An Empirical Study of Mnemonic Password Recall Errors

Aiping Xiong
The Pennsylvania State University

Huangyi Ge
Purdue University

Robert W. Proctor
Purdue University

Jeremiah Blocki
Purdue University

Ninghui Li
Purdue University

Abstract

A mnemonic password generation strategy with explicit instructions and an example of a personalized sentence was shown previously to increase the security of the resulting passwords. But the recall rate of the passwords was low. We report one online study quantifying *what* errors were made and *how often* they were made when participants confirmed and recalled passwords created with an instruction of mnemonic personalized example. The study also investigated whether an extra implementation-intention instruction improved the short-term and long-term recall of the passwords, but it was found to be ineffective. Error analyses revealed common failure types but varied rates across password confirmation and recalls. A handful of human memory limitations were also evident: 1) interference of association from common usage; 2) forgotten with a lack of encoding specificity; 3) forgetting/interference of the last letter position of passwords with limited memory span. Based on the findings, we provide suggestions to improve the mnemonic strategy.

1 INTRODUCTION

Passwords are widely used for authentication of users [3]. It is well known that passwords that are easy for someone to generate and remember often provide relatively little security, whereas stronger passwords are typically more difficult to remember. Previous interventions designed with the intention of guiding users toward strong passwords include use of policies describing necessary characteristics of a strong password and use of mnemonic strategies. Imposing numerous restrictions on acceptable passwords (e.g, requiring that a generated password must have symbols and numbers) was not found to be effective [10, 16], but mnemonic strategies for password generation have yielded promising results [4, 17].

One strategy that has been shown to be relatively effective is that of generating a sentence and then composing the password from the first letter of the words, along with insertion of a special character [16, 18]. Although the mnemonic sentence strategy can be effective for security, users often find complex strategies of this type unusable [18]. That is, the recall of the resulting password is not optimal.

Yet, little work has been done to understand why the recall rate is low. Prior work examined password recall errors, and measured kinds and rates of typographic errors made by users [5]. With a base of mnemonic sentences, we sought to quantify the nature of errors that participants make when they confirm and recall mnemonic passwords, i.e., whether those errors are due to forgotten, interference from other possible passwords, keystroke errors, or some other problems.

To this end, we conducted an online study on Amazon Mechanical Turk (MTurk). Participants were recruited to create a password using a mnemonic-personalized-example strategy, and to recall the password after a short distraction as well as one week later. We also examined whether an implementation-intention instruction in the form of an if-then plan helped participants recall passwords. We did not obtain any difference in recall between participants who received the extra implementation-intention instruction and those who did not. We obtained similar error types with varied rates across participants' password confirmation and short-term and long-term recalls. Analysis of errors also revealed a handful of limitations of human memory: 1) interference of association; 2) forgotten with a lack of encoding specificity; 3) forgotten/interference with limited memory span.

Our study results provide basic insights into the nature of errors that participants make during mnemonic password confirmation and recall, and underscore the importance of mitigating the limitations of human memory. Based on those findings, we suggest ways to improve the mnemonic strategy.

Who Are You?! Adventures in Authentication (WAY) 2019.
August 11, 2019, Santa Clara, CA, USA.

In Section 2, we review related work on mnemonic sentence strategies and implementation intentions. We describe the methodology of our study in Section 3 and present the results in Section 4. We conclude with a discussion the implications of our findings in Section 5.

2 RELATED WORK

Mnemonic Password Recall. Prior studies show that in general mnemonic sentences can be effective for the security of generated passwords based on both small and large sample sizes [9–11, 13, 16–18]. Nevertheless, low password recall rates, were still evident [16]. Yang et al. [18] examined the security and usability of passwords using 6 variants of mnemonic strategies in a series of online studies with 5,484 Amazon MTurk workers. They found an impact of the exact instructions on the security of the resulting passwords. Specifically, an explicit instruction to choose a personalized sentence, paired with an example, increased the security of the generated passwords significantly. However, only 40% of the passwords generated using the mnemonic strategy were recalled one week later.

Implementation Intentions. Implementation intentions specify a contextualized cue-action plan to enhance accessibility/activation of the anticipated situation and achieve the goal-directed behavior. For instance, if situation X is encountered, then I will initiate behavior Y in order to reach goal Z [6]. In a range of settings outside of cybersecurity, implementation intentions have been shown to be effective at enhancing the link between cue and action [8]. But whether implementation intentions are helpful at improving people's cybersecurity behavior, e.g., password recall, is unclear.

The goal of the current study was to quantify *what* errors were made and *how often* they were made when participants generated and recalled passwords using the mnemonic-personalized-example strategy. We also investigated whether an extra implementation-intention instruction in the form of an if-then action plan during short-term recall will be activated automatically in the long-term recall of passwords, and improve the recall success.

3 METHODOLOGY

Using a between-subject design, we studied two conditions, with the implementation-intention instruction presented at the short-term recall in the *II* condition but absent in the *Control* condition. Description of the extra implementation-intention instruction is:

- *Implementation Intentions (II)*: *If you need to use the password that you created, then you will retrieve the three steps of how you created the password to recall it.*

We examined the memorability of the generated passwords from two aspects: (1) short-term and long-term recall of the passwords; (2) errors made at both recalls. Time used for password generation, the success rate of password recall, password recall time, and the Levenshtein distance (also called edit distance) of failed password recall were measured. We also recorded the attempts that participants made during password confirmation and both recalls. Errors were categorized into *insertions*, *deletions*, and *substitutions* based on the edit distance analysis using the *adist* function in R. Because of our main interest being in errors due to using the mnemonic strategy, the three error types were classified into errors of *commission* and *omission* [12]. Commission errors refer to recalls that are different from what were encoded during password generation, and omission errors refer to recalls that forget part or the whole of generated password.

The effect of the implementation-intention instruction was evaluated by comparing the short-term and long-term recall of passwords between the two conditions. Participants in the *II* condition were expected to establish a *recall-three steps password generation* plan, and the main prediction was that the percentage of recalled passwords would be greater in the *II* condition than in the *Control* condition. We also evaluated the strength of the mnemonic sentences and passwords using $\tilde{\lambda}_1(\mathsf{top})$ and $\tilde{\lambda}_{10}(\mathsf{top}10)$ values [18] and the guessability of generated passwords for different guessing methods using the Password Guessability Service (PGS) data set [15].

3.1 Participants Recruitment

The study was conducted on Amazon MTurk. The human intelligence task (HIT) was posted with restrictions to US workers at least 18 years old and with 95% approval rate. The experiment complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at Purdue University. Implied consent was obtained for each participant. The experiment data were anonymized before analysis.

3.2 Procedure

After accepting the HIT on Amazon MTurk, all participants were directed from MTurk to a survey on Qualtrics, and were assigned to one condition randomly. Each participant was allowed to participate only once for the study. The study included two phases: 1) password generation and short-term recall; 2) long-term recall of the password.

In Phase 1, all participants were instructed to generate a password for their primary email account based on the mnemonic-personalized-example strategy that we adopted from [18]. The detailed instructions are as follows:

Imagine you are creating or updating a password for your primary email account. Create your password using the following three-step strategy:

1. *Think of a memorable sentence or phrase that is meaningful to you, and other people are unlikely to use. Make*

the sentence or phrase contain at least eight words. For example, "I went to London with my wife in June 2014."

2. *Select a letter (either uppercase or lowercase), number, or a special character to represent each word. For example, went –> w, London –> L, with –> &, four –> 4.*
3. *Combine them to create a password: iwtL&mWi*64

To mimic the password generation contexts in practice, we separated the sentence and password generation into two pages, so that the created sentence was not visible to participants during password generation. In the generation page, participants were also required to confirm the generated password. We explicitly asked participants not to use the example within the instructions. Participants who did not follow the instruction were excluded from data analysis. After password generation, participants filled out their demographic information as a distraction. Then, participants were asked to recall the password that they generated with two attempts. Participants in the *II* condition received the implementation-intention instruction before the recall whereas participants in the *Control* condition did not. Each participant was compensated $0.50 for completing Phase 1.

One week later, we invited each participant to return for Phase 2 to recall the password again with two opportunities. We sent the invitation through MTurk 6 days after Phase 1. We re-sent the invitation for another two days for participants who had not yet come back to the study. In Phase 2, returning participants were instructed to recall the password they generated and then to update the password. The implementation-intention instruction was not presented in either condition. Participants who finished Phase 2 received an extra $0.25.

4 RESULTS

Table 1 lists the general statistics of the study. Some participants' confirmation and recall attempts were not recorded, likely due to incompatibility of their browser and log code in the Qualtrics survey. We listed the number of those participants separately (see Table 1 w/o Log columns). The success rates and failure rates were calculated excluding those participants. For Phase 1, we list the number of participants, password creation time, and the statistics of short-term password recall. Those include the success rate of each recall attempt, failure rate, and password recall time. Results of Phase 2 were calculated and are presented in the same way as Phase 1. The numbers of participants who used the instructed strategy to update their passwords are also listed.

4.1 First Phase Results

We recruited 900 Amazon MTurk workers. Thirteen participants were excluded from the data analysis due to the incomplete results or failure to follow the password generation instructions. Additional 11 participants' results were excluded due to duplicated IP addresses. We also detected extra 104 participants with duplicated geographic locations[1]. Among the 104 duplications, 45 of them were repeated more than once. One identical geolocation generated 15 almost identical sentences and passwords like the example. Thus, we removed those participants from data analysis. A similar issue was also reported by Bai [2]. In the end, there were 376 participants in the *II* condition and 396 participants in the *Control* condition.

Participants' average age was 35.9 years, with 56.9% being less than 35 years. 57.0% of the participants were female, and 74.6% were college students or professionals who had associates, bachelors, or higher degrees. 81.2% of the participants claimed that they did not have a degree or work experience in computer science or related fields. The demographic distributions were similar for the two conditions.

Password creation time. The average time of password creation is listed in Table 1. With the same password generation instructions, the password creation time was similar between the two conditions, $t < 1.0$.

Password confirmation errors. Despite participants having just created the password, they began to forget password details during the confirmation. On average, participants spent 1.6 attempts to confirm the chosen password successfully. 25.6% (180) of them made confirmation errors with an average edit distance of 4.6. Confirmation error rates did not differ between the two conditions (Control: 25.6%, II: 25.6%). Errors of commission and omission collapsed across conditions were further categorized based on the mnemonic strategy and the most frequent errors are listed as follows. The first five are commission errors and the last one is omission error. Frequency of each error type is listed in Table 2.

- *Conversion error*: Participants followed the instructions and used a letter/number/special character to represent each word during mnemonic password generation, e.g., "t" → "Texas", "a" → "am", and "8" → "2018". But they recalled commonly used acronyms representing time, place, and digits during confirmation, e.g., "tx", "am", and "18". Participants also made errors when they confirmed symbols for converted words, including forgot the transfer, e.g., on: "@" → "o"; converted the unchanged ones, e.g., and: "a" → "&"; and confirmed different symbols for the converted words, e.g., born: "!" → "*".
- *Case error*: About half of the case errors were made at the first letter position of the created password. Participants also made a lot of case errors for letters representing time, people's name, and places, e.g., "Hawaii": "h" → "H", "Sissy": "S" → "s", and "June": "j" → "J".
- *Extra insertion*: Participants inserted extra letter/digit/symbol during the confirmation, e.g., "my sons mason and jackson smith": "msmjs" → "m**2**smjs",

[1] https://www.qualtrics.com/support/survey-platform/data-and-analysis-module/data/download-data/understanding-your-dataset/

Table 1: Statistics for user study. Succ1 and Succ2 mean the number of participants who successfully recalled the password on the first and second attempts, respectively.

Condition	Phase 1							Phase 2						
	Subject Number	Creation Time(s)	Short-term Recall					Number Returned	Long-term Recall					Update with Strategy
			w/o Log	Succ 1	Succ 2	Failed	Time(s)		w/o Log	Succ 1	Succ2	Failed	Time(s)	
Control	396	88.5	29	328 (89.4%)	21 (5.7%)	18 (4.9%)	19.8	183 (46.2%)	18	76 (46.1%)	9 (5.5%)	80 (48.4%)	75.6	72
II	376	85.6	40	304 (90.5%)	24 (7.1%)	8 (2.4%)	28.2	173 (46%)	19	63 (40.9%)	10 (6.5%)	81 (52.6%)	73.2	62
All	772	87.1	69	632 (89.9%)	45 (6.4%)	26 (3.4%)	24	356 (46.1%)	37	139 (43.6%)	19 (5.9%)	161 (50.5%)	74.4	134

and "three dogs: Sally, Joyce and Pretty": "3dsjp" → "3d**=**sjp". More than half of the extra insertions were obtained after the last letter position of passwords, e.g., "CNmYSomY" → "CNmYSomY**75**", "MdWi6yo" → "MdWi6yo**!**".

- *Keystroke error*: Participants mistakenly pressed neighbouring keys to the intended keys, e.g., "3" → "2", "o" → "0"; pressed same key twice, e.g., "o" → "oo"; or failed to press shift key, e.g., "%" → "5".

- *Another password*: Some participants forgot the whole generated password during confirmation, and recalled one that was totally different from the generated password, e.g., "cwdad@ro" → "R.oak225".

- *Forgotten*: Participants forgot part of the generated passwords. About half of such errors were due to omitting preposition words, e.g, "in": "iwtDW&mf**i**18" → "iwtDW&mf18". Participants also tended to forget the last letter position of the generated password during confirmation, e.g., "IwtSDiA**2**" → "IwtSDiA".

Table 2: Frequency of the most frequent errors for confirmation and recall errors. The first five are commission errors and the last one is omission error.

Error	Phase 1			Phase 2	
	Confirmation (180)	First Recall (71)	Second Recall (26)	First Recall (180)	Second Recall (161)
Conversion error	45.6%	18.3%	19.2%	21.7%	28%
Case error	31.1%	18.3%	23.1%	26.1%	27.3%
Keystroke error	12.8%	19.7%	0.0%	2.2%	3.7%
Extra insertion	15%	7.0%	23.1%	8.3%	13%
Another password	15.6%	0%	0%	34.4%	37.9%
Forgotten	15.6%	16.9%	19.2%	26.7%	26.1%

Short-term recall. After the demographic survey, participants in the *II* condition spent longer recalling their passwords (M=28.2 seconds) than did participants in the *Control* condition (M=19.8 seconds), $t_{(1,770)} = 37.96, p < .001$. This result provides evidence that participants in the *II* condition followed the instructions of retrieving the three steps of password generation. But the overall successful recall rate did not differ between the two conditions (*II*: 97.6%; *Control*: 95.1%), $\chi^2_{(1)} = 1.71$, $p = .192$. Neither the first recall success rates (*II*: 90.5%, *Control*: 89.4%), nor the second ones (*II*:7.1%, *Control*: 5.7%), showed any difference between the two conditions, $\chi^2_s < 1.0$.

Short-term recall errors. Excluding recalls of whole sentences (one in the first attempt and two in the second attempt), the average edit distance was 2.1 for all recall errors. For all first attempt errors (71), the edit distances of 1 or 2 took about 70.4% (50). 63.4% (45) of the first recall errors were corrected in the second attempt, 35 of which had distances of 1 and 2. 80% (20) of the second failed recall (25) also had edit distances of 1 or 2, suggesting that more short-term recall errors could be corrected if there were extra recall attempts.

The most frequent short-term recall errors were similar as those of confirmation errors except that the frequency distribution varied (see Table 2). All participants at least recalled part of the generated passwords. The conversion error rate of short-term recalls became smaller than that of confirmation, suggesting the effect of confirmation in helping participants consolidate the details of conversion. The average case error rate was about 20%, and more than half of the errors (12/19) were also obtained in the first letter position. Participants made more extra insertion errors when they attempted to correct first recall errors. The forgotten rates were similar as that of confirmation. Across all errors, only the keystroke errors obtained at the first recall attempt were all corrected during the second recall, indicating the effect of extra recall attempts in helping correct keystroke errors.

Password security. Average length of the generated passwords was 9.2, and did not differ between conditions (*II*: 9.3, *Control*: 9.1), $t < 1.0$. The obtained results were consistent with Yang et al. [18]. We did not observe any collisions among the generated sentences or the passwords (i.e., no repetition among chosen sentences and resulting passwords). Thus, $\tilde{\lambda}_1(\text{top})$ and $\tilde{\lambda}_{10}(\text{top10})$ values are also in agreement with the findings from Yang et al. [18] for this particular mnemonic strategy (see their Table 4).

PGS simulates password-guessing attacks using different approaches, including Markov models, a probabilistic context-free grammar (PCFG), John the Ripper (JTR), Hashcat, and neural network models [15]. PGS outputs the smallest guess number for each password using the above password-cracking approaches, and a minimum guess number (Min) across all cracking approaches (see Figure 1). From the figure, we can observe that all the cracking methods performed poorly. With Min, the most conservative security results, no more than 5% passwords were cracked within 10^{10} guesses, and less than 20% passwords within 10^{12} guesses, consistent with the results obtained by Yang et al. [18].

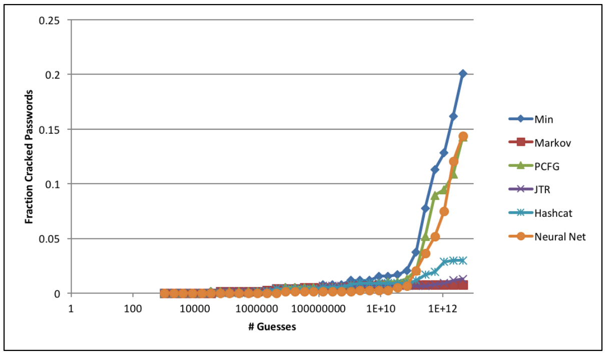

Figure 1: Guess number graph on passwords created by using mnemonic strategy with a personal example.

4.2 Second Phase Results

One week after completion of Phase 1, 356 (46.1%) participants returned to recall their passwords again. We did not present the implementation-intention instruction to participants, but expected that 173 participants in the *II* condition would retrieve the implementation intentions automatically.

Long-term recall. Participants spent similar time to recall the passwords in both conditions (*Control*: 73.2 s; *II*: 75.6 s), $t < 1.0$. The overall long-term success recall for the *II* condition (47.4%) was similar to that of the *Control* condition (51.6%), $\chi^2_{(1)} < 1.0$. The success rates of the first and second attempts for both conditions (see Table 1) also showed no significant difference, $\chi^2_s < 1.0$.

Long-term recall errors. One week later, for the first recall failure (180), the average edit distance was 6.3, with edit distances of 1 and 2 making up about 24.4% (44) of the errors. The average edit distance for the second failure was 6.6, with distances of 1 and 2 being 16.1% (26). Among the successful second recall (19), 68.4% (13) had edit distances of 1 and 2.

We analyzed commission and omission errors for both attempts as the short-term recall (see Table 2 last two columns). Large error rates of another password indicate that lots of participants totally forgot the generated password one week later and recalled those that they generated or used elsewhere. The forgotten rates were increased by about 10%. 29% of the forgotten were due to participants' failure to recall any item of the generated passwords, as indicated by inputs such as "no idea", "don'tremember", and "iforget". The conversion error rate of the first attempt was similar as those of short-term recalls but was increased in the second attempt. Compared to the short-term recalls, the case error rate was increased by about 6% on average. Like Phase 1, a lot of case errors (17/47) were still obtained in the first letter position of the passwords. We did not obtain a lot keystroke errors. Participants made more extra insertion errors in the second attempt, showing similar pattern as that of the short-term recall.

Different from confirmation and short-term recall, quite a few participants recalled passwords or part of the passwords that have similar meaning as the generated ones in the long-term recall (first recall: 20%, second recall: 21.7%). For example, "Yiwagm17" (Yeah I won a Gold Medal in 2017) → "IWNCI18" and "IWWCI2018"; "IjMua8WS" (I just made up a eight words sentence) → "Ijmua8WP"; "M2lcrLW" (My two lovely cats are Levi and Willow) → "ih2lcnlw".

Password update. At the end of the study, participants were asked to update the password without restriction. Except for four participants who chose their same passwords, 135 (38%) of the remaining participants updated their password with an extra mnemonic sentence and password. The results indicated that at least in the context of the present study, participants were willing to use the instructed strategy even though they were not forced to do so.

5 DISCUSSION

Different from our expectation, the extra implementation-intention instruction placed during the short-term recall did not improve participants' recall of generated password. The similar recall time obtained between the two long-term recall attempts indicates that participants in the *II* condition did not recall the three steps of password generation. Thus, future studies should evaluate when presenting the implementation-intention instructions will be helpful to establish the cue-action plan in an automatic manner.

We also quantified the kinds and rates of commission and omission errors that participants made when they confirmed and recalled passwords created with a mnemonic-personalized-example strategy. We found common error types with different rates across password confirmation and recalls. A prior study showed that case errors and keystroke errors can be effectively corrected by a typo-tolerant framework [5], so we focus on discussing conversion error, extra insertion, forgotten, and another password. Generally, those error rates were increased from short-term recall to long-term recall, with the largest increase obtained for errors of another password. Further examination revealed that those errors were made due to: 1) lacking an effective cue to recall the mnemonic password from other passwords; 2) interference from common usage, including acronym, typical symbol transfer, and capitalizing first letter of name/time/place; 3) forgetting exact information of mnemonic sentences; and 4) forgetting/interference of last letter position of generated passwords. Those patterns are closely related to human memory functions/limitations: interference with association [1], forgotten with a lack of encoding specificity [14], and forgetting/interference of the last letter position with limited memory span [7].

Across those errors, the error rates between the two attempts were similar for both short-term and long-term recalls, indicating that extra recall opportunities are not helpful to correct those errors. Thus, to mitigate those errors, improvements during password generation seem to be critical.

Analyses of confirmation and recall errors link our findings with basic memory theory, providing empirical grounding to improve the mnemonic strategies. For example, the average length of generated passwords in our study is 9.2. Prior studies showed that there is an upper limit on people's capacity (4 ± 2 chunks) to process information on interacting elements with reliable accuracy [7]. We suggest separating the mnemonic sentence into two or more chunks with less than 4 words in each, which is expected to accommodate the limitation of memory span and reduce some of the forgetting/interference errors. Also, commonly used acronyms and typical symbol transfers can be included in the mnemonic strategies to reduce conversion errors. 35% of the participants recalled another password during long-term recall, indicating that an effective cue to recall the generated password is missing. We conjecture that presenting the implementation-intention instruction during password generation will establish the encoding specificity [14] of the cue-action plan initially, which would improve the recalls afterwards.

6 Acknowledgement

This work was in part supported by NSF award #1704587.

References

[1] J. R. Anderson and G. H. Bower. *Human associative memory*. Psychology Press, New York, 1973.

[2] H. Bai. Evidence that a large amount of low quality responses on mturk can be detected with repeated gps coordinates, 2018. https://goo.gl/19KCHG.

[3] J. Bonneau, C. Herley, and F. M. Stajano. Passwords and the evolution of imperfect authentication. *Commun. ACM*, 58(7):78–87, 2015.

[4] J. Bonneau and E. Shutova. Linguistic properties of multi-word passphrases. In *International Conference on Financial Cryptography and Data Security*, pages 1–12. Springer, 2012.

[5] R. Chatterjee, A. Athayle, D. Akhawe, A. Juels, and T. Ristenpart. password typos and how to correct them securely. In *37th IEEE Symposium on Security and Privacy (S&P)*, pages 799–818, 2016.

[6] A.-L. Cohen, U. C. Bayer, A. Jaudas, and P. M. Gollwitzer. Self-regulatory strategy and executive control: Implementation intentions modulate task switching and simon task performance. *Psychological Research*, 72(1):12–26, 2008.

[7] N. Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114, 2001.

[8] P. M. Gollwitzer. Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54(7):493–503, 1999.

[9] J. Kiesel, B. Stein, and S. Lucks. A large-scale analysis of the mnemonic password advice. In *NDSS*, 2017.

[10] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2595–2604. ACM, 2011.

[11] C. Kuo, S. Romanosky, and L. F. Cranor. Human selection of mnemonic phrase-based passwords. In *Proceedings of the Second symposium on Usable Privacy and Security*, pages 67–78. ACM, 2006.

[12] H. L. Roediger and K. B. McDermott. Distortions of memory. In E. E. Tulving and F. I. M. Craik, editors, *The Oxford handbook of memory*, pages 149–162. Oxford University Press, New York, 2000.

[13] R. Shay, P. G. Kelley, S. Komanduri, M. L. Mazurek, B. Ur, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor. Correct horse battery staple: Exploring the usability of system-assigned passphrases. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 7. ACM, 2012.

[14] E. Tulving and D. M. Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5):352–373, 1973.

[15] B. Ur, S. M. Segreti, L. Bauer, N. Christin, L. F. Cranor, S. Komanduri, D. Kurilova, M. L. Mazurek, W. Melicher, and R. Shay. Measuring real-world accuracies and biases in modeling password guessability. In *24th USENIX Security Symposium*, pages 463–481. ACM, 2015.

[16] K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. B. Tai, J. Cook, and E. E. Schultz. Improving password security and memorability to protect personal and organizational information. *International Journal of Human-Computer Studies*, 65(8):744–757, 2007.

[17] J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password memorability and security: Empirical results. *IEEE Security & Privacy*, 2(5):25–31, 2004.

[18] W. Yang, N. Li, O. Chowdhury, A. Xiong, and R. W. Proctor. An empirical study of mnemonic sentence-based password generation strategies. In *Proceedings of ACM CCS*, pages 1216–1229, 2016.