Learning Assigned Secrets for Unlocking Mobile Devices

Stuart Schechter
Microsoft
stuart.schechter@microsoft.com

Joseph Bonneau
Stanford University & EFF
jbonneau@cs.stanford.edu

ABSTRACT

Nearly all smartphones and tablets support unlocking with a short user-chosen secret: *e.g.,* a numeric PIN or a pattern. To address users' tendency to choose guessable PINs and patterns, we compare two approaches for helping users learn assigned random secrets. In one approach, built on our prior work [16], we assign users a second numeric PIN and, during each login, we require them to enter it after their chosen PIN. In a new approach, we re-arrange the digits on the keypad so that the user's chosen PIN appears on an assigned random sequence of key positions. We performed experiments with over a thousand participants to compare these two repetition-learning approaches to simple user-chosen PINs and assigned PINs that users are required to learn immediately at account set-up time. Almost all of the participants using either repetition-learning approach learned their assigned secrets quickly and could recall them three days after the study. Those using the new mapping approach were less likely to write down their secret. Surprisingly, the learning process was less time consuming for those required to enter an extra PIN.

1. Introduction

Text passwords are no longer the dominant form of device authentication. Sales of smart phones, which almost universally support numeric PINs, far exceed sales of PCs. Tablets are also poised to overtake PCs in sales [53]. Regardless of whether these devices run Android, iOS, or Windows, they support authentication via simple device-unlock secrets, namely numeric PINs or graphical passwords. Even devices with fingerprint unlock typically fall back to secret-based authentication periodically for additional security or when fingerprints cannot be read.

While PINs and text passwords are both static user-chosen secrets, the reduced length and character set allow PINs to be entered in less time and on smaller screens, meeting usability requirements for mobile-device unlocking that text passwords cannot. Although mobile-device unlocking has stricter usability requirements, the consequences of having a mobile device compromised may be as dire as for computers with keyboards. Many companies allow their employees to access email and other services from their mobile devices. Mobile devices also now serve a critical role as second factors in website authentication, whether through access to users' emails and text messages or through dedicated applications for generating one-time codes.

To protect device-unlock secrets, most operating systems restrict guessing, either by limiting the frequency with which unlocks can be attempted, by erasing devices if too many consecutive unlock attempts fail, or both. Yet attackers may succeed even when restricted to a few guesses. They may exploit users' tendency to choose easy-to-remember but common numeric sequences (*1234*), repeat the same key (*9999*), or choose a path of adjacent keys (*2580*). They may guess the four-digit birth-year of the user or the users' loved ones—an estimated 25% of screen unlock codes are based on a date of some form, and 7% of respondents in a prior study admitted to using their own birthday as their banking PIN [18]. Leaked data on PINs chosen to unlock an iPhone application suggests an attacker with three guesses could expect as high as a 9.23% chance of success [18].

If, instead of choosing their own device-unlock secret, users were assigned a random four-digit PIN attackers would only have a 0.03% chance of guessing the PIN in three attempts. Furthermore, assigning secrets prevents users from re-using a PIN they have previously used elsewhere—though it does not prevent users from later re-using their assigned PIN elsewhere. If users are allowed to choose their own PINs, they may re-use the same PIN that they use for their ATM card, their voicemail, their frequent-flyer account, or their gym locker (in which they may store their phone or other mobile device). Indeed, over a third of respondents in a prior survey reported re-using their banking PINs for some other purpose [18].

We recently demonstrated a prototype ceremony for teaching users a random 56-bit secret, encoded as 12 characters or 6 words, using spaced repetition[16]. We integrated the ceremony into an existing website login process. Each time participants logged into the study website and verified their password, we displayed the text of the secret we had assigned to them and required them to type it into a text field. Each time they logged in, we added a progressively longer delay delay before revealing the secret for them to copy. Eventually, most participants began to recall and type their secrets reliably from memory.

While the prior approach was successful for participants using keyboards, we questioned whether users would accept the requirement to enter two PINs on a mobile device while learning an assigned code. Users expect their phones and tablets to unlock quickly, with minimal finger movement and delay. An in-situ study estimated typical users unlock their

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

phones nearly 50 times per day [35] and the time taken to unlock is already a significant annoyance to users.

In this paper we compare the prior approach to a new approach that does not require users to type additional keys to learn an assigned secret.

2. Learning Assigned Secrets via Mappings

Whereas our previous repetition-learning approach [16] requires users to enter the assigned secret after their chosen secret, our new approach teachers users a new assigned secret while they enter their chosen secret. The key idea is to assign users a random sequence, but provide their chosen secret as a guide to highlight the correct sequence. This is done by choosing a new random mapping from digits to positions on the keypad for each digit of the user's chosen PIN. The user simply needs to press each digit of their chosen PIN in sequence on four random-looking keypads, with the keypad appearing to shuffle before the entry of each digit. We illustrate our approach in Figure 5.

It is important that while the mapping of digits to keys changes with each key entered (between indexes in the sequence) it remains the same from login to login. This ensures that the user will be pressing the same sequence of key positions with each login. The assigned secret is this (random) sequence of positions on the keypad. We also provide letters for each position on the keypad which do not change—the assigned secret can equivalently be thought of as the sequence of letters needed to be pressed.

Essentially, our approach allows the user to enter their chosen secret and assigned secret at the same time. By using a random mapping, we ensure that even if users' chosen secret are drawn from an arbitrarily skewed distribution (including, as a degenerate case, if all users choose the exact same PIN) the distribution of assigned secrets will always be a uniform distribution over all possible sequences.

Since the pattern of keys in the assigned secret is the same each time a user logs in, we hypothesized that users would learn their assigned patterns from habit. To encourage users to learn, and to detect when learning had occurred, we add a delay before the digits appears that grows as the learning progresses. We draw arrows from key to key as users enter their assigned key sequences, making the visual pattern more salient. In the event that the same key appears two or more times in a row, we use within-key circular arrows. Line segments earlier in the sequence appear faded relative to those later in the sequence.

As with the prior scheme, attackers who obtain the device during the teaching period need only guess the user's chosen PIN in order to authenticate as the user. However, after learning the mapping can be destroyed and, unless attackers were already able to obtain it, knowledge of users' likely PIN choices will yield no benefit in guessing the assigned sequence of keys.

3. Related Work

3.1. Random passwords and PINs

It is now well-established in the research literature that humans will choose a skewed distribution of passwords or other secrets when given free choice [15, 42, 56]. This effect is robust across demographic groups [17, 49] and is impacted only marginally when users are more motivated to pick a strong password [15], are given stricter composition policies [42, 44] or are nudged to choose better passwords [70].

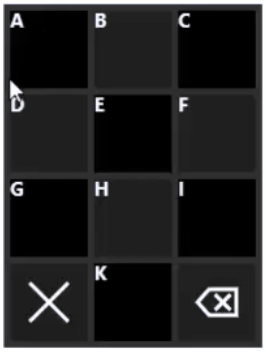
Random passwords. In response to persistent problems with weak human-chosen text passwords, a number of schemes have been proposed for encoding random passwords in such a way as to make memorization easier. Surprisingly, the few studies which have directly compared recall rates of user-generated passwords to assigned passwords have not found statistically strong evidence that users are less likely to remember assigned passwords than self-chosen passwords when no learning period is used [19, 63, 81, 85]. Various encodings have been proposed ranging from generating random but pronounceable nonsense words [1, 34, 80], choosing a list of random words from a dictionary [6, 46], generating a random grammatical sentence [7, 41] or even generating a random song [55]. No studies have actually validated that these encodings are more memorable than random character strings. Two studies which compared users' ability to recall random passwords under different encodings found no conclusive differences in memorability between random alphanumeric strings, random pronounceable strings or randomly generated passphrases [48, 62].

Spaced repetition. The hypothesis emerging from these results is that strong passwords, whether user-chosen or assigned, are not highly memorable without a learning period. Over a century of psychological research supports that spaced repetition [10, 20, 31] is the most reliable way to form long-term memories. While many other factors have been identified which affect the rate of memory formation, such as the depth of neural processing required during rehearsals [24] or the encoding of information in multiple forms [57], repetition is the most powerful and robust effect.

In recent work, we demonstrated the promise of spaced repetition for learning strong, 56-bit random text passwords for authentication [16], achieving 80–90% recall after a learning period of at most 15 days was followed by a period of at least three days during which the secret was not used. A subsequent study by Blocki et al. [13] demonstrated the effects over a longer period of time and with multiple passwords being memorized, observing recall rates close to 80% over a period of 180 days for a more complicated interface with graphical prompts for multi-word passphrases.

Numeric PINs. Relatively little has been published on numeric PINs. Historically, banking PINs were machine-chosen for technical reasons as well as security ones. Banks gradually began allowing user-chosen PINs in the 1980s as a marketing gimmick—they are now predominant. Bonneau et al. presented perhaps the only publicly-available estimates of the distribution of human-chosen PINs based on leaked data from an iPhone application developer, leaked web password data and surveys [18]. Their work highlighted that users' tendency to pick dates is the biggest source of skew in the data, consistent with research on dates chosen in text passwords [75]. Little work has focused on the memorability of random PINs; one exception is a pilot study by Huh et al. [38] which found memorability declines for longer PINs, although this could be improved by chunking them into smaller groups.

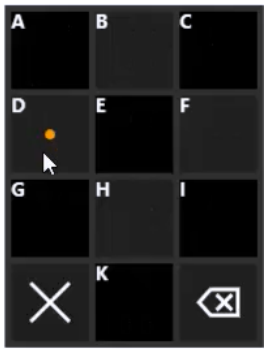
Graphical password schemes. A large variety of graphical passwords schemes have also been proposed [12, 59, 66]. Graphical passwords attempt to capitalize on the human brain's relatively strong visual memory, though different schemes have different goals. Traditionally the three ma-



(a) They keypad before the first digit is entered. We delay displaying digits.

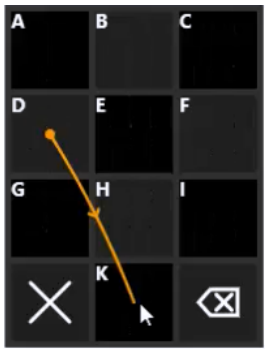
(b) We use a fade animation when displaying the mapping of digits to keys.

(c) The user presses the first digit, 1, at location D.



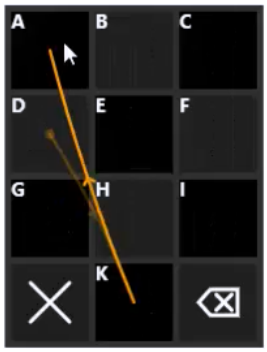
(d) After 1 is pressed, we again delay revealing digits for the second key.

(e) The users presses the second digit, 2, at location K.



(f) We immediately display an arrow from the first key pressed, D, to K.

(g) The user presses the third digit, 3, at location A.



(h) We connect the arrow path from $D \to K \to A$.

(i) The user completes the sequence by pressing digit 4 at location B.

Figure 1. In our mapping-based approach to learning a random secret, a user chooses her own PIN: in this example the four digits 1234. However, when she enters her PIN, the mapping of digits to keys on the keypad changes each time she enters a digit of her PIN; the letters at the top left of each key remain fixed. We choose the mapping of digits to keys so that the sequence of keys of the user's chosen PIN map to a randomly-chosen sequence of four keys: in this example the keys represented by $D \to K \to A \to B$. To encourage the user to learn to enter her key sequence without looking for the digits of her PIN, we increase the delay before we reveal the mapping of digits to keys with each login. Once the user has learned to enter the sequence of keys without seeing the digits, we can erase the mapping of digits to keys.

jor categories are recognition-based (searchmetric) schemes, in which a user recognizes previously-seen images [30, 64]; click-based (locimetric) in which users select points of interest in one or more images [21, 22, 79]; and recall-based or free-drawing schemes (drawmetric) in which the user draws an image or pattern [40, 67, 74, 78]. Note that secrets are typically assigned in recognition-based schemes, whereas recall- and click-based schemes tend to employ user-chosen secrets. A number of studies have demonstrated that user choice in graphical password schemes proposed for web authentication suffers from predictable choices much as do text password schemes [25, 68, 72, 73, 83].

Viewing a PIN-entry keypad as a visual stimulus with regions that users can press, our scheme could be considered a free-drawing recall-based graphical password. To our knowledge, no prior research has looked at how users memorize randomly-assigned secrets for free-drawing schemes.

3.2. Device authentication

User authentication for mobile devices, often simply called "device authentication" or "device (un)locking", is a burgeoning field of research. Our study appears to be the first utilizing spaced repetition to help users memorize a random secret for device authentication, but the literature suggests a number of interesting further research questions.

Device authentication habits. Several studies have looked at why, how, and how often users unlock their devices [32, 35, 71, 76] through a combination of surveys and telemetry on user devices. Collectively, these studies have found that between 40% [35] and 70% [32, 71] of users lock their phones, with a consistent preference for graphical unlock mechanisms over numeric PINs [32, 35, 71]. This dislike for numeric PINs was noted at least as early as 2002 [23], with most users choosing not to activate PIN activation on early generation (non-touchscreen) phones when this was the only option. Interestingly, despite this preference users are actually able to unlock more quickly and reliably using numeric PINs than graphical schemes [76].

A key challenge of unlock mechanisms compared to text passwords is the very high rate at which they are used; Harbach et al.'s telemetry study found an average of nearly 48 unlocks per day out of 80 total device activations (with some not requiring an unlock due to recent use or only performing a non-sensitive action) [35]. The frequency of unlocks, particularly as many are seen as unnecessary by users, motivates our goal of making learning as lightweight as possible. Parallel work in progressive and multi-level authentication [36, 60] aims to limit the number of unlock actions required of the user by delaying them until a security-critical action is attempted. This work is orthogonal to ours as our learning could be performed whenever a device authentication is needed, though we might note that if the rate of authentications per day were to become too low the speed of learning a new secret would decrease.

Security. Uellenbeck et al. [69] provide the only public estimates of the difficulty of guessing unlock patterns for Android's default scheme, a 3×3 variant of Tao et al.'s PassGo scheme [67]. By collecting patterns from a large number of users in an experimental setting and devising a dictionary to attack them, they estimate that this scheme provides roughly 8–10 bits of security for the median user and thus is roughly comparable to random 3-digit PINs. Thus all of our experimental treatments represent a security upgrade over the baseline Android scheme.

Windows touchscreen devices have used a modified click-based scheme, with a background image on which users enters a series of clicks, drags, or circles. Zhao et al. [83] studied the security of this scheme and found, depending on the background image in use, a dictionary with roughly 2^{18}–2^{20} items was sufficient to compromise the majority of users' patterns. They also found that a smaller dictionary of 2^{10} items was sufficient to compromise over 10% of user's patterns, indicating that even this stronger scheme still has many users picking predictable patterns for which any of our experimental treatments would be a security upgrade.

Strength meters have been proposed to nudge users towards choosing more difficult-to-guess unlock patterns [3, 65], but their effectiveness has not been established.

Other attacks on touchscreen authentication. Both PINs and graphical patterns are vulnerable to smudge attacks [9] or fingerprint attacks [82], in which residue from the user's fingers indicates where the user touches their screen during unlock. Defending against these attacks requires either randomizing the physical pattern input during any given authentication [2, 47, 61, 77], or switching to authentication schemes which do not require touching the screen such as gaze-based authentication [26, 45] or gesture-based authentication [8, 50, 51, 58]. PINs and graphical patterns are also both vulnerable to "shoulder-surfing" or physical observation attacks [54]. A number of authentication schemes have been proposed to defend against smudge attacks and shoulder-surfing [28, 29], though none of these schemes has seen practical deployment and they all appear to impose additional burden on the user.

Other device authentication mechanisms. Other research has attempted to replace explicit device unlocking completely. Physical biometrics deployed for smartphones include Apple's Touch ID fingerprint sensor and Android's Face Unlock face recognition scheme. User studies of these mechanisms find that users generally prefer using a fingerprint sensor and many find using face recognition annoying or impractical in certain situations (e.g. in a dark room) [11, 52]. Interestingly, convenience and perceived speed are the dominant factors, with increased security not being a major factor motivating adoption. Indeed, Apple's Touch ID is always configured to allow PIN or password authentication as a fallback; Apple's own documentation states that the primary goal of the feature is to allow users to use a stronger password since they won't have to enter it as often [5]. Similarly, Android's Face Unlock can be overridden by PIN entry. Thus, even with increased deployment of biometrics they are currently only a convenience and helping users remember stronger unlock codes is an important goal for security.

Behavioral biometrics [39, 43] capture a user's implicit actions using the device to detect if a different human appears to be using the device. For example, much research has shown that fine details of a user's touchscreen use can identify them [14, 27, 33, 37, 84]. However, this approach hasn't been deployed and it appears to inherently have sufficiently high false negative rate to require a more reliable backup authentication mechanism. Thus, our research is orthogonal to the challenge of using either explicit or implicit biometric signals to decrease the number of authentication requests imposed on the user.

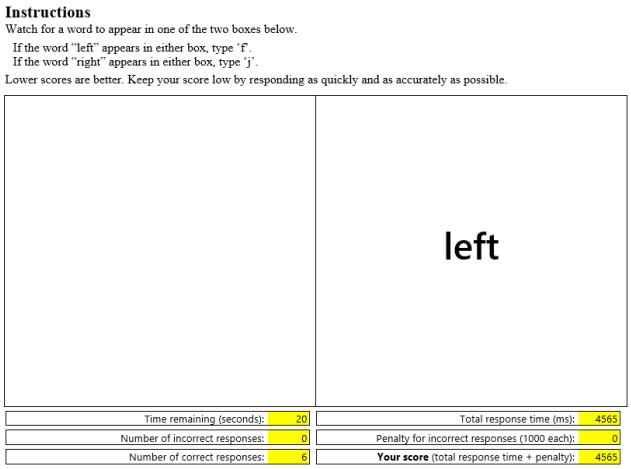


Figure 2. In the attention game, we asked players to press a key on the side of the keyboard that matches the word on the screen, regardless of which side the word appears on. Scores are based on response time and accuracy.

4. Methodology

In order to observe participants repeatedly entering a secret, we needed an excuse to cause them to authenticate. We asked participants to login to a study to perform the same distractor task used in our prior study: the attention game illustrated in Figure 2. We shortened the game to five attention trials (30 seconds). In each trial, we randomly choose a side of the screen (left or right) and a word ('left' or 'right'). We display the chosen word in the box on the chosen side. We ask participants to type a letter on the left side of the keyboard if they see the word 'left' and on the right side of the keyboard if they see the word 'right', and to ignore the side of the screen that the word appears in.

We recruited by offering the attention game as a Human Intelligence Task (HIT) on Amazon's Mechanical Turk. We paid US$0.10 and restricted our task to workers from the US. When workers completed the attention game we offered them the opportunity to become participants in our study. We offered $9 for 50 repetitions of the game: $0.10 per game plus a $4 bonus for completing all 50 and a survey at the end. We issued payments automatically on an hourly basis.

We required participants to wait 30 minutes between each game and gave them a total of 8 days to complete all 50 games. For each of the 50 games, participants would log in, play the attention-test game for 30 seconds, and then be shown a timer that counted down the 30 minutes until they could play again (another login would be required). This allowed us to collect data on no less than 49 authentications over 2–8 days. Participants could reload the study page and log in before they were allowed to play another game, but they would be forced to log in again after the time expired. As a result, the number of logins sometimes exceeded the number of games played.

4.1. Treatments

The process through which participants logged in to play the game depended on their treatment group. We created a total of ten treatments, which we list in Table 1 along with the probability that a participant would be randomly assigned to each treatment. We set the probabilities such that most participants (65%) would be placed in one of our four *primary* treatments: (1) used only a user-chosen PIN; (2) used only an assigned PIN to be memorized when setting up the account; (3) used a user-chosen PIN augmented with a second assigned PIN learned during a learning period, mirroring our prior work; and (4) used a user-chosen PIN mapped to a random sequence of keys—our new approach, as described in Section 2. These primary treatments were the focus of our a-priori hypotheses for which we perform statistical testing. The remaining six auxiliary treatments explore variants of the mapping-based approach.

Treatment	Length	Keys	p
Primary			
(1) ***User-Chosen***	**4**	**10**	.10
(2) ***Assigned***			.15
(3) ***Second-PIN***			.20
(4) ***Mapping***			.20
Auxiliary (variants of *Mapping*)			
(5) *4x20 Mapping*	4	20	.05
(6) *6x10 Mapping*	6	10	.05
(7) *6x20 Mapping*	6	20	.05
(8) *Instructionless*	4	10	.05
(9) *Arrowless*	4	10	.10
(10) *6x20 Arrowless*	6	20	.05

Table 1. Treatments followed by the probability that a participant would be assigned to that treatment (p). We assigned 65% of participants to our four *primary* treatments, shown in boldface: user-chosen PINs (1), traditional assigned PINs (2), a second assigned PIN entered after a chosen PIN (3), and our new mapping-based approach (4). Our a-priori hypotheses, for which we planned and ran statistical tests, focused on these four treatments. We assigned the other 35% of participants to variants of *Mapping*, examing such factors as the number of keys in the sequence (length) and the number of keys on the keypad (keys).

Regardless of treatment, participants used a standard 10-digit numeric keypad with digits placed in sequential order (left to right, top to bottom) when signing up for the study. (The same keyboard that appears in Figure 3.) The back arrow at the bottom right can be used to backspace one digit and the 'X' at the bottom left can be used to clear all digits. For consistency, small letters used in some treatments appear at the top left of each key regardless of treatment. For treatments (1)-(3), participants continued to use this standard 10-digit PIN-entry keypad throughout the study.

For all treatments, we provided a feature that would send participants a PIN reminder. This reminder contained the participant's assigned PIN only for the *Assigned* treatment (for which participants did not have a chosen PIN). For all other treatments (1,3-9), the reminder contained the participant's *chosen* PIN.

(1) User-Chosen

Participants in the *User-Chosen* treatment chose their own four-digit PIN and entered it on the standard keypad throughout the study. Since most users currently choose their own PINs, this treatment serves as a baseline of user-acceptability, speed of entry, and memorability.

Figure 3. The PIN-entry keypad for an example participant in the *User-Chosen* treatment. The keypad would look the same for participants in the *Assigned* treatment.

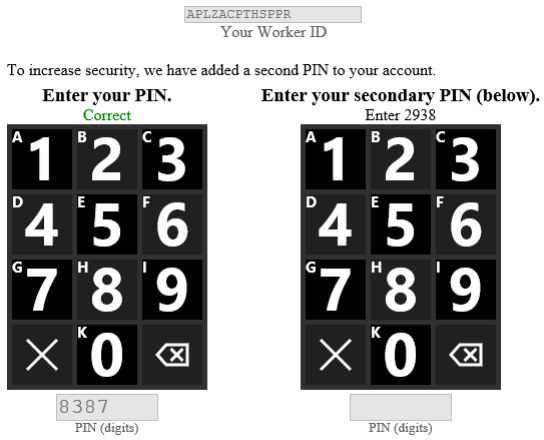


Figure 4. The PIN-entry keypad for the *Second-PIN* treatment after the example participant has entered her user-chosen PIN.

(2) Assigned

For each participant in our *Assigned* treatment, we randomly generated a four-digit PIN and instructed the participant to memorize it immediately. So as to not overly disadvantage this treatment for measures of memorability, we asked participants to enter their PIN twice during sign-up.

(3) Second-PIN

We created the *Second-PIN* treatment to mirror the two-secret approach from our prior work. This treatment appeared the same as the *User-Chosen* treatment when participants signed up. However, each time participants logged in with their chosen PIN, we asked them to copy their second secret, a four-digit PIN we had generated at random and assigned to them, using a second keypad—see Figure 4. Above the PIN-entry keypad, we provided participants the following *primary* instruction for each login: "To increase security, we have added a second PIN to your account." For each attention game they had completed after the first[1] (a lower bound on the number of prior logins), we added a 1/3 second delay before revealing the secret. So, by the login for the eigth game, a participant would need to wait 2 seconds before she could see the second PIN to copy it. If the participant entered the correct digit during the period before we revealed the digits to copy, we would conclude that she had entered the digit from memory. Each time a participant entered a correct key before the delay expired, we would start the delay over so that the participant would have as much time to enter the next key from memory as she had for the last key.

[1] We adjusted timings based on the number of games played, and not the number of prior logins, because participants might log in more than once per game. For example, a participant could refresh the study website, log in, and find they needed to wait until being allowed to play another game. They would need to log in again when the next-game timeout expired. Our decision to increase delays based on the number of games participants had played may have caused login-counts to grow without as much delays as we would have liked. The alternative, triggering on the number of prior logins, would have increased delays even when logins did not have sufficient spacing between them to reinforce learning.

After participants had completed five games, we added the following *supplemental* instruction:

> You do not need to wait for the second PIN to be written above the keypad to enter your PIN. If you recall the correct sequence of digits, you may enter it immediately.

To prevent participants from permanently tuning out the *supplemental* instruction text, we used a boldface font during logins following the 10th completed game and every 5 games after that (the 15th, 20th, ..., 45th). We removed the primary instruction after participants completed ten games. We removed both instructions immediately and permanently once the participant demonstrated that they had learned the assigned secret (by entering it before it was revealed).

(4) Mapping

The *Mapping* treatment is the baseline implementation of our new approach (Section 2), using a four-digit user-chosen secret (PIN) and an assigned secret (key/letter sequence) with 10,000 possible values. To assist learning, we employed the delay for *every* digit in the sequence. See Figure 1. As with the *Second-PIN* treatment, we used a 1/3 second additive delay.

We provided participants with the following *primary* instruction (with the same timing rules as the *Second-PIN* treatment).

> To increase security, we have changed the positions of the digits on the keypad you will use to sign in. We will use the same set of positions each time you sign in. That means that the pattern you make on the keypad when you enter the PIN will be the same each time. Entering your numeric PIN also creates a sequence of letters, representing the letters on each key of your PIN. This sequence of letters also stays the same each time you enter your PIN.

We used the following *secondary* instruction:

> You do not need to wait for the digits to appear on the keypad to enter your PIN. If you recall the

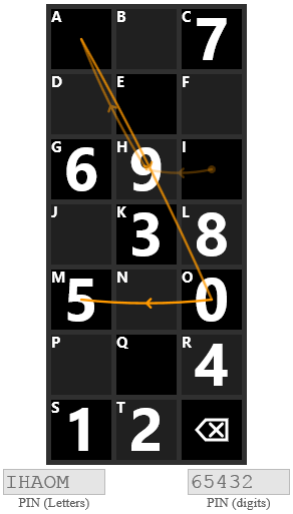


Figure 5. A six-digit mapping-based PIN on a 20-key pad. The user's chosen PIN was 654321. We assigned the user a random sequence encoded as letters $IHAOMS$. For each key the user enters, the mapping of letters to keys stays the same (see the top left corner of each key) but the digits move. We place digits so that the user's chosen PIN maps to the sequence of keys (letters) we assigned. In this screenshot we illustrate the moment at which a user had already typed 65432 at key positions $IHAOM$. Pressing the key labeled S with the 1 on it would complete the PIN. The arrows are an affordance to help users remember the pattern.

> correct pattern or sequence of letters before the digits appear on the keyboard, you may enter it immediately.

(5-10) Variants of Mapping

We created six more treatments to examine possible variations of the idea.

To test the effectiveness for memorizing more secure PINs (with more than 10,000 possible values), we created treatments (5)–(7) which encode larger secrets. In treatment (5), we expanded the PIN-entry keypad to include 7 rows. We also remove the clear function from the key on the bottom left so that digits could be placed on this key. This allowed us to double the number of usable keys from 10 to 20—see Figure 5 . The 10 digits were mapped onto these 20 keys with 10 keys left blank. Thus, while a participant would still choose a secret from a 4×10 space, the assigned secret (key/letter sequence) wass drawn from 4×20 (160,000) possible values—increasing security against guessing by a factor of 16. In treatment (6), participants chose a 6-digit PIN on a standard keypad and we assigned a six-digit secret with 6×10 possible values—increasing security against guessing by a factor of 100. In treatment (7), we combined the approaches of (5) and (6) to yield a 6×20 secret with 6,400,000 possible values—increasing security against guessing by a factor of 6,400 over the baseline *Mapping* treatment.

In treatments (8)–(10), we examined the impacts of taking away certain affordances in our design to see if they were actually needed, or just getting in the way.

When performing our prior work, we only told participants they *could* enter their secret without waiting for it to appear—we never asked them to. We were surprised how quickly they learned with such little guidance. We wondered whether it was necessary to provide any guidance at all. We created the *Instructionless* treatment (8) for which we removed both the primary instructions that explained the mapping of digits to keys and the secondary instructions that explained that the PIN could be entered before the digits appeared.

In treatments (9) and (10), we remove the arrow affordance from a 4-digit and 6x20 mapping treatments. As arrows may increase vulnerability to shoulder-surfing attacks, we would prefer to remove them if they had no benefit.

4.2. Study-completion survey

When participants logged in to complete their 50th attention game, we bypassed the game and immediately presented the completion survey.

Following standard demographic questions (language, gender, age, occupation, and level of education) we asked questions about the login process. We asked whether participants had entered their PIN "using a mouse, touch screen, or some other pointing device." We then asked if they had written or otherwise stored their PIN. To avoid confusion, we asked participants in the *Second-PIN* treatment only about their second (assigned) PIN. We then asked all participants in treatments using the mapping-based approach whether they had written/stored their assigned secret. We asked participants who reported storing their chosen or assigned secrets to explain how they had done so.

We asked participants using the mapping-based approach whether they remembered their secret as a visual pattern, a sequence of letters, or some combination of the two.

Finally, we asked participants in *all* treatments, "If you wanted to keep your phone or tablet secure, would you want to use a PIN like the kind you used to sign into our experiment's website?" For participants in treatments other than *User-Chosen*, we prefaced the question by explaining the security benefits of having an assigned PIN. For those in the *Second-PIN* treatment, we explained that a real system would allow users to discard their chosen PIN after learning—requiring only the second PIN.

We include our survey specification, with the exact wording used in the survey, as Appendix 10.

4.3. Follow-up study to test recall

Three days (between 72 and 73 hours) after each participant completed the main study, we emailed an invitation to participate in a follow-up for $0.50. The purpose of the follow-up was to determine if participants could recall their PIN after the learning period and having not used it for three days. We required only that participants log in to the study website, though we paid participants after a day's delay if they tried but failed to log in. As our goal was to measure memory after the learning period, participants in the mapping-based treatments were never shown the mapping of digits to keys, and those in the *Second-PIN* group were never shown their second PIN.

4.4. Hypotheses

We finalized the following seven hypotheses (four main hypotheses, three with two parts each) on the day we began our experiment, emailing a hash of the hypothesis statements to the program chairs as evidence that could be used

to prove these were a-priori hypotheses formed before examining experimental reuslts (see Appendix B). We present each hypothesis starting with the intuition behind it, then informally, and finally as a formal statement that can be used as a specification for a hypothesis test.

H1a/b. Study-completion rates

One of the motivations for the mapping-based approach is to reduce the frustration that may result when users are forced to attempt to memorize a secret in one session (without assistance) or if they are forced to type extra keys. We hypothesized that this frustration would cause higher drop-out rates for affected groups.

> A smaller proportion of participants in *Mapping* will drop out than of participants in (a) *Assigned* and (b) *Second-PIN*.

H2a/b. Written/stored PINs

Another motivation for the mapping-based approach is that users may be less likely to write them down. We intuited that users would be more likely to write down a PIN they were forced to memorize without a gradual learning period. We further assumed a second PIN, presented compactly as a numeric string above the PIN-entry keypad, would be easier to write down than a sequence of keys and so users would be more likely to do so.

> A smaller proportion of participants in *Mapping* will report having written down their assigned secret than of participants in (a) *Assigned* and (b) *Second-PIN*.

H3. Learning time

We hypothesized that, since the mapping approach requires users to press only half as many keys, learning an assigned sequence via a mapping would consume less of users' time than learning a second numeric PIN.

> Participants who complete the study in *Mapping* will have spent less time learning their secret than those in *Second-PIN*.

We define the time spent learning a secret as the sum of the PIN-entry time of each user's authentication sessions up to, but not including, the first session in which the user entered their assigned secret without assistance (the revealing of the positions of digits for *Mapping* or the display of the second PIN). We measure the PIN-entry time from the instant the PIN-entry keypad appears until authentication is complete. We cap the time consumed by any one PIN-entry session at 60 seconds.

H4a/b. Sentiment

Finally, we thought that participants would prefer learning an assigned secret via a mapping to receiving no assistance, or to having to enter extra keys during each login.

> When asked if they would want to use this system, participants in *Mapping* will answer more positively than those in (a) *Assigned* and (b) *Second-PIN*.

Since participants had three possible responses to this question, we use an using ordinal scale to measure positive sentiment: *no*<*maybe*<*yes*.

Hypothesis testing

We measure differences in proportion using a *two-tailed* Fisher's Exact Test (FET) and differences in both times and ordinal responses using a *two-tailed* Wilcoxon test with the Mann Whitney U statistic (U). To correct for multiple testing when examining these seven hypotheses, the conservative Bonferroni method yields a threshold of significance α=.05/7=.0071).

4.5. Ethics

Since some participants in our prior experiment had figured out that authentication was a focus of our study, in this experiment we revealed that the PIN was *a* component of our study. We used the study sign-up process to inform participants about the research in place of a standard informed consent form. We did not volunteer to participants that we would later give them the opportunity to participate in a follow-up study.

Unlike our prior work, we informed participants in advance that we would pay for each attention game they played even if they did not complete all 50—though we did provide a significant bonus to those who completed the study. This is more consistent with ethical guidelines that participants should know they may leave an experiment at any time without penalty.

We paid participants at a rate designed to ensure they received at least the highest minimum wage in the US. We identified that this was research being performed by Microsoft Research. We responded to workers' requests quickly and, where terms of service allowed, monitored worker forums to identify any participant concerns we might address.

Our study was approved by Microsoft Research's institutional review system. The second author, who was not employed by Microsoft Research, contributed after the last participant had already been recruited and was not involved in the conduct of the experiments or analysis of raw data.

4.6. Known limitations

While we strove to mimic as many aspects of a typical device-authentication experience as possible, we could not do so perfectly. While many users use a PIN to authenticate to their devices more frequently than once every thirty minutes [35], others may perform fewer than 50 PIN-based authentications over an eight-day period (such as those who bypass most PIN-authentications using their fingerprint). Participants in our study may have been more or less motivated to learn the PINs than real-world users would be.

Participants may have wanted to please the researchers by giving a more positive answer to our sentiment question, which asked whether they they would want to use the PIN from the experiment for their own mobile device. For this reason, we do not compare participants' reported sentiments for the *User-Chosen* scheme to others; Participants may be more likely to believe that the less-familiar schemes that assign PINs are schemes the researchers want to succeed.

Whereas PIN-entry on modern mobile devices uses a touch-screen keypad, participants in our study used our on-screen keypad via whatever computer and input device was available to them. Peeking ahead to our results, only 27 of 782 participants who completed the study (3%) reported that they primarily used a touchscreen to enter their PIN(s). The great majority of participants, 729 (93%), reported using a

mouse and 26 (3%) using some other device. We would anticipate device-unlock times would be shorter on touch screens, as users do not have to synchronize their hand movements with that of a mouse in order to press each key. While it's possible that using a device with a different form factor and input device impacted our between-group comparisons, we do not anticipate any reasons why one treatment group would disproportionately advantaged or disadvantaged.

5. Results

We offered our Human Intelligence Task (HIT) on Amazon's Mechanical Turk from 7:30PM EST on Sunday February 22, 2015 to 1:30PM on Wednesday February 25.[2] During this period, 1274 workers accepted the HIT and, of those, 1230 (97%) completed the HIT and saw the offer to sign-up for the study. Of those workers, 1016 loaded the sign-up page for the study (83% of those who completed the HIT). Since we assigned workers to a treatment group only after they arrived at the sign-up page, any departures prior to that should not be attributed to their treatment.

Of those workers who arrived at the sign-up page, 1001 (99%) completed the sign-up process to become study participants. Since more than four of every five workers who we paid to complete the attention-game HIT signed up to become study participants (1001/1230=81%), this recruiting strategy proved cost effective.

One factor contributing to the effectiveness of this recruiting approach was that many participants performed the HIT expressly because they had learned about the full study. The HIT appeared in discussions on forums for workers on Mechanical Turk, which ended up serving as feeders to our study. The forum that appeared to have the largest influence on our recruitment rate was MTurkGrind. After a pause in recruiting to ensure our funding would arrive in time to pay additional participants, we posted on that forum to let forum members know that the popular study was again open to new participants.

In monitoring these forums, we did not observe that any forum members had discovered that different participants were given different types of PINs, or any other "spoilers" that might have confounded the study. The one exception was that, towards the end, some posts revealed that the follow-up study was coming. For the most part, participants shared their best scores on the attention-game, encouraged each other, and shared their progress in completing the study.

5.1. Completion rates

Encouragement and competition on the forums may have caused a greater fraction of participants to complete the study, and possibly to do so at a faster rate, than they might have otherwise done. Another factor that may have raised completion rates is that we paid workers a higher wage than most requesters on Mechanical Turk. In the words of a participant who posted on the Turkopticon forum, "I think I got lucky to get in on this"[4].

With this in mind, it may not be surprising that we did not observe any significant differences in completion rates between treatment groups, meaning we had no support for Hypotheses 1a or 1b. As can be seen in Table 2, the proportion of workers arriving at the sign-up page who completed the study was *not* higher for the *Mapping* treatment (164 of 206, or 81%) than the *Assigned* treatment (128 of 158, or 82%, H1a: FET p=0.7395). In fact, a greater proportion of participants completed the *Assigned* treatment than the *User-Chosen* treatment (though the difference is well within the variance expected due to chance).

While fewer participants completed the *Second-PIN* treatment (132 of 179, 75%) as compared to *Mapping*, the difference was not significant (H1b: FET p=0.1731).

[2] There was a break in recruiting to ensure sufficient funds would be available for all participants.

Treatment	Didn't sign up	Quit quickly	Quit later	Finished
Assigned	2 (1%)	10 (6%)	18 (11%)	128 (81%)
User-Chosen	4 (4%)	17 (16%)	5 (5%)	83 (76%)
Second-PIN	4 (2%)	19 (11%)	24 (13%)	132 (74%)
Mapping	3 (1%)	16 (8%)	23 (11%)	164 (80%)
4x20 *Mapping*	1 (2%)	3 (7%)	1 (2%)	40 (89%)
6x10 *Mapping*	1 (2%)	7 (11%)	7 (11%)	50 (77%)
6x20 *Mapping*	0 (0%)	1 (2%)	3 (5%)	52 (93%)
Arrowless	0 (0%)	3 (3%)	8 (9%)	80 (88%)
6x20 *Arrowless*	0 (0%)	1 (2%)	2 (4%)	54 (95%)
Instructionless	0 (0%)	2 (4%)	5 (10%)	43 (86%)

Table 2. Participants' progress through the main study. We track all workers who arrived at the sign-up page, and were assigned a treatment, as participants assigned a PIN might abandon sign-up. We say that participants quit quickly if they completed no more than three attention tests (which would require two logins) or quit later if they otherwise failed to finish the main study.

5.2. Re-use and writing down of secrets

Recall that, for all treatments other than *Assigned*, the only observable differences in behavior at the sign-in page was the length of the PIN we asked participants to choose. We asked all participants who chose a PIN, with the exception of those in *Second-PIN*, whether they had chosen (re-used) a PIN they already used elsewhere and whether they had written down, or otherwise stored, their chosen PIN.
When asked if the PIN they chose was one they had used before, 229 of the 409 (56%) participants with a four-digit PIN reported that it was, as did 59 of the 146 (40%) participants with a six-digit PIN. Since six-digit PINs are less common than four-digit PINs, it is likely that fewer participants had a six-digit PIN already memorized to reuse.

Of the 180 participants who claimed not to have re-used an existing four-digit PIN, 41 (23%) reported that they had written down or stored their new chosen PIN, as opposed to 26 of 87 (30%) for six-digit PINs.

In Table 3 we examine the proportions of participants who wrote down the random secret assigned to them, those who needed a reminder of their PIN (their chosen PIN in all treatments except the *Assigned* treatment), and those who were unable to recall and enter their secret during the follow-up study.

In the *Assigned* treatment, 62 of 128 participants (48%) either wrote/stored their secret or later required a reminder. Surprisingly, if our participants are to be believed, the majority successfully memorized their PIN simply by entering it twice on the keypad of the study sign-up page! Still, we want to minimize the risk that nearly half of users will write or otherwise store their PIN, especially since they would

Treatment	Wrote assigned secret		Needed reminder		Never learned		Forgot later	
Assigned	57/128	(45%)	7/128	(5%)	~	~	0/110	(0%)
User-Chosen	~	~	0/83	(0%)	~	~	0/76	(0%)
Second-PIN	13/132	(10%)	4/132	(3%)	1/132	(1%)	0/124	(0%)
Mapping	2/164	(1%)	3/164	(2%)	1/164	(1%)	0/151	(0%)
4x20 *Mapping*	5/40	(13%)	0/40	(0%)	4/40	(10%)	0/30	(0%)
6x10 *Mapping*	3/50	(6%)	0/50	(0%)	2/50	(4%)	1/40	(3%)
6x20 *Mapping*	9/52	(17%)	0/52	(0%)	3/52	(6%)	2/43	(5%)
Arrowless	5/80	(6%)	3/80	(4%)	2/80	(3%)	1/70	(1%)
6x20 *Arrowless*	14/54	(26%)	0/54	(0%)	4/54	(7%)	2/48	(4%)
Instructionless	0/42	(0%)	0/43	(0%)	10/43	(23%)	0/28	(0%)

Table 3. The proportions of participants who reported writing down their assigned secret, requested and opened a PIN reminder, and of those who failed to login with their assigned secret during the follow-up. For the *needed reminder* column, note that our reminders contained users' *chosen* PIN for all treatments except *Assigned*. For the *forgot later* column, not that we exclude from our analysis those participants who never demonstrated learning their assigned secret (those in the *never learned* column).

likely carry a written reminder on them at the same time they were carrying their mobile device.

We had hypothesized that participants in the *Mapping* treatment would be less likely to write down their assigned secret than those in the *Assigned* and *Second-PIN* treatments. The proportion of participants who wrote or stored their assigned secret in the *Assigned* treatment (57/128, 45%) was significantly higher than the *Mapping* treatment, supporting Hypothesis 2a (2/164 1%): H2a FET $p<0.0001$. The proportion of those in the *Second-PIN* treatment (13/132, 10%) who wrote their assigned secret was also significantly higher than those in the *Mapping* treatment, supporting Hypothesis 2b: H2b FET $p=0.0008$.

5.3. Login and learning speed

We had hypothesized (naïvely, in retrospect) that participants in the *Mapping* group would spend less total time in learning – time learning their PINs – than those in the *Second-PIN* treatment. Whereas PIN-entry time for *Mapping* starts when the keypad appears and ends when the PIN is validated, the time for *Second-PIN* continues until the second PIN is validated. Recall that learning time is the sum of these PIN-entry times up to, but not including, the first login during which a participant enters their assigned secret before the secret or mapping is revealed. We present statistics elucidating the learning time in Table 4.
Turning to Hypothesis 3, comparing the learning times for the *Mapping* treatment and the *Second-PIN* treatment does reveal a significant difference: H3 $U=4{,}491.0$, $p<0.0001$. However, the direction of the difference was the opposite of what we had hypothesized!

Participants in the *Second-PIN* treatment required fewer logins to learn their secret, and thus had a lower learning time than those in the *Mapping* treatment despite having to enter twice as many keys (8 vs. 4) per login. We suspect participants in the *Second-PIN* required fewer treatments to learn because their assigned-secret was presented as a single chunk of four digits, whereas participants in the *Mapping* treatment were presented with their assigned-secret one key at a time (only seeing the final PIN as a chunk if they paid attention to the arrows or letters).

The impact of chunking goes beyond the number of logins required to learn the secret and also impacts the PIN-entry time for each login, which we infer from Figure 6. In fact, between the second and 13th logins, participants in the *Second-PIN* group were able to enter their 8 digits in less time, on average, than participants in the *Mapping* group could enter four digits! Again, *chunking* likely plays a role. We had employed a single delay before revealing the entire chunk of four digits to participants in the *Second-PIN* treatment, whereas we had employed four delays, one before revealing the positions of digits before *each* key of the PIN, for the *Mapping* treatment.

We suspect that participants in the *Mapping* treatment were also slowed down by their need to perform visual searching. Until they learned the positions of the digits of their PIN, they would have to perform four visual searches per login: one for each key. In contrast, participants in the *Second-PIN* treatment would find their four-digit secret displayed at an easy-to-find location (above the keypad) and the keys to enter these digits were at well-known positions—no visual searching was required.

Figure 6 does show that, once participants in the *Mapping* treatment learned their assigned key sequence, login speeds for the *Mapping* treatment decline rapidly and closely approach those of chosen PINs. Since users users often choose PINs with keys in close proximity, such as 1111 or 1212, we thought some fraction of participants might be slowed down by using keys at random positions. Yet, the $5th$ percentile times for the $49th$ login, representing the fastest 1 of every 20 participants, were tiny.

For the *Second-PIN* treatment, the login time per digit that needed to be entered was par with those of other treatments, and so we would expect performance equivalent to the *Assigned* treatment once users could skip their chosen PIN. While we had designed our *Second-PIN* treatment to mirror the approach in our prior paper, in retrospect we worried that this choice may have put it at an unfair disadvantage. For all other treatments, participants enjoyed post-learning login speeds the moment they learned their PIN. For the *Second-PIN* treatment, participants had to continue entering their chosen PIN even after they had learned their assigned PIN. We later addressed this methodological shortcoming, as we will explain in Section 6.

As expected, participants in treatments with more complicated secrets required longer learning periods, were more

Treatment	Logins to learn 50 %ile	95 %ile	Training time (sec) 50 %ile	95 %ile	Time for 49th login (sec) 5 %ile	50 %ile	95 %ile
Assigned	~	~	~	~	2.16	3.32	10.32
User-Chosen	~	~	~	~	1.82	2.97	7.09
Second-PIN	7.0	15.0	81	228	4.05	6.34	17.00
Mapping	12.0	23.0	172	507	1.98	3.09	7.67
4x20 *Mapping*	16.5	~	264	~	2.47	4.46	44.64
6x10 *Mapping*	16.0	38.4	346	1,579	2.70	4.65	20.28
6x20 *Mapping*	16.0	43.0	382	1,638	3.48	5.72	60.87
Arrowless	14.0	30.7	195	801	2.25	3.71	45.60
6x20 *Arrowless*	17.0	~	405	~	3.44	6.34	47.92
Instructionless	24.0	~	495	~	2.10	3.96	110.87

Table 4. Participants' performance on speed metrics, including (1) the number of logins prior to the first login in which they typed the code without seeing it, (2) the total learning time consumed by those learning logins, and (3) the time for participants' 49th login.

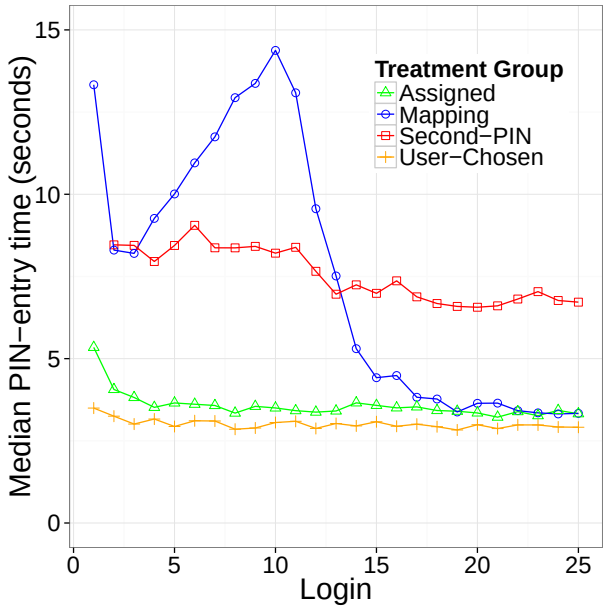


Figure 6. The median time to enter the login PIN(s) for the first through 25th login. The set of participants includes only those who completed the study. For the *Second-PIN* treatment, times include the time to enter both PINs.

Treatment	Pattern		Letters		Both		Did not	
Mapping	129	(80%)	14	(9%)	19	(12%)	2	(1%)
4x20 *Mapping*	18	(50%)	8	(22%)	10	(28%)	4	(11%)
6x10 *Mapping*	39	(81%)	5	(10%)	4	(8%)	2	(4%)
6x20 *Mapping*	22	(43%)	11	(22%)	18	(35%)	1	(2%)
Arrowless	26	(33%)	26	(33%)	28	(35%)	0	(0%)
6x20 *Arrowless*	7	(14%)	31	(62%)	12	(24%)	4	(8%)
Instructionless	31	(89%)	0	(0%)	4	(11%)	7	(20%)

Table 5. We asked participants in *Mapping* and its variants "If you learned how to enter your PIN on the keypad without waiting for digits to appear on the keys, how did you remember which keys to press?"

Treatment	No		Maybe		Yes	
Assigned	17	(13%)	44	(34%)	67	(52%)
User-Chosen	8	(10%)	25	(30%)	50	(60%)
Second-PIN	18	(14%)	43	(33%)	71	(54%)
Mapping	11	(7%)	53	(32%)	100	(61%)
4x20 *Mapping*	1	(3%)	14	(35%)	25	(63%)
6x10 *Mapping*	5	(10%)	13	(26%)	32	(64%)
6x20 *Mapping*	4	(8%)	19	(37%)	29	(56%)
Arrowless	4	(5%)	30	(38%)	46	(58%)
6x20 *Arrowless*	6	(11%)	13	(24%)	35	(65%)
Instructionless	2	(5%)	17	(40%)	23	(55%)

Table 6. We asked participants "If you wanted to keep your phone or tablet secure, would you want to use a PIN like the kind you used to sign into our experiment's website?"

likely to find a way to write or store their assigned secret, and were more likely to forget their assigned secret later. The *Instructionless* treatment had the greatest proportion of participants who never learned, required the most logins for those who did learn, and had the highest learning time—a clear indication that systems should provide *some* guidance.

Participants assigned to use mappings with the large keyboard were more likely to use the letters to help them memorize their secret, as can be seen from Table 5. Reliance on letters grew when we required a longer PIN or removed the arrow affordance. Those assigned the most difficult mapping (6x20 *Arrowless*) were the most likely to memorize the string of letters instead of a pattern of key positions. However, without instructions to inform them that the letters could be used to assist their memories, most participants learned their sequence as a pattern of key positions.

5.4. Participant sentiments

Table 6 summarizes participants' responses to the sentiment question, which asked whether they would want to use a PIN like the kind used in the study. For *Mapping*, 100 of 164 (61%) responded *yes*. Turning the three possible responses into an ordinal sentiment score (*no*=0, *maybe*=1, *yes*=2), participants in *Mapping* responded more positively than those in *Assigned* (52% yes) and those in *Second-PIN* (54% yes) as we posited in Hypotheses 4a and 4b, but the differences did not exceed our significance threshold: H4a: U=9,381.5, p=0.0772 ; H4b: U=9,805.5, p=0.1135

Written explanations in response to the sentiment question reveal that many participants were able to grasp what we were trying to accomplish in creating the mapping-based approach. In the words of one of our pilot participants:

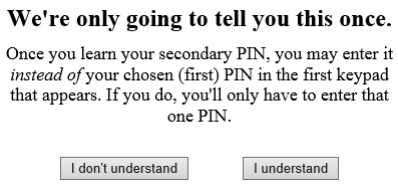


Figure 7. We presented this message on participants' first login after their third attention test. If they clicked on the "I don't understand" button we popped up an alert encouraging them to email us, then attempted to open a *mailto:* link to the study email address. None emailed us.

> This was pretty slick. I noticed I wasn't getting my PIN anymore but was still logging in. That's when I saw the pattern I had adapted to. I also realize that, in my head, I was repeating my actual credit card PIN which was not the PIN to get into the system. The numbers (and letters) were completely irrelevant and I thought that was awsome [sic].

6. Rematch: *Second-PIN* (v2) vs. *Mapping*

We conducted a second experiment to address concerns that we may have shortchanged the *Second-PIN* treatment. Users should be able to skip their chosen secret once they have learned their assigned secret. Whereas this design choice did not impact the hypotheses tested in our prior work, it may have put our *Second-PIN* treatment at an unfair disadvantage with respect to learning times and user sentiment.

We tested only two treatments in this experiment. We modified the *Second-PIN* treatment so that participants could enter their assigned secret in the first PIN-entry keypad and bypass the need to enter a second-PIN. For comparison we also included a *Mapping* treatment identical to that in our original experiment. For participants in the *Second-PIN* treatment, we presented the interstitial dialog in Figure 7 to participants just before presenting the PIN-entry keypad on the first login after the completion of their third attention test (which required a minimum of two prior logins).

Since, in the main experiment, 95% of participants in *Mapping* and *Second-PIN* had learned their secret by the 25th login, we shortened the study to 25 logins within four days for a total payment of $4.00.

Using data from our main experiment, we made one change to our calculation of learning time to better reflect the actual time lost to learning during each treatment. We subtracted 3 seconds for each learning login to account for time that would have been spent logging in even if no learning were taken place. We chose 3 seconds as it is approximately the median PIN-entry time for the user-chosen PIN treatment in the main study. This revised learning-period calculation better approximates the time users consumed due to the actual learning, excluding time the user would have had to spend logging in even if no learning were occurring.

We recruited participants for the rematch in bursts between 7:00PM EST on March 5 and 2:00PM EDT on March 9, excluding prospective participants who had participated in the earlier experiment.

We present the updated comparison of median PIN-entry times for these two groups in Table 7 and Figure 8, replicating the analyses that appear as Table 7 and Figure 6 from the main experiment. Since the *Mapping* treatment was unchanged, its results are similar to the same treatment in the previous experiment.

As expected, our improvements to the *Second-PIN* treatment made learning even faster. The median number of logins to learn (3) indicates that most participants were able to enter their assigned PIN from memory on their fourth login. In contrast, during the the main study it was only on the eighth login that we could determine that the majority of participants had learned their second PIN. It appears that, during the main study, participants in the *Second-PIN* treatment were either unable to enter the PIN quickly enough to prove knowledge of it (we revealed the digits for them to copy before they could enter the correct digit) or were less motivated to do so. For the *Mapping* treatment, in both the main and rematch experiments, the majority of participants did not enter their PIN without assistance until their 13th login.

For our rematch, the learning time for participants in the *Second-PIN* treatment was dominated by their first three logins. The first login required so much time, a median (over all participants) of 20.85 seconds, that it does not appear in our graph. The median learning time was only 40 seconds!

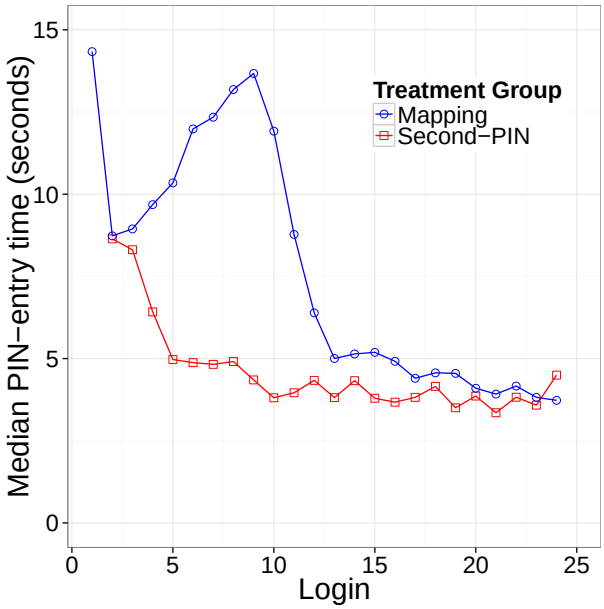


Figure 8. Rematch: median PIN-entry time for first 24 logins. Participants in the *Second-PIN* treatment could skip their chosen PIN and enter only assigned PIN. The first login time for the *Second-PIN* treatment (20.85 seconds) is outside the range of the graph.

In Table 8, we see that 20 of 73 participants in the *Second-PIN* treatment (27%) wrote down their assigned PIN. This proportion is not only greater than the *Mapping* treatment, as expected, but much greater than the *Second-PIN* treatment in the previous experiment. We fear that the cause of this difference, if not pure chance (**posthoc** FET: p=0.0011), was participants who wrote down their second PIN after learning they could use it to skip their first.

In Table 9 we summarize the rematch-study participants' responses to the sentiment question, which asked if partici-

Treatment	Logins to learn 50 %ile	95 %ile	Training time (sec) 50 %ile	95 %ile	Time for 24th login (sec) 5 %ile	50 %ile	95 %ile
Second-PIN	3.0	12.4	40	145	2.20	4.50	16.25
Mapping	12.0	26.0	117	412	2.51	3.73	33.93

Table 7. Participants' performance on speed metrics (see Table 4) for the rematch experiment.

Treatment	Wrote assigned secret		Needed reminder		Never learned	
Second-PIN	20/73	(27%)	0/73	(0%)	0/73	(0%)
Mapping	3/61	(5%)	1/61	(2%)	6/61	(10%)

Table 8. Rematch: Secret storage and recall. (Fewer *Mapping* participants learned their secret compared to the main experiment as they had half as many learning logins.)

Treatment	No		Maybe		Yes	
Second-PIN	8	(11%)	30	(42%)	34	(47%)
Mapping	7	(11%)	15	(25%)	39	(64%)

Table 9. Rematch: "If you wanted to keep your phone or tablet secure,would you want to use a PIN like the kind you used to sign into our experiment's website?"

pants would want to use the PIN scheme from the study on their mobile device. Surprisingly, we saw a drop in the desirability of the *Second-PIN*. Even had we hypothesized such a drop, we would not have the statistical strength to be able to dismiss the null hypotheses that both treatments inspire equally positive sentiment (**posthoc** U=1,874.0, p=0.1034). While there was insufficient evidence to prove a difference, it was enough for us to worry that our modifications had somehow made the *Second-PIN* treatment more annoying. To double-check, we examined participants' free-response answers and found no evidence to support this concern. The free-responses for *maybe* were consistently positive, suggesting that chance may have given us participants who rounded their scores down.

In fact, the most common concern focused on the trustworthiness of the party which generated the random PIN for the user. Since in most implementations the PIN would be generated by the device that the user is trusting to authenticate her correctly, this seems like an easy concern to overcome for any assigned secret.

7. Concluding discussion

Assigning users a random authentication secret, as opposed to letting them choose one, maximizes the difficulty of guessing (for a given alphabet/length) and prevents users from reusing prior secrets. We set out to test if spaced repetition, which we previously demonstrated for teaching users text passwords strong enough to resist extended brute-force [16], was also workable in the mobile device unlock setting for shorter, PIN-strength secrets.

We designed a new approach using randomly-assigned sequences (*Mapping*) hoping to make learning time as fast as possible which we feared would be a potential drawback of a direct application of our previous design for numeric PINs (*Second-PIN*). We in fact found the opposite, with users able to memorize a random PIN using our previous approach significantly faster than a random sequence using our new approach (particularly after the adjustment we made in the revision tested in Section 6).

However, both methods showed promise for use on mobile devices with very fast learning times. Both approaches also saw a smaller fraction of participants wrote down their assigned secret as the *Assigned* treatment, for which participants were asked to memorize a secret at sign-in time.

Our results do not yield a clear winner between *Second-PIN* and *Mapping* despite the shorter learning time for the former. *Mapping* offers the advantage that fewer users wrote their secret down, which may be attractive to system administrators who impose minimal-authentication requirements on devices used to access their systems (e.g., phones connecting to corporate email must have a PIN). Further, once participants had learned their secrets, login times for assigned secrets approached those for user-chosen secrets—remaining just a few percentage points higher.

With learning times of under a minute, the second-PIN approach requires surprisingly little effort. While a greater proportion of participants in our study reported wanting to use the *Mapping* treatment, we suspect that if prospective users knew which the approach required fewer learning logins and less visual searching, a substantial fraction of those might choose *Second-PIN*.

Acknowledgments

We are indebted to the hard-working and diligent participants from the Mechanical Turk marketplace who took part of the study, identified bugs, and took the time to write detailed explanations to survey questions. The presentation of this research benefited tremendously thanks to feedback from the anonymous reviewers, Simson Garfinkel (our shepherd), and Craig Agricola. Joseph Bonneau is supported by a Secure Usability Fellowship from Simply Secure and the Open Technology Fund.

References

[1] "Automated Password Generator (APG)". *NIST Federal Information Processing Standards Publication*, 1993.

[2] Irfan Altiok, Sebastian Uellenbeck, and Thorsten Holz. Graphneighbors: Hampering shoulder-surfing attacks on smartphones. In *Sicherheit*, pages 25–35, 2014.

[3] Panagiotis Andriotis, Theo Tryfonas, and George Oikonomou. Complexity metrics and user strength perceptions of the pattern-lock graphical authentication method. In *Human Aspects of Information Security, Privacy, and Trust*. Springer, 2014.

[4] AnotherPersona (Turkopticon Alias). Turkopticon review of microsoft research attention and pin study.

https://turkopticon.ucsd.edu/reports?id=A3A0N4OPTWRYPD, February 27, 2015.

[5] Apple, Inc. iOS Security. https://www.apple.com/business/docs/iOS_Security_Guide.pdf, April 2015.

[6] Reinhold G. Arnold. The Diceware Passphrase Home Page. , 2014.

[7] Mikhail J Atallah, Craig J McDonough, Victor Raskin, and Sergei Nirenburg. Natural Language Processing for Information Assurance and Security: An Overview and Implementations. In *Proceedings of the 2000 New Security Paradigms Workshop*. ACM, 2001.

[8] Md Tanvir Islam Aumi and Sven Kratz. AirAuth: Evaluating In-Air Hand Gestures for Authentication. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, pages 309–318. ACM, 2014.

[9] Adam J Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M Smith. Smudge attacks on smartphone touch screens. *WOOT*, 10: 1–7, 2010.

[10] Alan D Baddeley. *Human memory: Theory and practice*. Psychology Press, 1997.

[11] Chandrasekhar Bhagavatula, Blase Ur, Kevin Iacovino, Su Mon Kywe, Lorrie Faith Cranor, and Marios Savvides. Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption. *USEC*, 2015.

[12] Robert Biddle, Sonia Chiasson, and Paul C Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys (CSUR)*, 44 (4): 19, 2012.

[13] Jeremiah Blocki, Saranga Komanduri, Lorrie Faith Cranor, and Anupam Datta. Spaced Repetition and Mnemonics Enable Recall of Multiple Strong Passwords. *NDSS*, 2015.

[14] Cheng Bo, Lan Zhang, Xiang-Yang Li, Qiuyuan Huang, and Yu Wang. Silentsense: Silent user identification via touch and movement behavioral biometrics. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, pages 187–190. ACM, 2013.

[15] Joseph Bonneau. The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *2012 IEEE Symposium on Security and Privacy*, May 2012. URL http://www.jbonneau.com/doc/B12-IEEESP-analyzing_70M_anonymized_passwords.pdf.

[16] Joseph Bonneau and Stuart Schechter. Towards reliable storage of 56-bit secrets in human memory. In *Proceedings of the 23rd USENIX Security Symposium*. USENIX, August 2014. URL http://research.microsoft.com/apps/pubs/default.aspx?id=216728.

[17] Joseph Bonneau and Rubin Xu. Of contraseñas, sysmawt, and mìmă: Character encoding issues for web passwords. In *Web 2.0 Security & Privacy*, May 2012. URL http://www.jbonneau.com/doc/BX12-W2SP-passwords_character_encoding.pdf.

[18] Joseph Bonneau, Sören Preibusch, and Ross Anderson. A birthday present every eleven wallets? The security of customer-chosen banking PINs. In *FC '12: Proceedings of the the 16th International Conference on Financial Cryptography*, March 2012. URL http://www.jbonneau.com/doc/BPA12-FC-banking_pin_security.pdf.

[19] Julie Bunnell, John Podd, Ron Henderson, Renee Napier, and James Kennedy-Moffat. Cognitive, associative and conventional passwords: Recall and guessing rates. *Computers & Security*, 16 (7): 629–641, 1997.

[20] Nicholas J Cepeda, Harold Pashler, Edward Vul, John T Wixted, and Doug Rohrer. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132 (3): 354, 2006.

[21] Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. Graphical password authentication using cued click points. In *Computer Security–ESORICS 2007*, pages 359–374. Springer, 2007.

[22] Sonia Chiasson, Alain Forget, Robert Biddle, and Paul C van Oorschot. Influencing users towards better passwords: persuasive cued click-points. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1*, pages 121–130. British Computer Society, 2008.

[23] Nathan L Clarke, Steven M Furnell, Phihp M Rodwell, and Paul L. Reynolds. Acceptance of subscriber authentication methods for mobile telephony devices. *Computers & Security*, 21 (3): 220–228, 2002.

[24] Fergus IM Craik and Robert S Lockhart. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11 (6): 671–684, 1972.

[25] Darren Davis, Fabian Monrose, and Michael K Reiter. On User Choice in Graphical Password Schemes. In *USENIX Security Symposium*, volume 13, pages 11–11, 2004.

[26] Alexander De Luca, Martin Denzel, and Heinrich Hussmann. Look into my eyes!: Can you guess my password? In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 7. ACM, 2009.

[27] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 987–996. ACM, 2012.

[28] Alexander De Luca, Emanuel Von Zezschwitz, Ngo Dieu Huong Nguyen, Max-Emanuel Maurer, Elisa Rubegni, Marcello Paolo Scipioni, and Marc Langheinrich. Back-of-device authentication on smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2389–2398. ACM, 2013.

[29] Alexander De Luca, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. Now you see me, now you don't: protecting smartphone authentication from shoulder surfers. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, pages 2937–2946. ACM, 2014.

[30] Rachna Dhamija and Adrian Perrig. Deja vu-a user study: Using images for authentication. In *USENIX*

Security Symposium, volume 9, pages 4–4, 2000.

[31] Hermann Ebbinghaus. *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Duncker & Humblot, 1885.

[32] Serge Egelman, Sakshi Jain, Rebecca S Portnoff, Kerwell Liao, Sunny Consolvo, and David Wagner. Are you ready to lock? In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 750–761. ACM, 2014.

[33] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *Information Forensics and Security, IEEE Transactions on*, 8 (1): 136–148, 2013.

[34] Morrie Gasser. A random word generator for pronounceable passwords. Technical report, DTIC Document, 1975.

[35] Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[36] Eiji Hayashi, Oriana Riva, Karin Strauss, AJ Brush, and Stuart Schechter. Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 2. ACM, 2012.

[37] Grant Ho. Tapdynamics: strengthening user authentication on mobile phones with keystroke dynamics. Technical report, Technical report, Stanford University, 2014.

[38] Jun Ho Huh, Masooda Bashir, Hyoungshick Kim, Konstantin Beznosov, and Rakesh B Bobba. On the memorability of system-generated pins: Can chunking help? 2014.

[39] Markus Jakobsson, Elaine Shi, Philippe Golle, and Richard Chow. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX Conference on Hot Topics in Security*, pages 9–9. USENIX Association, 2009.

[40] Ian Jermyn, Alain Mayer, Fabian Monrose, Michael K Reiter, Aviel D Rubin, et al. The design and analysis of graphical passwords. In *Proceedings of the 8th USENIX Security Symposium*, volume 8, pages 1–1. Washington DC, 1999.

[41] Sundararaman Jeyaraman and Umut Topkara. Have the cake and eat it too—Infusing usability into text-password based authentication systems. In *Computer Security Applications Conference, 21st Annual*. IEEE, 2005.

[42] Patrick Gage Kelley, Saranga Komanduri, Michelle L Mazurek, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Julio Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *2012 IEEE Symposium on Security and Privacy*, pages 523–537. IEEE, 2012.

[43] Hassan Khan and Urs Hengartner. Towards application-centric implicit authentication on smartphones. In *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, page 10. ACM, 2014.

[44] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011.

[45] Manu Kumar, Tal Garfinkel, Dan Boneh, and Terry Winograd. Reducing shoulder-surfing by using gaze-based password entry. In *Proceedings of the 3rd Symposium on Usable Privacy and Security*, pages 13–19. ACM, 2007.

[46] Stanley A. Kurzban. Easily Remembered Passphrases: A Better Approach. *SIGSAC Rev.*, 3 (2-4): 10–21, September 1985. ISSN 0277-920X. doi:10.1145/1058406.1058408. URL http://doi.acm.org/10.1145/1058406.1058408.

[47] Taekyoung Kwon and Sarang Na. Tinylock: Affordable defense against smudge attacks on smartphone pattern lock systems. *Computers & Security*, 42: 137–150, 2014.

[48] Michael D Leonhard and VN Venkatakrishnan. A comparative study of three random password generators. In *IEEE EIT*, 2007.

[49] Zhigong Li, Weili Han, and Wenyuan Xu. A large-scale empirical analysis of chinese web passwords. In *Proc. USENIX Security*, pages 1–16, 2014.

[50] Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. User evaluation of lightweight user authentication with a single tri-axis accelerometer. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 15. ACM, 2009a.

[51] Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5 (6): 657–675, 2009b.

[52] Alexander De Luca, Alina Hang, Emanuel von Zezschwitz, and Heinrich Hussmann. I Feel Like I'm Taking Selfies All Day! Towards Understanding Biometric Authentication on Smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014.

[53] Ingrid Lunden. Gartner: Device Shipments Break 2.4B Units In 2014, Tablets To Overtake PC Sales In 2015, July 6 2014. URL http://techcrunch.com/2014/07/06/gartner-device-shipments-break-2-4b-units-in-2014-tablets-to-overtake-pc-sales-in-2015/.

[54] Federico Maggi, Alberto Volpatto, Simone Gasparini, Giacomo Boracchi, and Stefano Zanero. A fast eavesdropping attack against touchscreens. In *Information Assurance and Security (IAS), 2011 7th International Conference on*, pages 320–325. IEEE, 2011.

[55] Pascal C. Meunier. Sing-a-Password: Quality Random Password Generation with Mnemonics. 1998.

[56] Robert Morris and Ken Thompson. Password Security: A Case History. *Communications of the ACM*, 22 (11): 594–597, 1979. ISSN 0001-0782. doi:10.1145/359168.359172.

[57] Allan Paivio. Mental imagery in associative learning and memory. *Psychological Review*, 76 (3): 241, 1969.

[58] Shwetak N Patel, Jeffrey S Pierce, and Gregory D Abowd. A gesture-based authentication scheme for untrusted public terminals. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, pages 157–160. ACM, 2004.

[59] Karen Renaud and Antonella De Angeli. Visual passwords: cure-all or snake-oil? *Communications of the ACM*, 52 (12): 135–140, 2009.

[60] Oriana Riva, Chuan Qin, Karin Strauss, and Dimitrios Lymberopoulos. Progressive authentication: Deciding when to authenticate on mobile phones. In *USENIX Security Symposium*, pages 301–316, 2012.

[61] Stefan Schneegass, Frank Steimle, Andreas Bulling, Florian Alt, and Albrecht Schmidt. Smudgesafe: Geometric image transformations for smudge-resistant user authentication. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 775–786. ACM, 2014.

[62] Richard Shay, Patrick Gage Kelley, Saranga Komanduri, Michelle L Mazurek, Blase Ur, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Correct horse battery staple: Exploring the usability of system-assigned passphrases. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 7. ACM, 2012.

[63] Elizabeth Ann Stobert. Memorability of Assigned Random Graphical Passwords. Master's thesis, Carleton University, 2011.

[64] Adam Stubblefield and Dan Simon. Inkblot authentication. *Microsoft Research*, 2004.

[65] Chen Sun, Yang Wang, and Jun Zheng. Dissecting pattern unlock: The effect of pattern strength meter on pattern selection. *Journal of Information Security and Applications*, 19 (4): 308–320, 2014.

[66] Xiaoyuan Suo, Ying Zhu, and G Scott Owen. Graphical passwords: A survey. In *Computer Security Applications Conference, 21st Annual*, pages 10–pp. IEEE, 2005.

[67] Hai Tao and Carlisle Adams. Pass-Go: A Proposal to Improve the Usability of Graphical Passwords. *IJ Network Security*, 7 (2): 273–292, 2008.

[68] Julie Thorpe and Paul C van Oorschot. Human-seeded attacks and exploiting hot-spots in graphical passwords. In *16th USENIX Security Symposium*, pages 103–118, 2007.

[69] Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. Quantifying the security of graphical passwords: The case of android unlock patterns. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, pages 161–172. ACM, 2013.

[70] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, et al. How does your password measure up? the effect of strength meters on password creation. In *USENIX Security Symposium*, pages 65–80, 2012.

[71] Dirk Van Bruggen, Shu Liu, Mitch Kajzer, Aaron Striegel, Charles R Crowell, and John D'Arcy. Modifying smartphone user locking behavior. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 10. ACM, 2013.

[72] P. C. van Oorschot and Julie Thorpe. Exploiting Predictability in Click-based Graphical Passwords. *Journal of Computer Security*, 19 (4): 669–702, 2011.

[73] Paul C van Oorschot and Julie Thorpe. On predictive models and user-drawn graphical passwords. *ACM Transactions on Information and System Security (TISSEC)*, 10 (4): 5, 2008.

[74] Christopher Varenhorst, MV Kleek, and Larry Rudolph. Passdoodles: A lightweight authentication method. *Research Science Institute*, 2004.

[75] Rafael Veras, Julie Thorpe, and Christopher Collins. Visualizing semantics in passwords: The role of dates. In *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, pages 88–95. ACM, 2012.

[76] Emanuel Von Zezschwitz, Paul Dunphy, and Alexander De Luca. Patterns in the wild: a field study of the usability of pattern and pin-based authentication on mobile devices. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 261–270. ACM, 2013a.

[77] Emanuel Von Zezschwitz, Anton Koslow, Alexander De Luca, and Heinrich Hussmann. Making graphic-based authentication secure against smudge attacks. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 277–286. ACM, 2013b.

[78] Roman Weiss and Alexander De Luca. Passshapes: utilizing stroke based authentication to increase password memorability. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction*, pages 383–392. ACM, 2008.

[79] Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, 63 (1): 102–127, 2005.

[80] Helen M Wood. *The use of passwords for controlled access to computer resources*, volume 500. US Department of Commerce, National Bureau of Standards, 1977.

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

[82] Yang Zhang, Peng Xia, Junzhou Luo, Zhen Ling, Benyuan Liu, and Xinwen Fu. Fingerprint attack against touch-enabled devices. In *Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 57–68. ACM, 2012.

[83] Ziming Zhao, Gail-Joon Ahn, Jeong-Jin Seo, and Hongxin Hu. On the security of picture gesture authentication. In *USENIX Security*, pages 383–398, 2013.

[84] Nan Zheng, Kun Bai, Hai Huang, and Haining Wang. You are how you touch: User verification on smartphones via tapping behaviors. In *Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, pages 221–232. IEEE, 2014.

[85] Moshe Zviran and William James Haga. Passwords Security: An Exploratory Study. Technical report, Naval Postgraduate School, 1990.

10. Post-experiment survey

The remainder of this submission contains the survey we presented to participants in place of the 50th attention game, at the end of the main study (but before the follow-up).

Page 1

Congratulations!

You have completed all of your required attention tests. (We're not going to ask you to do the 50th.)

All you need to do now is complete this final survey.

Page 2

Is English your native language?

- Yes (811, 98%)
- No (14 2%)
- I don't understand the question (0, 0%)
- Decline to answer (0, 0%)

What is your gender?

- Female (329, 40%)
- Male (494, 60%)
- Decline to answer (2, 0%)

What is your age?

What is your current occupation?

Page 2

What is the highest level of education you have completed?

- Did not complete high school; High school/GED (7, 1%)
- High school/GED (83, 10%)
- Some college High school/GED (232, 28%)
- Associate's degree; High school/GED (94, 11%)
- Bachelor's degree (319, 39%)
- Master's degree (60, 7%)
- Doctorate degree (4, 0%)
- Law degree (8, 1%)
- Medical degree (7, 1%)
- Trade or other technical school degree (10, 1%)
- Decline to answer (1, 0%)

Page 3

The following question(s) are about how you logged into the attention study using your username and PIN.

During the study, did you enter your PIN using a mouse, touch screen, or some other pointing device? (If you used more than one input method, choose the one you used the most.)

- Mouse (729, 93%)
- Touch Screen (27, 3%)
- Other device (26, 3%)

[*If not in Assigned*]
Was the PIN you chose one you have used before, such as to protect a locker, debit card, credit card, or website?

- Yes (280, 50%)
- No (285, 50%)

[*Unless participant in Second-PIN*]
Did you store your PIN for the study website, such as by writing it down, emailing it to yourself, or adding it to a password manager?
[*If participant in Second-PIN*]
During the course of the study we assigned you a second numeric PIN to enter. Did you store that PIN, such as by writing it down, emailing it to yourself, or adding it to a password manager?

- Yes (148, 18%)
- No (677, 82%)

[*If anwered 'Yes' above*]
Please explain how and where you stored your PIN.

[*If in a mapping treatment*]
During the course of the study, as the delay before the digits of your PIN appeared grew longer, it became faster to sign in by pressing the keys before the digits appeared. In order to do so, did you store that pattern or sequence of letters, such as by writing it down, emailing it to yourself, or adding it to a password manager?

- Yes (44, 9%)
- No (438, 91%)

[*If answered 'Yes' above*]
Please explain where you stored this information, and whether you stored it as a pattern or as a sequence of letters.

[*If in a mapping treatment*]
If you learned how to enter your PIN on the keypad without waiting for digits to appear on the keys, how did you remember which keys to press?

- I remembered the position of each key on which the correct digit would eventually appear, which formed a pattern. (272, 56%)
- I remembered the letter of each key on which the correct digit would eventually appear, which formed a sequence of letters. (95, 20%)
- I remembered both.; (95, 20%)
- I never learned to enter my PIN without waiting for the digits to appear. (20, 4%)

Page 4

As we explained at the start of the study, we are experimenting with a new login system using a PIN.

[*If in a mapping treatment*]
Security researchers have found that computer users often choose predictable PINs, such as those that represent important dates or easy-to-enter patterns. One reason users choose predictable PINs is that it is hard to memorize less-predictable codes without practice.

[*If in a mapping treatment*]
With the PIN system used in this study, you practiced learning a more-secure a random code (the sequence of positions/letters on the keypad) each time you entered your numeric PIN (the digits that you chose when you signed up for the study). Once you learned the positions or letters, the sign-in system could discard the digits of your PIN and never show them again, leaving you with the secure random code that you had memorized through repetition.

[*If in Assigned* or *Second-PIN*]
Security researchers have found that computer users often choose predictable PINs, such as those that represent important dates or easy-to-enter patterns. In this study we assigned you a more secure randomly-generated PIN.

[*If in Second-PIN*]
For the purpose of the following question, assume that you had the option to remove the PIN that you had initial chosen once you had learned the more secure randomly-generated PIN that we assigned you. This option would allow you to login more quickly, using only four digits, instead of eight.

[*If not Second-PIN*] If you wanted to keep your phone or tablet secure, would you want to use a PIN like the kind you used to sign into our experiment's website? (If you also use a fingerprint reader, this would be the code you use when your device needs a stronger proof of your identity.)
[*If in Second-PIN*] If you wanted to keep your phone or tablet secure, would you want to use a randomly-generated second PIN like the kind you learned when signing into our experiment's website? (If you also use a fingerprint reader, this would be the code you use when your device needs a stronger proof of your identity.)

- Yes (478, 58%)
- Maybe (271, 33%)
- No (76, 9%)

Please explain your preference.

Page 5

Last question!

If you encountered any problems during the study, or any bugs in our study website, please let us know about them.

Page 6

You have now completed the entire study. Thank you so much for your time and attention. We will process payment within the next two business days. If your payment does not arrive within that time, please contact us at msrstudy@microsoft.com. (If you forget that address, you can also find it at the bottom of all the web pages on this site.)

You may close this tab at any time.

User ID	Treatment	Explanation
2255	*Mapping*	I simply remembered the pattern.
2400	*Mapping*	I just remembered it
2407	*Mapping*	I just remembered the patter [SIC] based on what my pin was.
2819	*Instructionless*	I just remembered the sequence
2836	*Arrowless*	I memorized the sequence of letters through rote memorization.
2849	4x20 *Mapping*	I just remembered where my numbers appeared

Table 10. Participants who answered *yes* when asked if they wrote down or stored their PINs, but then explained that they had only stored it in their memory.

A. Corrections to multiple-choice responses

We followed many of our multiple-choice questions with follow-up questions that asked participants to explain their multiple-choice response in written form. We used these responses to determine how well participants understood our questions and identify situations in which participants clearly misunderstood a question when answering it.

We discovered that, when we asked participants if they had written their pattern, some reported *yes* but then provided answers that clearly and unambiguously indicated otherwise. We expect this is because participants didn't realize our goal was to teach them the pattern, and interpreted that the question asked them to report storing the key in their own memory and using their memory to enter the PIN before the digits appeared. In presenting our results and performing our analyses, we disregarded a response of *yes* from participants in Table 10, substituting a *no* to reflect their explanation.

We audited responses to questions asking about whether participants in the *Second-PIN* treatment had stored their assigned (second) PIN, and did not find any evidence to suggest that any of those participants had misunderstood the question.

B. Evidence of a priori hypotheses

At 12:06PM eastern time on February 23, we sent the SOUPS program chairs the following base64 encoded SHA256 hash:
xVHlPGs/WkKQZvmAHxhWrwpjy/WCH9oB1GMupwzLx+E=

That hash was generated from the string below. The presence of "\r\n" indicates a carriage return and line break. Any white space, including line breaks produced by the formatting of this document (those not following "\r\n", indicates the presence a single ASCII space character. The numbering of hypotheses 3 and 4 were switched to facilitate exposition.

Hypothesis 1a/1b\r\n
Participants in PATTERN will be less likely to drop out of than those in (a) ASSIGNED and (b) SECOND_PIN\r\n
Statistic: Of participants who reached the sign-in page, the proportion who finish the study\r\n
Test: Fisher's Exact test\r\n
\r\n
Hypothesis 2a/2b\r\n

Participants in PATTERN will be less likely to write down their secret than those in (a) ASSIGNED and (b) SECOND_PIN\r\n
Statistic: Of participants who finished the survey, the proportion who reported writing their PIN or pattern\r\n
Test: Fisher's Exact test\r\n
\r\n
Hypothesis 3a/3b\r\n
Participants in PATTERN will report being more willing to use this authentication system than those in (a) ASSIGNED and (b) SECOND_PIN\r\n
Statistic: Of participants who completed the survey, the their answer to a question about whether they would want to use it was \r\n
scored 'no'=0, 'maybe'=1, 'yes'=2\r\n
Test: Mann Whitney U a.k.a. Wilcoxon\r\n
\r\n
Hypothesis 4: Participants in PATTERN will spend less time learning their secret than those in SECOND_PIN\r\n
Statistic: Of participants who completed the study, the aggregate pin-entry time in seconds from appearance of the PIN to completion, with no login taking more than 60s) from the first login until (but not including) the first session in which the participant entered the code correctly (but no more than the first 49 logins).\r\n
Test: Mann Whitney U a.k.a. Wilcoxon\r\n