to guess. To prevent users from insecure coping behaviours, we hope to offer them an approach to better memorize and recall system assigned passphrases. Our instantiation of "Implicitly Reinforced Passphrases" (i.e., CC-SP) can be offered for such cases that users are looking for guaranteed protection against online attacks for their accounts, but with less memory burden. CC-SP has the potential to also be used to replace a PIN number, and/or be used as a second factor in some environments that require high security. Future work could look at mobile devices (a variant modified for mobile screens). Some future possibilities to increase the keyspace are discussed in Section 8.

6.4 Interference

Previous CC studies [47] have confirmed that CC resists retrospective interference, meaning the CC effect diminishes when the target is re-positioned elsewhere in a display. Thus, it may prevent interference of multiple passphrases with different arrangements. However, testing authentication scalability needs a separate study.

6.5 Determining Important Factors for CC-SP

Including two other conditions in our study; i.e., the CC and SP conditions, allowed us to find if these two methods can provide significant memorability benefits when they are employed solely. Including the Repetition condition allowed us to determine whether any improvements in CC-SP and CC are due to the implicit learning techniques, or repetition and/or recognition. Finally, the Recognition condition allowed us to determine whether any improvements offered by SP are due to recognition or semantic priming.

The pairwise comparison of CC-SP and Repetition confirmed a statistically significant difference between the two conditions, indicating that the improvements were not due to repetition (and thus exposure time) and recognition effects. Login time was also significantly lower for CC-SP than for Recognition ($p < .001$). Interestingly, the tendency for storing passphrases was lowest (7%) for CC-SP. This can also be an indication that in CC-SP, the training is more effective.

The pairwise comparison of CC vs. Repetition did not show statistically significant difference for the authentication success rates with the p-value (0.03) and HC corrected alpha (0.01).

We also evaluated if there is any significant difference for the login time of CC versus Repetition. Our analysis confirms that the

login time for the third session of CC and Repetition had statistically significant difference (p = .008), meaning that CC's improved performance comes from more than just repeated exposures.

Including the Recognition condition in the experiments allowed us to find out if providing SP alone can result in better memorability. Comparison of this condition with SP did not show any significant improvement for memorability; however, it resulted in decreased login time. In summary, our results show that the combination of CC and SP (CC-SP) offers significant memorability and login time improvements, independent of repetition and recognition factors. While CC and SP on their own do not appear to offer significant memorability improvements, they do offer improvements in login time and as such may also be worth further investigation.

7 ECOLOGICAL VALIDITY

There are different factors that may affect the ecological validity of the study. Passphrases are not as well-known as passwords to the users which could affect the way they interact with the system. Another ecological validity issue related to authentication studies is that participants do not put as much effort as for their valued sensitive accounts. This may result in less effort to memorize or recall their assigned passphrase.

The participants performed the study through an online system where they were involved in their usual physical environments, without the intervention of any experimental equipment or person. While this may be better than a lab environment in some ways, using MTurk means that our participants may have been less motivated and/or more rushed than usual. Regardless of any issues associated with the use of MTurk, our comparison to control groups should still provide useful evidence of whether our approach yields an improvement.

8 CONCLUSION AND FUTURE DIRECTIONS

Our results indicate that implicit learning techniques can be used to reinforce memory for system-assigned passphrases. We anticipate that our findings will stimulate further research into other authentication designs that harness implicit learning.

Our proposed approach aims at overcoming an effortful authentication experience for system-assigned secrets. Our CC-SP system offers an improved balance of usability and security compared to system-assigned PINs [19] and passwords [36].

The usability of CC-SP is much better than these systems as it has high memorability (88% success rate one week later) and infrequent login errors. It also has faster login times (mean 13 seconds 1 week later) which is much faster than MTurk studies of system-assigned 6-digit PINs (mean 41.7 seconds [19]) and system-assigned 5-char passwords (mean 27.5 seconds [36]) 2 days later. It is worth noting however that user-chosen PINs or passwords have shorter login times, but these are not comparable in terms of security. CC-SP also involves a relatively short one-time training cost of 64 seconds on average. Since forgetting passwords can have consequences in terms of money (e.g., IT helpdesk costs) and time, and can take up to two hours before it has propagated to all the systems [20, 43], one can view this approx. 1 minute training as worthwhile given the reduced number of forgotten passphrases, which reduces the number of resets. Our CC-SP system also has the added benefit of phishing resistance and according to Thomas et al. [40], phishing and leaks from other verifiers are currently two important threats that lead to credential theft. Leaks from other verifiers are also protected with our approach, as system-assigned secrets are not reused across sites.

In future work it would be interesting to improve resilience to phishing, even against targeted phishing attacks, by using the technique of Cued Click-Points (CCP) [11]. CCP uses one click-point on each image (from a sequence of five images). The next image is displayed based on the location of the previously chosen click-point. CC-SP can be considered in the same way wherein images are word displays and click-points are the word of the passphrase. Depending on what word they click on each display, the next display shown would be different. Choosing the correct word of the passphrase results in the next display being of the next passphrase word, whereas the wrong selection presents a different, unfamiliar display. Such an amendment complicates targeted phishing attacks when an attacker tries to harvest displays as they will most likely wind up with an incorrect sequence of displays to present the user with.

Our goal was to determine whether implicit learning (IL) techniques can improve recall for system-assigned tokens, more than commonly studied aids of repetition and recognition. Our study evaluates CC, SP, or the combination is most effective. Previous work on CC and SP were applied independently (not compared) to detect IL, not its potential to reinforce explicit memory. We included the basic passphrase condition to understand the usability of our system, independent of the type of tokens to be recalled; each condition used the same pool of 10 randomly chosen passphrases.

There are some other conditions that are of interest to study in future work. For instance, as mentioned in 4.2, due to the difference in the keyspace for CC-SP and the Control condition, future study is needed to compare CC-SP performance with a system-assigned passphrase with the same keyspace. Another condition of interest is to have the user go through a different training phase wherein they rehearse the passphrase the same number of times as in CC-SP, rather than leaving it up to them to spend as much time as they want memorizing. In future work, we are also interested to see if users are not shown the word displays during login, whether they could successfully recall their passphrases. If the results turn out to be promising, this would be an indication that the implicitly learnt contexts can be retrieved without the need for cues. Such an amendment could improve the security of the scheme so that it is secure against offline attacks, as it is still system-assigned yet would not expose any cues for the login session.

Another avenue for future work is to improve security against offline attacks, by investigating whether increasing the number of displays and/or distractors on each display (e.g., having 6 displays instead of 4, or 5 displays with twice as many distractors) would result in increased keyspace and maintain the usability improvements we observed in this work.

9 ACKNOWLEDGMENTS

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number 402500-2013 and *RGPIN*-2018-05919.

REFERENCES

[1] Mahdi Nasrullah Al-Ameen, Kanis Fatema, Matthew Wright, and Shannon Scielzo. 2015. The Impact of Cues and User Interaction on the Memorability of System-Assigned Recognition-Based Graphical Passwords. In *Symposium on Usable Privacy and Security*. USENIX Association.

[2] Mahdi Nasrullah Al-Ameen, Matthew Wright, and Shannon Scielzo. 2015. Towards Making Random Passwords Memorable: Leveraging Users' Cognitive Ability Through Multiple Cues. In *ACM Conference on Human Factors in Computing Systems*. 2315–2324.

[3] Hristo Bojinov, Daniel Sanchez, Paul Reber, Dan Boneh, and Patrick Lincoln. 2012. Neuroscience Meets Cryptography: Designing Crypto Primitives Secure Against Rubber Hose Attacks. In *Presented as part of the 21st USENIX Security Symposium (USENIX Security 12)*. USENIX, Bellevue, WA, 129–141.

[4] Joseph Bonneau. 2012. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *IEEE Symposium on Security and Privacy*. 538–552.

[5] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. 2012. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *IEEE Symposium on Security and Privacy*. 553–567.

[6] Joseph Bonneau and Stuart Schechter. 2014. Towards Reliable Storage of 56-bit Secrets in Human Memory. In *USENIX Security Symposium*. 607–623.

[7] Joseph Bonneau and Ekaterina Shutova. 2012. Linguistic Properties of Multi-Word Passphrases. In *International Conference on Financial Cryptography and Data Security*. Springer, 1–12.

[8] James R. Brockmole and John M. Henderson. 2006. Using Real-World Scenes as Contextual Cues for Search. *Visual Cognition* 13, 1 (2006), 99–108.

[9] John Brooke. 1996. SUS-A Quick and Dirty Usability Scale. *Usability Evaluation in Industry* 189, 194 (1996), 4–7.

[10] Claude Castelluccia, Markus Duermuth, Maximilian Golla, and Fatma Deniz. 2017. Towards Implicit Visual Memory-Based Authentication. In *Network and Distributed System Security Symposium*. San Diego, United States.

[11] Sonia Chiasson, Alain Forget, Robert Biddle, and Paul C. van Oorschot. 2008. Influencing Users Towards Better Passwords: Persuasive Cued Click-Points. In *Proceedings of the British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1*. 121–130.

[12] Marvin M Chun and Yuhong Jiang. 1998. Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology* 36, 1 (1998), 28–71.

[13] Marvin M Chun and Yuhong Jiang. 2003. Implicit, Long-Term Spatial Contextual Memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 29, 2 (2003), 224–234.

[14] Tamara Denning, Kevin Bowers, Marten Van Dijk, and Ari Juels. 2011. Exploring Implicit Memory for Painless Password Recovery. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2615–2618.

[15] Dinei Florêncio, Cormac Herley, and Paul C. Van Oorschot. 2014. An Administrator's Guide to Internet Password Research. In *Proceedings of the USENIX Conference on Large Installation System Administration*. 35–52.

[16] Google. 2013. word2vec. Available at: https://code.google.com/archive/p/word2vec/, last accessed November, 2017.

[17] Annabelle Goujon, André Didierjean, and Sarah Poulet. 2014. The Emergence of Explicit Knowledge from Implicit Learning. *Memory Cognition* 42, 2 (2014), 225–236.

[18] SM Taiabul Haque, Mahdi Nasrullah Al-Ameen, S Scielzo, and M Wright. 2017. Learning System-Assigned Passwords (up to 56 bits) in a Single Registration Session with the Methods of Cognitive Psychology. *Proc. USEC. The Internet Society* (2017).

[19] Jun Ho Huh, Hyoungshick Kim, Rakesh B Bobba, Masooda N Bashir, and Konstantin Beznosov. 2015. On the Memorability of System-Generated PINs: Can Chunking Help?. In *Proceedings of the Symposium on Usable Privacy and Security*. 197–209.

[20] Philip G Inglesant and M Angela Sasse. 2010. The True Cost of Unusable Password Policies: Password Use in the Wild. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, 383–392.

[21] Zeinab Joudaki, Julie Thorpe, and Miguel Vargas Martin. 2017. System-Assigned Passwords You Can't Write Down, but Don't Need to. In *Privacy, Security, and Trust*.

[22] Jerrold Katz and Jerry Fodor. 1963. The Structure of a Semantic Theory. *Language* 39 (1963), 170–210.

[23] Mark Keith, Benjamin Shao, and Paul Steinbart. 2009. A Behavioral Analysis of Passphrase Design and Effectiveness. *Journal of the Association for Information Systems* 10, 2 (2009).

[24] Mark Keith, Benjamin Shao, and Paul John Steinbart. 2007. The Usability of Passphrases for Authentication: An Empirical Field Study. *International Journal of Human-Computer Studies* 65, 1 (2007), 17–28.

[25] Alejandro Lleras and Adrian Von Mühlenen. 2004. Spatial Context and Top-Down Strategies in Visual Search. *Spatial Vision* 17, 4-5 (2004), 465–482.

[26] Timothy McNamara. 2005. *Semantic Priming : Perspectives from Memory and Word Recognition*. Psychology Press, New York.

[27] Ken McRae and Stephen Boisvert. 1998. Automatic Semantic Similarity Priming. *Journal of Experimental Psychology* 24, 3 (1998), 558.

[28] David E. Meyer and Roger W. Schvaneveldt. 1971. Facilitation in Recognizing Pairs of Words: Evidence of a Dependence Between Retrieval Operations. *Journal of Experimental Psychology* 90, 2 (1971), 227.

[29] Craig M Mooney. 1957. Age in the Development of Closure Ability in Children. *Canadian Journal of Psychology* 11, 4 (1957), 219.

[30] Randall Munroe. 2012. xkcd: Password Strength. https://www.xkcd.com/936/

[31] James H Neely. 1977. Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention. *Journal of Experimental Psychology* 106, 3 (1977), 226.

[32] Arthur S. Reber. 1989. Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology: General* (1989).

[33] Daniel J Sanchez, Eric W Gobel, and Paul J Reber. 2010. Performing the Unexplainable: Implicit Task Performance Reveals Individually Reliable Sequence Learning Without Explicit Knowledge. *Psychonomic Bulletin & Review* 17, 6 (2010), 790–796.

[34] Jeff Sauro. 2011. Measuring Usability with the System Usability Scale (SUS). https://measuringu.com/sus/, accessed Sep 30, 2017.

[35] Bernhard Schlagbauer, Hermann J Müller, Michael Zehetleitner, and Thomas Geyer. 2012. Awareness in Contextual Cueing of Visual Search as Measured with Concurrent Access-and Phenomenal-Consciousness Tasks. *Journal of Vision* 12, 11 (2012), 25–25.

[36] Richard Shay, Patrick Gage Kelley, Saranga Komanduri, Michelle L. Mazurek, Blase Ur, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2012. Correct Horse Battery Staple: Exploring the Usability of System-Assigned Passphrases. In *Symposium on Usable Privacy and Security*. 7:1–7:20.

[37] Andrea C Smyth and David R Shanks. 2008. Awareness in Contextual Cuing with Extended and Concurrent Explicit Tests. *Memory & Cognition* 36, 2 (2008), 403–415.

[38] Yishay Spector and Jacob Ginzberg. 1994. Pass-Sentence—A New Approach to Computer Code. *Computers & Security* 13, 2 (1994), 145–160.

[39] Michael A. Stadler and Peter A. Frensch. 1998. *Handbook of Implicit Learning*. Thousand Oaks, CA: Sage.

[40] Kurt Thomas, Frank Li, Ali Zand, Jacob Barrett, Juri Ranieri, Luca Invernizzi, Yarik Markov, Oxana Comanescu, Vijay Eranti, Angelika Moscicki, et al. [n. d.]. Data Breaches, Phishing, or Malware?: Understanding the Risks of Stolen Credentials. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*. 1421–1434.

[41] Yuan-Chi Tseng and Alejandro Lleras. 2013. Rewarding Context Accelerates Implicit Guidance in Visual Search. *Attention, Perception, Psychophysics* 75, 2 (2013), 287–298.

[42] Blase Ur, Jonathan Bees, Sean M Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. Do Users' Perceptions of Password Security Match Reality?. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 3748–3760.

[43] Kim-Phuong L Vu, Robert W Proctor, Abhilasha Bhargav-Spantzel, Bik-Lam Belin Tai, Joshua Cook, and E Eugene Schultz. 2007. Improving Password Security and Memorability to Protect Personal and Organizational Information. *International Journal of Human-Computer Studies* 65, 8 (2007), 744–757.

[44] Ding Wang, Zijian Zhang, Ping Wang, Jeff Yan, and Xinyi Huang. 2016. Targeted Online Password Guessing: An Underestimated Threat. In *ACM Conference on Computer and Communications Security*. 1242–1254.

[45] Nicholas Wright, Andrew S. Patrick, and Robert Biddle. 2012. Do You See Your Password?: Applying Recognition to Textual Passwords. In *Proceedings of the Symposium on Usable Privacy and Security*.

[46] Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. 2004. Password Memorability and Security: Empirical Results. *IEEE Security & Privacy* 2, 5 (2004), 25–31.

[47] Martina Zellin, Adrian von Mühlenen, HermannJ. Müller, and Markus Conci. 2014. Long-Term Adaptation to Change in Implicit Contextual Learning. *Psychonomic Bulletin and Review* 21, 4 (2014), 1073–1079.

[48] Eleni Ziori and Zoltán Dienes. 2012. The Time Course of Implicit and Explicit Concept Learning. *Consciousness and Cognition* 21, 1 (2012), 204–216.

[49] Moshe Zviran and William J. Haga. 1993. A Comparison of Password Techniques for Multilevel Authentication Mechanisms. *Comput. J.* 36, 3 (1993), 227–237.

A SCREENSHOTS OF SAMPLE CC-SP DISPLAY

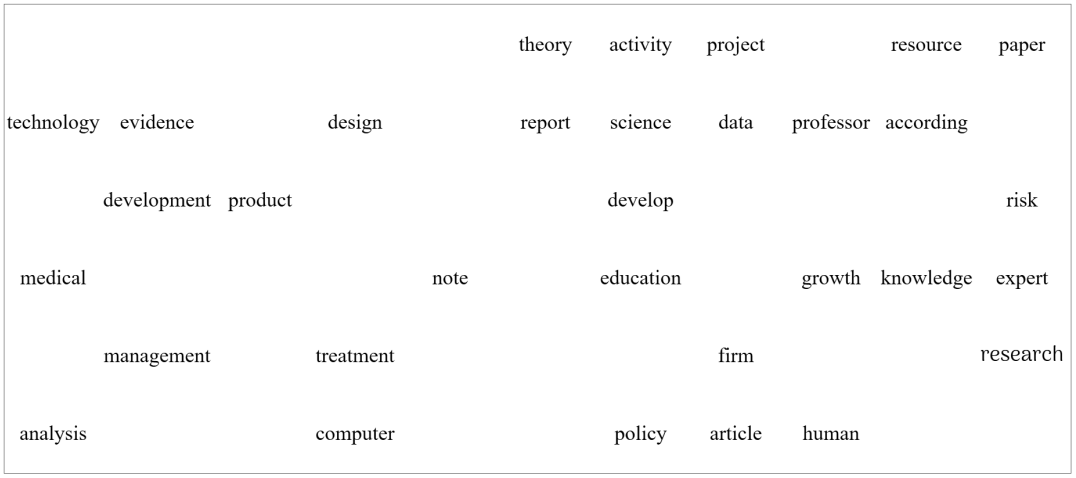


Figure 5: Sample CC-SP display for the Implicitly Reinforced Passphrase Experiment. The target word is 'research'.

B EXIT SURVEY

Please indicate your answer using the following 5-point scale where: (1. = Strongly disagree , 2. = Disagree, 3. = Neutral , 4. = Agree , 5. = Strongly Agree)						
		1	2	3	4	5
1	Did you find the first part of the experiment, i.e., training phase, boring?					
2	Given that such a training session is provided for system-assigned passphrases which provide more security, would you use it instead of a text-based password for?					
3	Online Banking					
4	Email					
5	Social Networks					
6	Accounts you access infrequently (e.g., at most once/week)					
5	I think it was difficult to remember the order of the words in the assigned passphrase.					
6	I think that I would like to use this system frequently in order to have a more secure authentication token*.					
7	I found the system unnecessarily complex*.					
8	I thought the system was easy to use*.					
9	I think that I would need the more instructions to be able to use this system*.					
10	I thought there was too much inconsistency in this system*.					
11	I would imagine that most people would learn to use this system very quickly*.					
12	I found the system very cumbersome to use*.					
13	I felt very confident using the system*.					
14	I needed to learn a lot of things before I could get going with this system*.					
15	Did you, at any time during the study, write down or record your passphrase in any way? Please be honest in your answer, it is OK if you did.	Yes			No	
15.1	If your answer to the above question is, Yes, what did you record?					
17	We are interested in any other comments you might have concerning the entire experiment. Please write in the space below any thoughts you would like to share with us.					

Figure 6: The Implicitly Reinforced Passphrase experiment - Post-experiment questionnaire. Highlighted with *, are the questions from the SUS scale.