# System-Assigned Passwords You Can't Write Down, But Don't Need To

Zeinab Joudaki Julie Thorpe Miguel Vargas Martin
*University of Ontario Institute of Technology*
{Zeinab.Joudaki, Julie.Thorpe, Miguel.VargasMartin}@uoit.ca

***Abstract*—We explore the feasibility of *Tacit Secrets*: system-assigned passwords that you can remember, but cannot write down or otherwise communicate. We design an approach to creating Tacit Secrets based on *Contextual Cueing*, an implicit learning method previously studied in the cognitive psychology literature. Our feasibility study involving 30 participants indicates that our approach has strong security properties: resistance to brute-force attacks, online attacks, phishing attacks, and some coercion attacks. It also offers protection against leaks from other verifiers as the secrets are system-assigned. Our approach also has a high login success rate and low false positive rates. We explore the trade-offs of different configurations of our design and provide insight into valuable directions for future work.**

## I. Introduction

The security of user-chosen passwords has become a serious concern to organizations and individuals alike. Dramatic improvements have been made in offline guessing (or trawling) attacks [1], [2] and targeted attacks that exploit a user's reused passwords [3]. The threat of these attacks is growing with the increasing amount of publicly leaked password data. Perhaps the most damning are attacks that combine leaked password data with personal information—such online targeted password guessing attacks have been shown to guess over 32-73% of passwords within 100 attempts [4].

Password managers offer one solution to these problems by allowing users to generate and securely store random passwords. However, many users distrust them given recent password manager data breaches [5] and software vulnerabilities[6]. Another solution, for a small number of accounts with high security requirements, is to assign users a random, *system-assigned* password; however, these are well-known to have significant problems with memorability [7] and thus users writing them down. Writing down passwords is only secure in some situations, e.g., when they are stored in a physically secured location such as a safe. For example, an organization that uses a password or PIN to access an important safe or server room is unlikely to have secure physical storage nearby. This problem motivates our research into a completely new approach for system-assigned passwords.

In the present work, we investigate the feasibility of random, system-assigned passwords that can be 'remembered' without being written down. We explored literature on implicit learning and identified a promising method called *Contextual Cueing (CC)*. In CC, users are trained to implicitly learn the location of a target item on a display full of many distractors. Each display can be thought of as a character in their password, and the entire password is a set of such displays. Knowledge of the password is demonstrated by a challenge-response system that authenticates based on performance metrics that indicate the password was implicitly learnt.

The result is a method of creating what we call *Tacit Secrets*: system-assigned passwords that can be remembered, and also cannot be written down or explained to others. The use of CC may also have interesting properties for accessibility; for example, it has been found to remain intact in several neurological and mental disorders [8], and to work with subjects having dyslexia [9], [10]. Our feasibility study indicates that our design has high authentication success rates (86-90%, depending on the configuration), and low false positive rates (0.8-9.2%, depending on the configuration). Our security analysis indicates that our approach is resistant to offline guessing attacks, online guessing attacks, phishing attacks, and some types of coercion attacks. It is also resilient to leaks from other verifiers due to the Tacit Secret being system-assigned. Finally, it also provides some resistance to observation attacks, such that a successful attack would require multiple observations.

**Use Case.** Tacit Secrets could be used for any system requiring the strong security guarantees offered by system-assigned passwords. However, our current design has long login times that limit its practicality. We believe the current design we studied would still be useful in some environments with high security requirements, e.g., unlocking a critical system configuration terminal, unlocking a high-security vault or room, unlocking an encrypted file, etc. If future work shows the implicit memory effect lasts for longer time periods, it may also be useful for fallback authentication.

**Contributions.** Our contribution is the design and feasibility study of a method for producing Tacit Secrets, which the user can remember, despite the fact they cannot write them down. This design should be of interest for use in the environments discussed above. Also, our positive feasibility study results demonstrate that implicit learning can be used to produce a user authentication system with high accuracy, and strong security properties, and as such might be employed in future authentication systems research.

The remainder of this paper is organized as follows. Related work is discussed in Section II. Our Tacit Secrets design is presented in Section IV-C. Our feasibility study design is

Natural Sciences and Engineering Research Council of Canada (NSERC).

described in Section IV, the results of which are presented in Section V. Section VI presents our analysis of the performance of different configurations of our design and Section VII analyzes the security of the recommended configurations. Section VIII discusses limitations to consider when interpreting our results. We conclude the paper with a discussion of our results and in Section IX and future work in Section X.

## II. Related Work

We focus on the related work most relevant to our approach to creating Tacit Secrets: system-assigned secrets, authentication systems that employ implicit memory, and authentication systems that have coercion-resistant properties.

### A. System-Assigned Secrets

System-assigned passwords are much stronger than user-chosen passwords, but the practice is well-known to lead to problems such as poor memorability and requiring a written copy for a long period of time [7]. Writing down passwords is insecure unless the written copy is stored in a physically secure location, or using strong encryption on a device. Some attempts have been made to improve the memorability of system-assigned secrets so they may be usable. Schechter et al. [11] examined the impact of a training period of a few weeks that employed spaced repetition. 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. Shay et al. [12] investigated the potential of using random system-assigned words as opposed to randomly assigned characters, and encouraging users to imagine a scene that links them. Their results were unfortunately not very positive; only 51% could recall the passphrases after 2-5 days. Jeyaraman and Topkara [13] proposed random generation of a password and automatically creating a mnemonic phrase to help recall, but its efficacy is unknown. Al-Ameen et al. [14] proposed a series of cues to aid recall of system-assigned passphrases; pilot studies show this method holds promise as all users recalled their phrase after one week. However, the security offered by the system tested is limited as it has only a 28-bit key space, is vulnerable to coercion, phishing, and observation attacks involving a single session. We study Tacit Secrets with the goal of achieving strong security desired from system-assigned secrets, but that can also be recalled.

### B. Authentication Systems That Employ Implicit Memory

Denning et al. [15] 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. Since the scheme involves the conscious learning of the images, it does not provide any resistance to coercion attacks. Furthermore, the requirement to provide a large set of images makes the system less deployable.

The work most similar to our approach is the scheme of Bojinov et al. [16], as it also offers 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. [17]. 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. The authentication process in this scheme is based on the users' performance (the percentage of the correct responses and RT) on the learnt sequence versus random ones. Only 71%, 47%, and 62% of participants could successfully authenticate using this method immediately, 1 week, and 2 weeks later respectively. No further investigations of this system have been performed. Tacit Secrets has substantially better authentication success rates, registration times, and login times. Other relevant security properties, such as false positive rates and resistance to observation attacks were not evaluated for this system.

### C. Coercion Resistant Authentication

Since one of the properties of Tacit Secrets is its resistance to certain types of coercion, we discuss other approaches that provide some degree of coercion attack resistance. Authentication based on physical tokens (i.e., "what you have") can be given to a threatening attacker and is thus highly vulnerable to coercion. Most knowledge-based forms of authentication (i.e., "what you know" ) are explicitly memorized and can be communicated, thus can also be given to an adversary. One way to protect users in such systems is through panic passwords [18], where any user has a regular password and another, panic, password. If input, the panic password communicates a duress situation to the server. While this approach can help, it can lead to more cognitive load for the user to memorize both passwords and the panic password could be forgotten in a stressful situation.

Some static biometrics (i.e., "who you are"), are vulnerable to coercion whereby the attacker makes a copy of the user's biometric data for use later on (e.g., fingerprints, iris, and facial recognition [19]). Some behavioural biometrics can resist some coercion attacks. Babu et al. [20] propose a method that uses users' transaction time behaviours for authentication. De Luca et al. [21] propose an implicit authentication method for touch screen smart phones whereby they authenticate users based on how they interact with the device using a sequence of time series of touch screen data. Gupta et al. use voice [22] and skin conductance [23] measurements to provide a key generation mechanism with reduced accuracy while the user is under duress. They showed these measures can reveal the user's emotional states and recognize if he/she is under the attacker's control; however, for the suggested voice solution, some people may not be able to speak due to injuries or mental deficiencies, and a person can lose his/her voice temporarily due to illness such as cold, cough, etc. Skin is also affected by several external factors such as temperature, illness, etc. Some advantages Tacit Secrets have over these mechanisms

is that they are system-assigned (and thus have configurable security for higher-security environments), are more difficult to observe in a way the user cannot detect (e.g., through social engineering), and can be changed more easily if compromised.

## III. Tacit Secrets Design

Our design of a Tacit Secrets approach uses implicit learning (IL). IL is the acquisition of skills through the repetition of a task; these skills are acquired unconsciously, unintentionally, and without having declarative knowledge about what has been learnt [24], [25], [26]. IL is associated with complex features or probabilistic patterns, whereas explicit learning is most probable when stimuli are salient [27]. Examples of where IL is involved include perceptual-motor skills, language acquisition, social intuition, and detecting a target in a complex scene [25]. We are inspired by a method known to trigger IL for spatial contexts, called Contextual Cueing (CC) [28], as it has been found to be robust over time (lasting for at least six weeks [29]). We provide a description of CC in Section III-A and how our approach makes use of CC in Section III-B.

### A. Contextual Cueing

Contextual Cueing (CC) is a mechanism [30] through which visual attention can be guided by implicitly learnt knowledge [31]. CC was first developed by Chun and Jiang [30] 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. This context tends to be predictable, because one's visual experience is not based on a random sample of objects; it is structured and repetitive. 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 the location of traffic signals are most often to the right of street signs.

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. See Figure 2 for an example of a display used in CC experiments.

In cognitive psychology experiments of CC, subjects are shown a set of displays where some subset are *repeated* (i.e., shown more than once in the session). For each display, the subject is asked to find the target, and given a time limit of 3 seconds. Over time, for repeated displays, subjects' performance in finding the target improves [26], [32], [28]. Chun and Jiang [28] found that the difference of reaction time between previously unseen (*novel displays*) and seen (*repeated displays*) was significantly different. *Reaction time* (RT) refers to the time it takes a participant to find the target; see Figure 3 for this effect on our experiment (as described in Section IV). Chun and Jiang [28] showed that participants were typically unable to explicitly recognize learnt contexts through a post-experimental classification task.

### B. Tacit Secret Design Overview

On a high level, our design trains users to implicitly learn a secret set of displays, which becomes their *Tacit Secret*. We call user $i$'s secret set of displays $K_i$. The training/registration of a Tacit Secret is explained in section III-C. The login/verification process is described in Section III-D.

### C. Training of Tacit Secrets

The goal of training is to ensure the user $i$ implicitly learns a set of displays; this is accomplished during account registration. Let $D$ be the full set of displays that can exist under the system parameters. $K_i$ (user $i$'s secret) is a set of displays, drawn at random from $D$. Note that $|K_i| << |D|$. We use $N_i$ to refer to a sequence of *novel displays* for $i$; i.e., displays that are not in $K_i$; these are drawn at random from the pool of possible novel displays (i.e., $D \setminus K_i$). We also use the notation $R_i$ to refer to a sequence of *repeated displays* shown to $i$, where each display is drawn at random from $K_i$.

In the training session, user $i$ is shown $R_i$ and $N_i$, which are interleaved at random. For each display, $i$ must search for a single rotated 'T' (the target) among many 'L's (the distractors; see Figure 1). Once the target is found, $i$ must report the target orientation as quickly as possible by pressing the corresponding keyboard arrow key. Pressing the incorrect key, or not pressing any key, results in an invalid response for that display. There is a time limit of 3 seconds for each display that if the user does not answer, the display is removed and the new one is shown. At the end of the training session, the user is expected to have implicitly learnt the configuration for the displays in $K_i$, due to repeated exposure to these displays.

### D. Verification of Tacit Secrets

To be authenticated at a later time, a user $i$ is provided with the same task as in training. The sequence of novel displays in $N_i$ are once again drawn randomly from $D \setminus K_i$, so they are unlikely to have been seen before. The sequence of displays in $R_i$ are drawn again randomly from $K_i$. By demonstrating better performance on the displays in $R_i$ over the displays in $N_i$, the user $i$ is demonstrating knowledge of the displays in $R_i$ (and thus $K_i$). For each display, a response is considered incorrect if the target orientation is not correctly input within the time limit (i.e., 3 seconds). We only consider performance data for the responses labelled as correct. Users only have one chance to input a target orientation for each display.

**Performance Data**. For performance data used in making authentication decisions, we use RT. We note that it may be possible in future work to incorporate further metrics, e.g. related to eye tracking, mouse movements, or touch screen behaviours, depending on the environment.

**Verification Method**. We consider login success to occur if the Mann-Whitney (MWU) test is significant with $\alpha = 0.05$. The null hypothesis for the MWU test is that the distribution of the performance metric, RT, for $R_i$ is the same as for $N_i$, against the alternative hypothesis that the distribution of the performance metric for $R_i$ is significantly different than for $N_i$. This test was chosen as it is non-parametric and

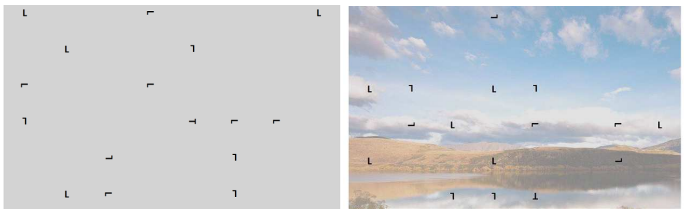

Fig. 1: Illustration of different displays with and without background image.

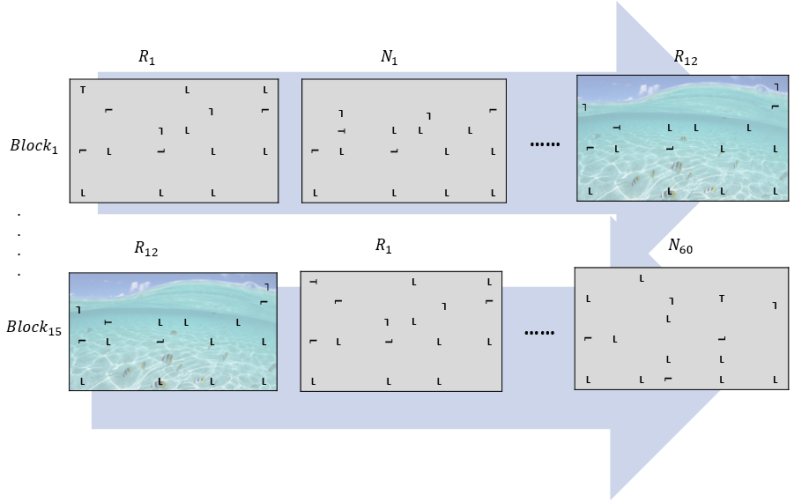


Fig. 2: An example display arrangement during the training session.

the performance data is not normally distributed. Also, the performance data is ordinal.

## IV. Feasibility Study

Here we describe our study to test the feasibility of our approach to Tacit Secrets. The experimental procedure was approved by the Research Ethics Board at our university. The study ran over two weeks in a laboratory environment in order to collect eye-tracking data for future analysis. Participant demographics are described in Section IV-A, study structure in Section IV-B, and design considerations for our implementation and study in Section IV-C.

### A. Participants

Thirty participants (18 males and 12 females, aged between 18 and 25 years) were recruited through email and posters which were distributed across the university campus. These participants were paid $10 each to participate in our lab study and entered into a draw for $50. The inclusion/exclusion criteria consisted of being with normal or corrected-to-normal vision acuity, and not to be registered in any computer security-related program. All of the participants were students, where 67% of the participants had a high school degree (or equivalent) and 33% had a university or college degree. 30% of the participants majored in engineering and applied science, 30% science, 23% business and IT, and the other 17% majored in health and social science. 53% of our participants had normal and 47% had corrected-to-normal vision.

### B. Study Structure

The participants were asked to attend three sessions. The sessions were scheduled according to the participant's convenience, within the following constraints: the second session is two days after the first training session, and the third session happens a week after the second session. The procedures for all three experimental sessions were the same except that the pre-experimental questionnaire is only presented during the first session.

Participants were instructed to sit approx. 60 cm from a 23-inch LCD display monitor with a sample rate of 85 Hz and to press keyboard arrow keys in response to stimuli. In the first session, participants were asked to sign the consent form and then were provided written and oral instructions. They were calibrated with the eye-tracker and started using the application after they agreed to their participation in the experiment. The study purpose (in the consent form and invitation letter) was left intentionally vague, so they were not informed about the exact process of learning that the experiment was testing until after the end of the experiment. The reason for this was that we wished to avoid the possibility of this knowledge affecting their performance and thus the unconscious learning that the experiment aims to test. The experiment's purpose was debriefed at the end of the third session.

During pilot testing, we realized that in addition to a mandatory break that is given between the training and testing phase, the task should allow users to initiate optional rest-breaks when they felt tired. Breaks were initiated by pressing the 'Esc' key and the experiment resumed by pressing 'Enter'.

### C. Study Considerations and Parameters

Here we explain the initial design considerations and parameters of our implementation and feasibility study.

**Number of Displays per Session**. Our primary goal was to ensure we had sufficient data to test feasibility of the approach. Thus, we leaned toward longer training sessions than was likely necessary. To decide on an appropriate number of displays per session, we referred to previous studies on CC [33] and found they suggest that the cueing effect arises after the fourth block of 16 displays, and there are no reliable trends in RT before this block. The decreasing trend for the RT would exist until block 15 and 16 [34], [35]. Thus, our training phase consists of 240 trials (i.e., displays), divided into 15 blocks. Each block contains 16 displays, where 12 are from $R_i$, and 4 are from $N_i$. Figure 2 shows an example of how displays are presented in each block during the training session. In each display, there are 48 (i.e., an invisible matrix of $6 \times 8$) possible target locations. A look at the RT trends in our training data indicate that implicit learning effects are relatively stable after 7 blocks, so it is possible we could reduce the number of displays in the training session accordingly, and thus the training time.

For verification/login, there is a trade-off related to the length of the sequences $R_i$ and $N_i$; for accuracy, there must be enough performance data recorded for each sequence, but longer sequences means a longer login time. We aimed to gather sufficient data to test the feasibility of the approach, and simulate the feasibility of shorter sequences using the data collected. We present the simulation results in section VI to evaluate how optimized these sequence lengths can be in future implementations. To ensure sufficient data, we tested verification/login sessions containing 100 trials where for each user $i$, $R_i$ contains 50 displays drawn randomly from $K_i$, and $N_i$ contains 50 random displays drawn from $D \setminus K_i$. On the day of training, we also performed a short session where $R_i$ contains 20 displays drawn randomly from $K_i$, and $N_i$ contains 20 random displays drawn from $D \setminus K_i$.

**Display Variations**. The training and login tasks contain two variations of displays of size $1440 \times 900$ pixels: array-based (standard CC; see left side of Figure 1) and scene-based which contained a background image (see right side of Figure 1). Scene-based displays elicit scene-based cueing, which is related to a background scene and array-based cueing occurs based only on the position of distractors in the context. Brooks et al. [36] suggest that when a particular repeated array had been consistently associated with a particular scene background, it produces more robust contextual cueing. They found that training with scene-array displays led to joint learning of the two cues, such that cueing was disrupted when either the scene or the array is changed. In our experiment, we used natural scenes as backgrounds for half of the repeated displays. These images were randomly chosen from our database. Participants searched for a target that was predicted by both the background scene and the locations of distractor items. We also adjusted the luminance of the target and distractors across displays in order to increase search items' contrast against the background scene. In all displays, the target appears equally likely in each of four quadrants of the screen to eliminate learning of location frequencies for the repeated stimuli.

**Search Strategy**. To facilitate access to implicit knowledge, thereby allowing a consistent Contextual Cueing Effect to develop, we asked our subjects to use a passive strategy while searching for the target. We notified them that the best strategy for this task is to be as receptive as possible and asked them to "let the unique item pop into your mind as you look at the screen". Lleras et al. [26] hypothesized that using different search strategies: active (an active effort to find the target) vs. passive (intuitive search, wherein they need to be as receptive as possible, let the unique item 'pop' into their mind while looking at the screen, let the display and intuition determine the response, and tune into 'gut feeling'), can have different results while performing the CC task. They experimentally showed that those subjects who used a passive strategy for the search task had more substantial CC effects. We do not know what strategy users really used; however, providing a set of precise and consistent instructions helps us guide users from arbitrarily choosing a search strategy.

**Positive/Negative Feedback**. To indicate that a user's response has been recorded by the system, after pressing a key, a border appears around the display which is either green (when the correct arrow is pressed) or red (when an incorrect arrow is pressed). This decision follows Lleras et al. [37], who investigate how contextual learning is considerably sensitive to external rewards associated with the search interactions.

## V. Results

Here we report the results of our feasibility study. An overview of performance data trends for the whole participant sample is reported in Section V-A. Our results for authentication success rate, false positives, and speed are presented in Section V-B. We report the results of our simulations to determine optimal configurations separately in Section VI.

### A. High-Level Overview of Performance Data Trends

To confirm the CC effect, we first analyze the search RT for the entire sample of participants. Figure 3 indicates the overall RT performance for the repeated displays compared to novel ones for all our participants.

### B. Using RT Performance Data

*1) Authentication Success Rate:* Running the MWU test (with $\alpha = 0.05$) on the recorded RT data for all participants for sessions 1, 2, and 3 revealed that 100%, 88%, and 86% of users had a significant difference between the RT for the displays in $R_i$ versus $N_i$ (recall the verification method described in Section III-D).

*2) False Positives:* Here we evaluate the false acceptance rate (FAR), i.e., the proportion of attempts that would be wrongly classified as legitimate. To evaluate this threat, we used each user's display sequence labels (i.e., 'novel' and 'repeated') to re-label each other user's sequence and see if the

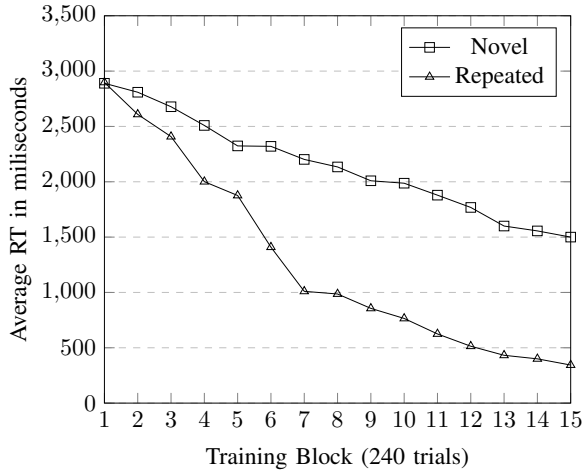


Fig. 3: RTs for novel and repeated displays, for progressive blocks in the training session. The CC effect is evident after 4 blocks and stabilizes after 7 blocks.

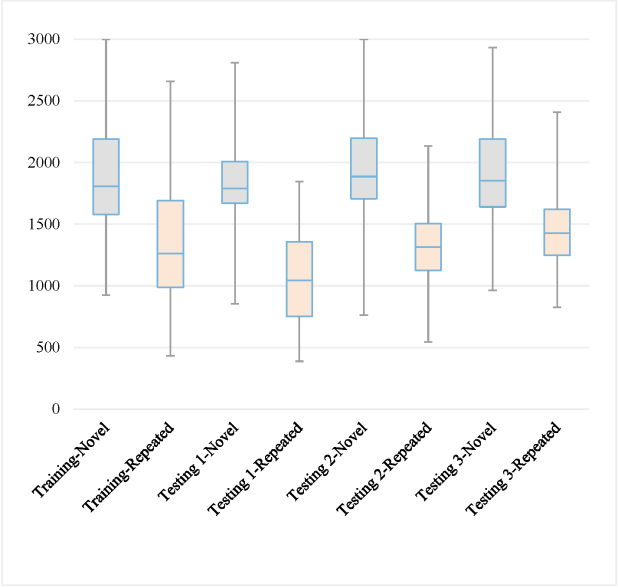


Fig. 4: RT in milliseconds, for each session, for the set of all 30 participants.

newly-re-labelled sequences passed or failed authentication. In our implementation, different users have different sequences, containing a different order of display types. In this scenario, we assume attackers try to use their own performance data to login to another user's account. Our three authentication sessions had a different number of displays: 40, 100, and 100 for Session 1, 2, and 3 respectively. Thus, we performed the analysis through labelling each user's display sequence for Session 1 with the Session 1 display sequence of all other users, and the display sequence of Session 2 and 3 of each user with the display sequence of Session 2 and 3 of all other users. As shown in Table I, through the first run of the test, we considered all types of displays, including array-based and scene-based. Then, we excluded scene-based displays to see if the results changed. The exclusion was due to the possible complexity that displays with background images might have impacted performance of the users. As the results show, there is a negligible improvement of 0.2% in the FAR when we excluded background displays. We further improve the FAR in Section VI-A.

| Display Types | $S_1$-$S_1$ | $S_2$-$S_2$,$S_3$ | $S_3$-$S_2$,$S_3$ | Total | Passed |
|---|---|---|---|---|---|
| All Types | 70/870 | 137/1404 | 147/1566 | 3840 | 9.21% |
| Exclude BG | 60/870 | 127/1404 | 160/1566 | 3840 | 9.03% |

TABLE I: The number of cases the MWU-test passed -False Positives ($S_1$: Session 1, $S_2$: Session 2, $S_3$: Session 3).

*3) Speed:* The mean training time was 14.5 minutes. As noted in Section IV-C, our data indicates that this could likely be reduced by half or more. The login times are explained in Table II. We show how different configurations can substantially improve the login time in sections VI-A.

| | Training | Session 1 | Session 2 | Session 3 |
|---|---|---|---|---|
| Mean | 14:49 | 01:08 | 04:46 | 05:53 |
| Median | 14:18 | 02:09 | 04:14 | 05:40 |
| Std. Dev. | 02:46 | 00:29 | 01:36 | 00:24 |

TABLE II: Completion times for each session.

## VI. Simulating Different Configurations

Our feasibility study verification/login sessions used long sequences of novel and repeated displays, for the purpose of ensuring we had sufficient data to analyze. However, it may be possible that shorter display sequences are required for an effective Tacit Secret design. To determine whether a more optimal configuration of display sequences might exist, we simulate different configurations of our system design (i.e., using different numbers of displays)

For the purpose of these simulations, we sampled data randomly from each user's session 2 and 3 datasets. We are not sampling from session 1 as it is the session with the best results and we wish to avoid biasing our results. For sessions 2 and 3, while the measurements may be influenced by higher learning effects from more repetitions, they may also be influenced by stronger fading effects due to time delays. We note that there is not a remarkable performance improvement of the users from session 2 to 3. Since we are sampling at random from the data collected in sessions 2 and 3, we note that the sequences of novel and repeated displays for every user differs in these simulations, than from the ones actually provided during the testing sessions. In these simulations, we also considered using fewer repeated displays, which would reduce the risk of observation attacks.

### A. Results for Different Configurations

We show our results in a Receiving Operating Characteristics (ROC) graph, to demonstrate the trade-off between

True Positive Rate (TRP or Sensitivity) and False Positive Rate (FPR or 1-Specificity). For Figure 5, we display the configurations that had an authentication success rate over 70%. The closer the points are to the northwest of the graph, the better performance the configuration has. The graph shows the configuration with 25 repeated and 25 novel displays outperforms the other configurations with a TPR of 0.897 TPR and 0.008 FPR. Given this 25-25 configuration, we could keep strong accuracy and have a shorter session duration. The average login time for the 25-25 configuration would be 2.5 minutes which is comparably shorter than the 50-50 configuration average login time (5 minutes).

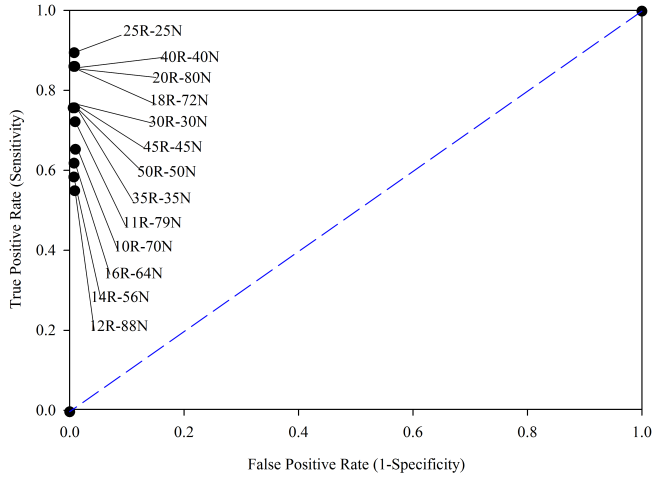


Fig. 5: ROC graph showing performance given different login configurations. Configurations are described by the number of novel (N) and repeated (R) displays they contain.

## VII. Security Analysis

In this section, we first provide our threat model in Section VII-A and then analyze how our approach to Tacit Secrets would fare against five different attack scenarios. These attacks include: (1) offline brute-force in Section VII-B, (2) online guessing using population statistics in Section VII-C, (3) coercion attacks in Section VII-D, (4) observation (shoulder-surfing) attacks in Section VII-E, and (5) phishing attacks in Section VII-F. Our security analyses are performed for different configurations of our approach to Tacit Secrets.

### A. Threat Model

Our threat model is based on the assumption that an adversary wishes to obtain the user's Tacit Secret in order to either decrypt previously collected data and/or gain access to a high security system, room, or administration task. We consider online, offline, coercion, observation, and phishing attacks. Here we list the assumptions our analysis builds upon: (1) The attacker has software that is capable of (*i*) detecting background scene change, (*ii*) detecting display/context objects' orientations, and (*iii*) responding with a chosen true delay, (2) the attacker is able to collect data from the population on the task in general (i.e., for both novel and repeated displays) to obtain response time distributions, and (3) the attacker does not know what the display types are (novel/repeated) for the target user.

### B. Offline Brute-Force Attack

To determine the efficacy of an offline brute-force attack, we must enumerate the size of the key space for our approach. We can consider a random, system-assigned Tacit Secret as a set of size 12 (i.e., $|K_i| = 12$). Each element in $K_i$ could be any display in $D$, with equal probability as it is system-assigned. To enumerate the key space, we must first determine $|D|$. Since each display is a $6 \times 8$ matrix, there are 48 possible positions on each display where objects (distractors or targets) can be placed. Each display contains 16 objects; 15 distractors ('L') and 1 target ('T'). First, the position of the target is chosen: ${}_{48}C_1$. Then the position of each of the 15 distractors is chosen: ${}_{47}C_{15}$. Thus, $|D| = 48 \times {}_{47}C_{15} = 2^{45}$. Given that there are 12 displays to be chosen from $D$, the total number of possible keys is: ${}_{2^{45}}C_{12} \approx 2^{510}$. Thus, a brute-force offline attack is expected to succeed only after approximately $2^{509}$ guesses.

### C. Online Attack Using Population Statistics

For an online attack to succeed, the attacker must correctly guess the type of all displays presented in a login session (i.e., if they are novel or repeated). If, as assumed in Section VII-A, the attacker knows the time distribution of novel/repeated displays, he/she can submit a legitimate guess for each display, and the attack success is determined by correctly guessing the type of each display. To calculate the probability of correctly guessing all the display types in a session for user $i$, consider that there are $|R_i|$ positions from the sequence of $|R_i| + |N_i|$ displays that could contain the repeated displays. Then there are ${}_{(|R_i|+|N_i|)}C_{|R_i|}$ possible positions for the repeated displays. If the attacker has one attempt at guessing this particular sequence, since it changes on each login attempt, the probability of a successful guess of the entire display sequence is $1/({}_{(|R_i|+|N_i|)}C_{|R_i|})$.

**(50-50 Configuration)**. Here $|R_i| = 50$, $|N_i| = 50$, and $|R_i|+|N_i| = 100$. Thus, the probability of a successful online guess is $2^{-96}$.

**(25-25 Configuration)**. We evaluate this configuration as we found it to outperform other configurations (recall Section VI-A). Here $|R_i| = 25$, $|N_i| = 25$, and $|R_i| + |N_i| = 50$. Thus, the probability of a successful online guess is $2^{-47}$. While this indicates this configuration is not as resistant to attacks as the 50-50 configuration, it is still sufficient to be considered resistant to online attacks.

### D. Coercion Attack

Imagine a scenario whereby a motivated attacker threatens a legitimate user with a weapon or using blackmail. The attacker can ask the victim to hand over his/her key, or tailgate the user, e.g., through a physical access control point or forcing the user to login while he/she is present in order to take over

the account after authentication is complete. Below we further explain these attack scenarios.

*1) Communicating the Secret:* This describes when a victim is forced to hand over his/her secret key so the attacker can masquerade as the user at a later time. Since our approach to Tacit Secrets is based on implicit knowledge, even if the trainee is coerced and willing to reveal the key, she/he is not able to do so as she does not have explicit and conscious knowledge of the key. The implicit nature of the acquired knowledge allows protection against such coercion attacks.

*2) Tailgating:* This describes when an attacker tailgates the user to the authentication station, coerces the user to login to the system, and then follows them past the authentication point. In this scenario, we have no evidence that our approach will protect the user's account, as the user may have no choice but to login out of fear for their life. To protect against such an attack scenario, we suggest using a type of panic password [18]. E.g., this could be a simple recognition test from a set of items, whereby the user is trained on a decoy (e.g., an image) to select from the set presented, in the event of this type of coercion. If no coercion is taking place, another item can be selected instead. If the user selects the decoy, the system can detect suspicious activity by a masquerader. It raises an alert to the system of a potential attack in which case the system will not expose the user's real key. For such a system to work, it is important that the user understands there is no way for the attacker to determine the decoy was selected; thus, it is important for the system to behave as though the user had logged in normally (yet with limited access to sensitive data and functionality). We note that it is also conceivable that our approach might provide some protection against coercion even in this scenario. Although it is not yet tested, it is possible that a user might fail to do the task properly as their subconscious system might be affected under duress (e.g., being stressed) [38]. Gauging the stress level of users and how duress influences the measurements of our approach is out of the scope of our feasibility study and is left as future work.

### *E. Observation Attack*

Another type of attack can occur through an attacker's observations. Assuming that the training session is performed in a secure location, the attacker attempts to pass the login test using obtained knowledge through observations of single or multiple testing sessions. Given that he does not have any prior knowledge, he tries to recover the user's dataset through observation. So to have a probabilistic view of this threat, the following scenario should be considered. Each user has a learnt dataset $K_i$ containing 12 displays. Through an authentication session, a sequence of repeated displays $R_i$ will be randomly drawn from the $K_i$ set. If a display is shown at least twice, an observer can understand that it is a part of the user's learnt dataset. To find how successful an attacker might be, we need to know how many sessions are required for the attacker to acquire the knowledge of all the learnt displays $K_i$ for a user $i$. To calculate this number we refer to the "double dixie cup problem" [39], which is a well-known type of the "coupon collector's problem". Given that there are $n$ different types of coupons, the coupon collector problem finds the waiting time for a coupon collector to collect all $n$ coupons. Each coupon is equally likely and would be randomly selected at each trial. The double dixie cup problem is an extension to the coupon collector problem and it determines the expected number of dixie cups which must be purchased in order to complete *m* sets of *n* existing different dixie cups in time *t*. Using the following formula we can calculate this number:

$$E_m(n) = n \cdot \int_0^\infty \left[1 - (1 - e^{-t} \sum_{k=0}^{m-1} \frac{t^k}{k!})^n\right] dt.$$

Given $n = 12$ and $m = 2$, the expected number of displays required to be exposed in order to show the entire set of user's learnt displays would be 58.04. We discuss the implications for each configuration below.

**(50-50 Configuration)**. With a testing session containing 50 repeated displays in $R_i$ (the length of $R_i$ is 50), the attacker is expected to need to observe 2 login sessions in order to see all learnt displays at least twice.

**(25-25 Configuration)**. Given that each login session contains 25 repeated displays in $R_i$, the attacker is expected to need to observe 3 login sessions in order to be able to acquire the knowledge of the user's key.

**Discussion.** There are different amendments to the experiment configuration we can apply in order to decrease the chances of success of the observation attack while keeping the same accuracy. By exposing fewer repeated displays $R_i$, we increase the number of sessions the attacker needs to observe (e.g., for 25-25 configuration, it is 3 login sessions). We can also increase the length of each user's learnt key. By increasing this number, we have more displays to select from and thus the attacker needs to learn more displays in order to know the user's whole set. This would result in the user needing to learn more displays; however, since the CC effect can be observed after the fourth block in training, we may be able to decrease the number of repetitions during the training session from 15. Another possibility is to provide the user some new displays to be learnt during each login session. Once user $i$ learns the display configuration through a few testing sessions, the display can be added to $K_i$ (i.e., their key). Such a mechanism may allow us to update the user's key and prevent attackers from acquiring sufficient knowledge during a series of observations. Finally, we note that if our approach to Tacit Secrets is infrequently used (e.g., for password resets), then it may take a very long time for an attacker to observe the required number of sessions.

### *F. Phishing Attack*

For the purpose of this discussion, we assume that a future version of our Tacit Secrets approach may be useful for a web environment (e.g., for fallback authentication). For an attacker to launch a phishing attack, he/she must create a phishing site that mimics the Tacit Secret login process. In order to gain information about whether a given challenge display $d$ is in user $i$'s $K_i$, the phishing site would need to provide $d$

as a challenge to user $i$, record $i$'s performance data for $d$, and compare it to $i$'s performance data for other displays to determine whether it has better performance. If $d$ has better performance than the majority of displays in the session, the attacker can assume $d \in K_i$. Since there are $2^{45}$ possible displays to challenge the user with, and each login session should only contain 50 displays, we expect it would take over 10 billion phishing attempts on the same target user $i$ to successfully recover $i$'s Tacit Secret.

### G. Security Discussion

Overall, the results of our security analysis suggest that our approach to Tacit Secrets has strong potential to offer security from various attacks: offline and online guessing, certain types of coercion, observation attacks given a small number of observations, and phishing. Our results also highlight how varying the system configuration can result in even stronger security from observation attacks, e.g., by increasing the number of learnt displays and also decreasing how often they are revealed. Of most interest is our 25-25 configuration, which offers the best accuracy, is expected to require 3 observations for an advanced observation attack to succeed, has sufficient security to protect against online guessing attacks, and provides excellent security against offline and phishing attacks.

Due to the inherent features of our approach, the Tacit Secret is protected against coercion attacks involving communication of the secret. However, if the attacker forces the user to login to the system while she is present, the user may feel they need to login to the system as usual. For such systems where this threat is of concern, we suggest employing panic passwords as discussed in Section VII-D2. Finally, since our approach is based on fine grained performance metrics, it has the potential to naturally deteriorate under user duress conditions; however, evaluating whether this is the case is left to future work.

## VIII. Limitations

As our feasibility study was performed in a laboratory environment, our results may not describe how well the approach would work in a non-laboratory setting, especially since the training task requires participant focus. Also, our study participants are university students who may have improved focus than the broader population. Further research is needed to study this approach to Tacit Secrets in other populations and in other settings.

## IX. Discussion

**Feasibility of Tacit Secrets**. We found our approach has much better performance than a previously proposed scheme for Tacit Secrets, SISL. The authentication process in SISL is based on the users' performance (the percentage of the correct responses and response time) on a learnt sequence versus random ones. This data can be used to prevent the same coercion attacks as our approach; however, only 71%, 47%, and 62% of participants could successfully authenticate using this method immediately, 1 week, and 2 weeks later respectively. SISL's first experiment aimed to confirm the existence of implicit learning through an authentication session immediately after training; Their second experiment had two groups of participants: the first group did the SISL task one week after training. The second group did the SISL task two weeks after training, where the length of the login session was doubled (from 5-6 minutes to 10-12 minutes) to see if this change could affect their performance. For this second group of participants, 62% exhibited better performance on the trained sequences. The improvement in the authentication success rate (from 47% to 62%) was due to doubled length of the testing session for the 2-week delay group (from approx. 5-6 minutes to approx. 10-12 minutes).

Our results showed that our approach offers substantial improvements, increasing success rates from 71% to 100% and 47% to 90%, immediately and one week later respectively, reducing training times from 30-45 to 14.5 minutes (and could be further reduced, according to Figure 3), and reducing the login times from 6-12 minutes to approximately 2.5 minutes.

**Interference Between Multiple Systems**. There is a very low probability for interference between the Tacit Secrets assigned and novel displays of different systems using the same approach. This scenario would occur when the novel displays (randomly generated by system A), happen to be part of the user's key for another system (e.g., system B). Consider that the possible number of displays in our configurations is $2^{45}$, and each system has 12 learnt displays (randomly assigned). For the purpose of our discussion, we assume a user has Tacit Secrets for 100 systems. Then the probability that in a login session for user i on system A, a given novel display belongs to $K_i$ on any of the other 99 systems, is $p = \frac{12*(99)}{2^{45}}$. To compute the probability of interference in any display in a given login session, we first compute the probability of no interference using the binomial distribution's probability mass function with number of successes $k = |N_i|$, number of trials $n = |N_i|$, and the probability of success (i.e., of no interference in a given trial) being $1 - p$. Then the probability of interference is one minus the probability of no interference. Then the probability of interference under these assumptions is as follows for the configurations we consider: 25R-25N $= 8.44 \times 10^{-10}$, and 50R-50N $= 1.69 \times 10^{-09}$.

## X. Future Work

We conclude the paper by discussing directions for future research. The design of our feasibility study is based upon related work on the contextual cueing paradigm. To be consistent with those studies, for the training session, we considered 15 blocks containing 16 trials. However, we found the learning effect is detectable after the fourth block, and appears to stabilize after the seventh block. This implies that we may be able to substantially reduce the training session time. In future work, it would be worth investigating how much the training session can be shortened while the learning is still effective and durable.

Since the knowledge acquired through CC has been found to last for delays of at least six weeks [40], [29], it would be

interesting to determine whether it exists for longer duration to evaluate its suitability for fallback authentication.

We are also interested in exploring other enhancements that may improve accuracy and login times, such as using eye tracking performance metrics and eye movements as a behavioral biometric.

Our work suggests that directly using implicitly learnt secrets in authentication may be a viable approach for some contexts. Future work also includes exploring whether implicitly learnt information could be used indirectly to facilitate memorization of traditional authentication secrets.

## XI. Acknowledgments

We thank the participants of our feasibility study. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

## References

[1] R. Veras, C. Collins, and J. Thorpe, "On semantic patterns of passwords and their security impact," in *NDSS*, 2014.

[2] W. Melicher, B. Ur, S. M. Segreti, S. Komanduri, L. Bauer, N. Christin, and L. F. Cranor, "Fast, lean, and accurate: Modeling password guessability using neural networks," in *25th USENIX Security Symposium*, 2016, pp. 175–191.

[3] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang, "The tangled web of password reuse." in *NDSS*, vol. 14, 2014, pp. 23–26.

[4] D. Wang, Z. Zhang, P. Wang, J. Yan, and X. Huang, "Targeted online password guessing: An underestimated threat," in *ACM CCS*, 2016, pp. 1242–1254.

[5] A. Greenberg. (2015) Hack brief: Password manager lastpass got breached hard. https://www.wired.com/2015/06/hack-brief-password-manager-lastpass-got-breached-hard/, accessed May 30, 2017.

[6] J. Siegrist. (2017) Security update for the lastpass extension. https://blog.lastpass.com/2017/03/security-update-for-the-lastpass-extension.html/, accessed May 30, 2017.

[7] J. Yan, A. Blackwell, R. Anderson, and A. Grant, "Password Memorability and Security: Empirical Results," *IEEE Security and Privacy Magazine*, vol. 2, no. 5, pp. 25–31, 2004.

[8] E. C. Merrill, F. A. Conners, Y. Yang, and D. Weathington, "The acquisition of contextual cueing effects by persons with and without intellectual disability," *Research in Developmental Disabilities*, vol. 35, no. 10, pp. 2341 – 2351, 2014.

[9] J. H. H. Jr., D. V. Howard, K. C. Japikse, and G. F. Eden, "Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning," *Neuropsychologia*, vol. 44, no. 7, pp. 1131 – 1144, 2006.

[10] G. Jimnez-Fernndez, J. Vaquero, L. Jimnez, and S. Defior, "Dyslexic children show deficits in implicit sequence learning, but not in explicit sequence learning or contextual cueing," *Annals of Dyslexia*, vol. 61, no. 1, pp. 85–110, 2011.

[11] J. Bonneau and S. Schechter, "Towards reliable storage of 56-bit secrets in human memory," in *USENIX Security Symposium*, 2014, pp. 607–623.

[12] R. Shay, P. G. Kelley, S. Komanduri, M. L. Mazurek, B. Ur, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor, "Correct horse battery staple: Exploring the usability of system-assigned passphrases," in *Symposium on Usable Privacy and Security (SOUPS)*, 2012, pp. 7:1–7:20.

[13] S. Jeyaraman and U. Topkara, "Have the cake and eat it too - Infusing usability into text-password based authentication systems," *Annual Computer Security Applications Conference (ACSAC)*, pp. 473–482, 2005.

[14] M. N. Al-Ameen, M. Wright, and S. Scielzo, "Towards making random passwords memorable: Leveraging users' cognitive ability through multiple cues," in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2015, pp. 2315–2324.

[15] T. Denning, K. Bowers, M. van Dijk, and A. Juels, "Exploring implicit memory for painless password recovery," in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2011, pp. 2615–2618.

[16] H. Bojinov, D. Sanchez, P. Reber, D. Boneh, and P. Lincoln, "Neuroscience meets cryptography: Designing crypto primitives secure against rubber hose attacks," in *21st USENIX Security Symposium*, Bellevue, WA, 2012, pp. 129–141.

[17] D. J. Sanchez, E. W. Gobel, and P. J. Reber, "Performing the unexplainable: implicit task performance reveals individually reliable sequence learning without explicit knowledge." *Psychonomic bulletin and review*, vol. 17, no. 6, pp. 790–796, 2010.

[18] J. Clark and U. Hengartner, "Panic passwords: Authenticating under duress," in *Hot Topics in Security (HOTSEC)*, 2008, pp. 8:1–8:6.

[19] Y. Xu, T. Price, J.-M. Frahm, and F. Monrose, "Virtual u: Defeating face liveness detection by building virtual models from your public photos," in *USENIX Security Symposium*, 2016, pp. 497–512.

[20] B. Babu and P. Venkataram, "Transaction based authentication scheme for mobile communication: A cognitive agent based approach," in *Parallel and Distributed Processing Symposium*, 2007, pp. 1–8.

[21] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, "Touch me once and i know it's you!: Implicit authentication based on touch screen patterns," in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2012, pp. 987–996.

[22] P. Gupta, X. Ding, and D. Gao, "Coercion resistance in authentication responsibility shifting," in *ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, 2012, pp. 97–98.

[23] P. Gupta and D. Gao, "Fighting coercion attacks in key generation using skin conductance," in *USENIX Security Symposium*, 2010, pp. 469–484.

[24] A. Reber and B. Winter, "Implicit learning and tacit knowledge," *Journal of Experimental Psychology: General*, vol. 118, pp. 219–235, 1989.

[25] M. A. Stadler and P. A. Frensch, *Handbook of implicit learning*. CA: Sage: Thousand Oaks, 1998.

[26] A. Lleras and A. von Mühlenen, "Spatial context and top-down strategies in visual search," *Spatial vision*, vol. 17, no. 4-5, pp. 465–482, 2004.

[27] E. Ziori and Z. Dienes, "The time course of implicit and explicit concept learning," *Consciousness and Cognition*, vol. 21, no. 1, pp. 204 – 216, 2012.

[28] M. M. Chun and Y. Jiang, "Implicit, long-term spatial contextual memory," *Journal of experimental psychology. Learning, memory, and cognition*, vol. 29, no. 2, pp. 224–234, 2003.

[29] A. Goujon and J. Fagot, "Learning of spatial statistics in nonhuman primates: Contextual cueing in baboons (papio)," *Behavioural Brain Research*, vol. 247, pp. 101 – 109, 2013.

[30] M. M. Chun and Y. Jiang, "Contextual cueing: Implicit learning and memory of visual context guides spatial attention," *Cognitive Psychology*, 1998.

[31] A. C. Smyth and D. R. Shanks, "Awareness in contextual cuing with extended and concurrent explicit tests." *Memory & Cognition*, vol. 36, no. 2, pp. 403–415, 2008.

[32] C. J. Vaidya, M. Huger, D. V. Howard, and J. H. Howard, "Developmental differences in implicit learning of spatial context." *Neuropsychology*, vol. 21, no. 4, pp. 497–506, 2007.

[33] T. Geyer, M. Zehetleitner, and H. J. Müller, "Contextual cueing of pop-out visual search: When context guides the deployment of attention." *Journal of vision*, vol. 10, p. 20, 2010.

[34] J. R. Brockmole and J. M. Henderson, "Using real-world scenes as contextual cues for search," *Visual Cognition*, vol. 13, no. 1, pp. 99–108, 2006.

[35] A. Goujon, A. Didierjean, and S. Poulet, "The emergence of explicit knowledge from implicit learning," *Memory Cognition*, vol. 42, no. 2, pp. 225–236, 2014.

[36] D. I. Brooks, I. P. Rasmussen, and A. Hollingworth, "The nesting of search contexts within natural scenes: evidence from contextual cuing." *Journal of experimental psychology. Human perception and performance*, vol. 36, no. 6, pp. 1406–18, Dec. 2010.

[37] Y.-C. Tseng and A. Lleras, "Rewarding context accelerates implicit guidance in visual search," *Attention, Perception, Psychophysics*, vol. 75, no. 2, pp. 287–298, 2013.

[38] M. Luethi, B. Meier, and C. Sandi, "Stress effects on working memory, explicit memory, and implicit memory for neutral and emotional stimuli in healthy men," *Frontiers in Behavioral Neuroscience*, vol. 2, p. 5, 2009.

[39] D. J. Newman, "The double dixie cup problem," *The American Mathematical Monthly*, vol. 67, no. 1, pp. 58–61, 1960.

[40] M. Zellin, A. von Mühlenen, H. Müller, and M. Conci, "Long-term adaptation to change in implicit contextual learning," *Psychonomic Bulletin and Review*, vol. 21, no. 4, pp. 1073–1079, 2014.