Is Convenient Secure? Exploring the impact of Metacognitive beliefs in password selection

by

Mukund Choudhary, K V Aditya Srivatsa, Ishan Sanjeev Upadhyay, Priyanka Srivastava

in

CogSci 2021

Report No: IIIT/TR/2021/-1

Centre for Cognitive Science
International Institute of Information Technology
Hyderabad - 500 032, INDIA
July 2021

Is Convenient Secure?
Exploring the impact of Metacognitive beliefs in password selection

Mukund Choudhary and **K V Aditya Srivatsa** and **Ishan Sanjeev Upadhyay**
International Institute of Information Technology Hyderabad, Hyderabad, Telangana, India
{mukund.choudhary, k.v.aditya, ishan.sanjeev}@research.iiit.ac.in

Priyanka Srivastava
priyanka.srivastava@iiit.ac.in
Cognitive Science Lab, International Institute of Information Technology Hyderabad, Hyderabad, Telangana, India

Abstract

Recently, there has been research on what factors influence a user's password setting practices, which include various types of emotions such as anger, risk-taking tendencies, etc. However, research has shown that factors such as memorability and perceived memorability have a greater influence on password choice. Some recent research has shown a negative correlation between the *perceived memorability* and the *perceived security* of passwords, particularly passphrases (that are technically more secure). However, it is unclear whether this effect can be extended to groups with good experiences with digital spaces (IT professionals, entrepreneurs, etc.). Furthermore, it has not been determined whether random, uncommonly-worded, or complex structure passphrases would also maintain the correlation, as opposed to relatively less secure, common/simple passphrases. This study examines this problem using a diverse demographic and different categories of passphrases.

Keywords: Metacognition; Passwords; Perception; Memorability; Security

Introduction

Password strength is critical for healthy digital interaction, especially in recent years, with growing trends in the use of applications, digital devices, and other highly secure digital interfaces. It is becoming increasingly important to understand user's views on password setting practices. Nordpass (2020) reported that last year 2,543,285 people set their passwords to `123456`. This is a vulnerable password to choose, even though previous studies have shown that users know about standard safe password setting practices like using different types of characters, not using personal information, dictionary words, common combinations, etc. (Woods & Siponen, 2019).

This is not just limited to passwords, there is also a particular interest in passphrases. Recent studies have specifically found that the best way to create strong and memorable passwords is to use four or more words (Kävrestad et al., 2020), which implies that passphrases and their memorability is proving to be an increasingly important area.

There are numerous situational, theoretical, judgement-based factors found behind the unsafe password setting behaviour over the years. An interesting method is to link metacognitive theories to explain the same. Metacognition is thinking about thinking itself. The two major processes of "monitor" and "control" (Nelson, 1990), were interpreted in this context as "Users **monitor** passwords and decide on their security". This in turn "**control**s their decision" on whether to use the password in a particular environment.

This interpretation was tested by Luna (2019), which builds on research showing that memorability and security have a negative correlation, and the study examines whether perceived memorability (PM) has a similar correlation with perceived security (PS), i.e. do users believe "an easy to remember password is not secure"?

The study surveyed 40 Portuguese university students and found that the more heterogeneous a password is, the more secure it is perceived (PS), and the less memorable (PM). For example passwords with just lowercase characters (like `jfhdnele`) are less heterogeneous than passwords with a mix of lower and uppercase characters, symbols, and numbers (like `hR5@io88`).

They also found that passphrases were not considered (like `no longer freshman` (Luna, 2019)) as the most secure type of passwords. The authors concluded that the PS values for passphrases were ranked lower than some other categories. This is because their PM values were higher than most categories perceived as secure such as the category with a mix of lowercase, uppercase and numbers.

They concluded that PM & PS have a negative correlation and these results were reinforced by analysing intention of use levels in a critical vs. a non-critical website scenario. However, upon closer examination, we found that this study and other similar studies, did not focus on what could be these **other factors** (that affect the PS of passphrases), the participants **were not diverse** in terms of experience, and the passphrases that were used were **limited** to meaningful/easy to remember sentences.

In this study, we address these issues and hypothesize that a diverse participant base will **show trends** in the behaviour but also that it would generally be consistent. This is because even if a user knows these factors, they would ultimately act on perceptions and **not** to **facts**. Second, since we study pass**phrases**; the order of the words, their commonality, etc. may also be responsible for how users perceive their security. Therefore, we would **not** observe a strong negative correlation between PM and PS across *well-structured and simple passphrases* and *complex/uncommonly worded passphrases*. Finally, in order to understand the population that uses mobile devices frequently, we have also included use cases such as non-critical **mobile applications**.

Method

This study was conducted in 2020 and therefore had to be conducted "completely online" (due to the coronavirus pandemic). We selected Psytoolkit (Stoet, 2010, 2016) to script and float it to a diverse demographic of participants for 20 days in November 2020. This section describes the demography, the resources used for the survey, and the conduct of the survey.

Participants

We collected a total of 118 complete responses and after looking at our response times for the pilot ($N = 12$), we found that the minimum time required to complete the survey was approximately *10 minutes*. The 7 responses that took lesser time to finish and one response that indicated that the participant was uncomfortable with English were not included in our analysis. Finally, we had a set of 110 responses to analyse (42 Females), from a broad age group (range: 14-72 years; M=29.74 years; SD=13.3 years), a wide range of educational backgrounds, (12th grade or below: 6 (5.45%); college degree (current/completed): 63 (57.27%) and Masters/PhD, etc: 41 (37.27%)) and a diverse professional background (student: 55 (50.0%); unemployed: 4 (3.63%); retired: 6 (5.45%); employed: 45 (40.91%)). Since the survey material was exclusively in English, we asked the participants to report their knowledge of English (basic: 6 (5.45%); Good: 14 (12.72%); Professional: 48 (43.63%); Fully-Professional: 25 (22.72%) and Native-Speaker: 17 (15.45%)).

Material

Materials used for the survey, how they were collected, etc. are described below.

Passwords For the experiment 45 passwords were used, which were divided into 9 categories, with 5 passwords each. The 9 categories were: (***LF***): Low-Frequency Words (such as `meteoric`), (***HF***): High-Frequency Words (such as `children`), (***PD***): Pseudowords (such as `dwaughts`), (***LC***): Lowercase (such as `mjzxxvyt`), (***+U***): Lowercase + Uppercase (such as `ShpzczSo`), (***+N***): Lowercase + Uppercase + Numbers (such as `47Qn3nUD`), (***+S***): Lowercase + Uppercase + Numbers + Special Characters (such as `qy∼c)Aw4`), (***CP***): Common Phrases (such as `the book is under the table`), and (***RP***): Random/Complex Phrases (such as `shake medicine read floor`).

Categories 1 to 7 were all 8 characters long and the last two categories were 21-23 characters long with an average length of 22 characters. Methods of acquiring these passwords are fully reproducible and randomized where they could be. All the following passwords and passphrases were selected based on a normalised, averaged, and aggregated total of their security ratings by multiple websites as referenced (Kaspersky, 2019; My1Login, 2019).

For **Categories 1 and 2**, we used the MRC Psycholinguistic Database (Wilson, 1988) with filters on Brown Frequency, Kucera-Francis Frequency, and Thorndike-Lorge Frequency apart from length and then selected the results accordingly (for exact filtering methods, see the linked shared folder: `https://bit.ly/isconvenientsecure`). Similarly, for the 3rd category, we used the ARC Nonword Database (Rastle et al., 2002).

For **categories 4 to 7**, we used KeePassX 2.0.3 to generate passwords filtered by length, entropy, etc. For **Category 8**, we used English learning websites such as EnglishSpeak (EnglishSpeak, n.d.), to find introductory sentences in English and filtered them by length.

Finally, for **Category 9**, we used passphrase generators (randomised) such as "Use a Passphrase" (Hearn & Wheeler, n.d.), etc. and filtered some according to their length and whether they contained known but rare words.

IMS section Since the experiment was conducted **online**, the Immediate Mood Scaler (IMS) (Nahum et al., 2017) helped us determine the mental state of the participants and analyze whether they responded in a stable mood. It was the standard 24-item inventory with a 1-7 scale for mood pairs such as "depressed" or "happy", etc. Some items were: distracted or focused, hopeless or hopeful, etc.

Security Awareness section The main judgement tasks, were followed by a short questionnaire consisting of 10 objective questions. 8 of which were a basic security health and awareness assessment through questions such as "How often do you change your passwords?" and "Using characters of different types in a password is more secure than characters in the same category." (Yes/No). These were selected on the basis of previous studies and from inventories used in other password preference studies (Loutfi & Jøsang, 2015; Stainbrook & Caporusso, 2018; Luna, 2019).

The other two were binary answer questions that helped us understand whether particpants' beliefs matched the judgements done in the previous sections. The questions were *"A complicated/difficult to remember password is more secure."* (like `Tr0ub4dor&3`) and *"An easy to remember password is a safe password."* (like `correct horse battery staple`).

Procedure

We shared the link to the Psytoolkit form with willing participants who were informed that it took about 30 minutes to complete. The form consisted of 6 parts, which were presented to them in the following order:

Consent & Demographics In this part, the anonymity of the data and its use were clearly explained. The participants were also informed that this should be done without interruption except between some sections. We then asked for basic demographic details such as age, gender, profession, fluency in English, etc.

IMS After filling in the demographic data, participants read a description of the IMS scale and had to scale their emotions to the 24 items, according to their current behaviour.

Table 1: Security Awareness section Responses

#	Item	Response
1	Approximately how many passwords do you use on a daily basis?	M: 5.45 ; SD: 3.71
2	Frequently used passwords are easier to remember.	96.40% agree
3	An easy to remember password is a safe password.	36.94% agree
5	Using characters of different types in a password is safer than characters of the same category.	81.98% agree
6	Have you been hacked before?	10.81% say yes
7	A shorter password is less secure than a longer one.	59.46% agree
8	A password based on personal information or dictionary entries is secure.	16.22% agree
10	A complicated/difficult to remember password is more secure.	83.78% agree

Memorability Judgement This section was the first of three sections which presented the 45 passwords for user judgement. We asked participants to rate each of these passwords on a scale from 0 to 100% (less memorable to more memorable), by reflecting on the following prompt for each password: *"How likely are you to remember this password 2 days from now?"*.

Security Judgement Security was the second judgement task. We asked the participants to rate each of the 45 passwords on a scale from 1 to 6 (not secure to very secure) while thinking about the prompt: "How secure is this password?". The passwords were in the same order as in the last section.

Usability Judgement The final judgement task was to select all possible use cases for the given password. We listed 5 use cases for each of the 45 passwords and asked users to select all possible cases in which the displayed password could be used. A sample prompt: *For the password "sample_password", select all scenarios you could use this for: (Please select all situations that apply for the particular password. Do not select situations that do not apply for this password.)*

- *In an Important Online Service like banking online on SBI*
- *In a Casual Online Service like reading an article on Medium or some e-newsletter*
- *When registering is time-bound and you need to fill in a password quickly.*
- *Using Personal/Private Accounts on apps like Instagram or Facebook.*
- *Utility Apps/Gaming Apps like Calculator or Candy Crush, Temple Run, etc.*
- *None of the above.*

Security Awareness As mentioned in the Materials section, participants were asked 10 objective questions about their opinions and awareness of secure password setting practices. After completing this part of the survey, the participants were thanked for their participation and forwarded to the Google homepage via Psytoolkit.

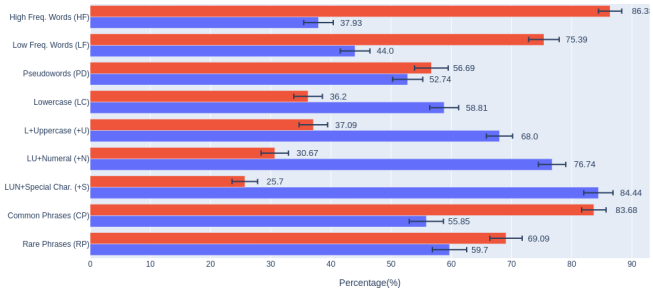

Figure 1: Mean ratings for PM and PS for each type of password. Key: (Blue: Perceived security, Red: Perceived memorability)

Results

For this section, Analyses of the variance (ANOVAs) were calculated across the seven measures (PS, PM, and usability in 5 environments) for each of the 9 password types, Pairwise Student's t-tests between the values of the 7 metrics for each password type (total 63 comparisons) are mentioned, indicating a trend. Pearson correlations are averages over the respective values retrieved from all 110 participants and the full record of the data as well as the statistics are available at: `https://bit.ly/isconvenientsecure`.

Perceived Memorability

An ANOVA showed significant differences, $F(8,848) = 119.02$, $p < .001$ (Figure 1). In decreasing order of PM ratings, HF and CP are ranked first, followed by LF, RP, and PD. LC, +U, +N and +S ranked lowest. In our study, however, the **passphrases** were additionally branched into CP and RP, revealing a significant difference between them in terms of their PM ($t = 12.17, p < .001$).

Perceived Security

An ANOVA showed significant differences, $F(8,848) = 96.89$, $p < .001$ (Figure 1). The general order of PS within password types showed the opposite trend compared to PM (except for CP and RP). In addition to previous studies, the highly negative Pearson's correlation $r = -0.92$ also supports this trend.

Table 2: question-4: "When do you normally change passwords?" Responses

Response	% of participants
On forgetting	28.82%
As the service reminds	20.72%
Depends on the service	20.72%
Regularly (every month)	7.21%
Rarely (annually)	17.12%
Never	5.41%

Table 3: question-9: "Check all options where you have passwords stored now:" Responses

Option	% of participants
Sticky notes on digital devices	13.51%
Noting offline	23.42%
Password Manager	15.31%
Nowhere (rely on memory)	48.65%
Rely on OTP/Forgot Password	18.02%
Some other place	21.62%

Usability in Specific Environments

Critical Services (CritWeb) An ANOVA showed significant differences, $F(8,848) = 190.08$, $p < .001$ (Figure 2). Of the 10 top-rated passwords for critical services, 5 were "+S" and 4 were "+N". All of these passwords are *character-level* and not dictionary entries. In Figure 2 as we move from top to bottom, we see a sharp rise in the *"Intention of use"* for Critical Services with the addition of more character classes (+U, +N), peaking at +S. These ratings closely follow the PS ratings, with a Pearson correlation of r = 0.93.

Non-Critical Services (NonCritWeb) Unlike in previous studies, the ANOVA showed significant differences in usability share *even* for non-critical services, $F(8,848) = 11.29$, $p < .001$ (Figure 2). Of the 10 top-rated passwords, 4 were PD and 2 of HF.

Time-Bound Services (Time) An ANOVA showed significant differences, $F(8,848) = 22.61$, $p < .001$ (Figure 2). Of the 10 top-rated passwords, 5 were LF and 4 HF. With a Pearson correlation of r = 0.82, the usability ratings for time-bound services resemble non-critical services.

Critical Apps (CritApp) An ANOVA showed significant differences, $F(8,848) = 15.35$, $p < .001$ (Figure 2).

Non-Critical Apps (NonCritApp) An ANOVA showed significant differences, $F(8,848) = 16.32$, $p < .001$ (Figure 2). With 4 PD and 3 HF among the 10 top-rated passwords, usability in non-critical applications shows a very similar behaviour to non-critical and time-bound services (Pearson's correlation of $r = 0.95$ and $r = 0.84$ respectively).

Finally, The 10 top-rated passwords *per usage environment* did not consist of Common or Random **Passphrases**.

Security Awareness

Tables 1, 2, and 3 (of the Security Awareness section) show that the majority of participants were aware of common safe password setting practices and had not yet been hacked. Question 7 confirms that not everyone knows that length is important for a technically more secure password.

79.28% of the participants stated that they use between 0 and 10 passwords daily (question 1), 47.7% rely exclusively on their memory to store the passwords (question 9), and 18.18% prefer to use the 'forgot password' option over memory. 50% of the participants who use 0-10 passwords daily change passwords only when reminded of it by the service, and a further 23.86% rarely change their passwords.

Demographics

We decided to do a correlational analysis between Demographics & passwords' Usability Environments, between Demographics & PM, and between Demographics & PS. We observed violation of normality assumptions by using Shapiro-Wilk test for each of the (above mentioned) series (p-values for all series were found to be less than 0.05). Thus we selected Spearman's correlation coefficient to study these correlations.

The participants were aged between 14 - 72, (M: 29.745

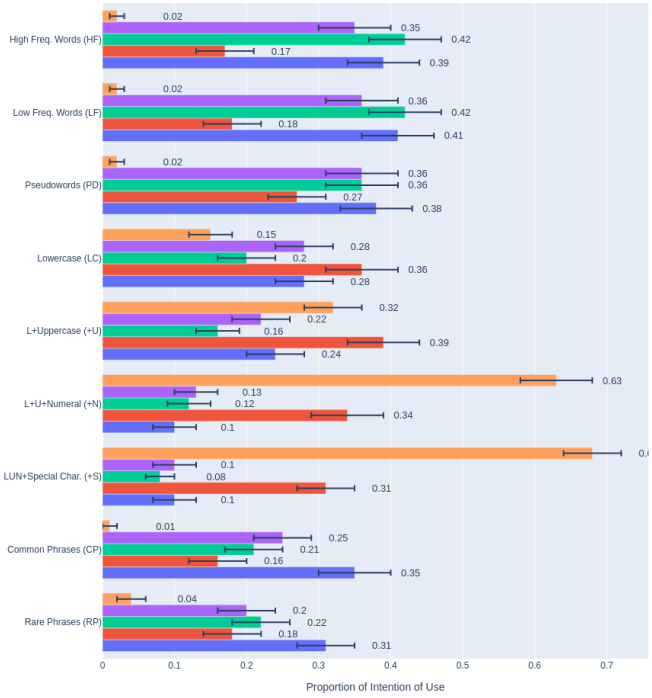

Figure 2: Mean ratings for Usability for each category. Key: (Red: Critical apps, Blue: Non-critical app, Orange: Critical services, Green: Time-bound services Purple: Non-critical services)

Table 4: Spearman correlation between user-demographics and usability-(password-type) pairs

Demo-graphic	Use Case(s)	Password Type(s)	Spearman Rho; p-value
eng	webapp	heterogeneous	0.2449 ; **
	critapp	common	-0.3109 ; ***
	critapp	rare	-0.2412 ; *
	crit	HF	-0.3755 ; ***
	crit	LF	-0.369 ; ***
	app	HF	-0.2851 ; **
	crit	+S	0.2675 ; **
edu	time	passwords	-0.2663 ; **
	noncritapp	common	-0.2493 ; **
occ	critapp	heterogeneous	-0.2693 ; **
	webapp	+U	-0.2464 ; **
age	critapp	heterogeneous	-0.2747 ; **
	webapp	+U	-0.2807 ; **

; SD: 13.3). There were 42 women (38.18%), 66 men (60%), and 2 preferred not to say. 6 participants were currently enrolled in a school (5.45%), 63 in a college/university (57.27%), and 41 have graduated or are pursuing a higher degree (37.27%). There were a total of 55 students (50%), 4 were unemployed currently (3.64%), 45 were employed (40.91%), and 6 were retired (5.45%). Finally 6 participants reported *Basic* understanding of English (5.45%), 14 reported *Good* (12.73%), 48 reported *Professional* (43.64%), 25 reported *Fully Professional* (22.73%), and 17 reported themselves as *Native Speakers* (15.45%).

We also used some groupings of password categories for the results in this section (see Tables 4 & 5), they are as follows: *"words"* includes High & Low Frequency words, *"heterogeneous"* includes Lowercase, Lower+ Uppercase, Lower+ Upper+ Numerals, and Lower+ Upper+ Numerals+ Special Characters, *"common"* includes High frequency words and Common Phrases, *"rare"* includes Low frequency words and Rare Phrases, and finally *"passwords"* includes all categories except Common and Random Phrases.

Similarly, some usability environments were also grouped: *"crit"* includes Critical Apps and Services, *"webapp"* includes all (critical or not) Apps and Services, and *"app"* includes Critical and Non-Critical Apps.

The demographics (apart from *age*) were numericalised using the following mapping:

- *English Proficiency (eng)*- 1: No knowledge, 2: Basic, 3: Good, 4: Professional, 5: Fully Professional, and 6: Native Speaker.
- *Profession (occ)*- 1: Student, 2: Unemployed, 3: Employed, and 4: Retired.
- *Education (edu)*- 1: No Schooling, 2: 12th Grade or below, 3: College Degree, and 4: Masters/Doctorate etc.

Table 5: Spearman correlation between user-demographics and (PM-PS)-(password-type) pairs

Demo-graphic	Rating Type	Password Type(s)	Spearman Rho; p-value
eng	PM	all	0.2774 ; **
	PS	words	-0.2715 ; **
	PS	LF	-0.2601 ; **
occ	PM	LF	-0.2616 ; **
	PM	PD	-0.3173 ; **
age	PM	all	-0.2635 ; **
	PM	LF	-0.2771 ; **
	PM	PD	-0.3537 ; **

Table 6: t-test value experiment

Password-Type Pair	t-value (PM)	t-value (PS)
PD and +S	$10.1 \approx 10.0$	17.099
LC and +S	$9.955 \approx 10.0$	20.74
CP and RP	*12.17*	*4.638*
HF and RP	14.45	40.034

Discussion

In this section we explain the results obtained and present our inferences.

Perceived Memorability & Security

The basic results for PM and PS are consistent with the previous studies. However, we saw that the results from additional branching in the PM section underscored the need to consider different types of passphrases based on their structure and vocabulary.

Passphrase Experiment In order to determine an expected variation in PS ratings, we found pairs whose t-values (for PM) are close to the CP-RP pair (Table 6). These pairs were ordered by PM. We observe that the PS are also in ascending order, except for the CP-RP pair. It shows a much lower t-value for PS ($t = 4.39$, $p < .001$) compared to the t-values of the closest pairs ($t = 20.74$, $p < .001$ and $t = 40.03$, $p < .001$ in order).

This suggests that PS is not influenced by PM only. *Other factors also play a role*, otherwise, we would have seen a much larger variation between the PS ratings for CP and RP.

Usability in Specific Environments

We see that the *Critical Services* results show that PS is the major control variable for the usability of a password in a "critical service" and that the type of distribution in *Non-Critical Services* suggests a shift towards the use of word-like passwords, suggesting that PM becomes the deciding factor as the relative severity of the usage environment declines.

However as compared to Critical Services (not mobile applications), the usability distribution of *Critical Apps* is much more distributed across the categories. Since the criticality of

the environments is equivalent, the preferred password-types are the same (***means** of distribution* of the two distributions show negligible differences), but the difference in the environment (web services vs. mobile applications) influences the general agreement on the preferred password-types (variance for critical-applications is much higher).

We can also see that results from *Time-Bound Services* and *Non-Critical Apps* suggest that participants in these use cases give a higher preference to retrievability than PS i.e. password types that have significantly higher PM than those preferred for critical use.

Finally, *Passphrases* not being considered usable across different use cases can be explained by observing that Common and Random passphrases fail due to their low PS ratings in use cases dominated by high PS passwords.

Demographics

We discuss a few significant correlations (as obtained from Tables 4 & 5) between some demographics and use cases, PM or PS, below:

We can see that the participants who are more proficient in "English" (as self reported), also have lower preference for meaningful words and phrases in critical scenarios. This is reinforced by the similarly low preference of high frequency passwords in even non-critical applications and the preference to use highly heterogeneous passwords (that don't have a meaning), in critical environments. Education levels give slightly ambiguous results where the participants show a low preference for generic password types in a time constrained scenario, this might be because of how difficult it is to retrieve such a string on a short notice. Finally, from Table 4 we can also see that "profession" and "age" show similar results, an older participant shows higher correlation with "simpler" passwords in most scenarios.

Moving onto the correlations with PM and PS (Table 5), we see that reported-proficiency in "English" seems to correlate with higher memorability ratings, while security ratings follow the opposite trend. This result seems to be aligned with the gradient of password types, ranging from heterogeneous strings to meaningful words and phrases, which allows participants of different language proficiency level to gauge the passwords accordingly. The "profession" and "age" demographics indicate an opposite trend compared to English-proficiency. This may be supported by the fact that a younger and/or working (not retired) participant will be both exposed to many password types, and would use more passwords on a daily basis.

Conclusion

In short, our results show that the negative correlation between PS and PM in passwords is strong for a large and varied demographic. Combined with previous studies, this also shows its true for different languages, experiences, etc. This also correlates with password choice in a few different use cases, e.g. Passphrases are not a popular choice for any use case, but the password `@?kUGS8o` was almost unanimously the best choice for a critical website because of its heterogeneity, and that Pseudo**words**, Low-frequency **words**, and High-frequency **words** were the most popular choices for use in Non-critical websites, mobile applications, and in a time-bound scenario.

A majority of participants disagreed with question 3 *"An easy to remember password is a safe password."* and agreed with question 10 *"A difficult to remember password is a more secure password."*. This is consistent for the shorter passwords but *not for the passphrase categories*, as participants also acknowledged that CP are more memorable than RP but ultimately rated CP and RP similarly in terms of Security. From this, we conclude that we need to go beyond PM as the *only* influencing factor and pay more attention to the factors that could make passphrases appear safer to users. As shown, it could be due to "randomness" in the way the phrase was formed. This randomness in turn can be due to the *syntax* (if the words strung together make grammatical sense) and *semantics* (if the words make sense when they are put together in any order) of the passphrase.

Finally, the our results regarding passphrases show the need for such metacognition based studies on th and informed that people regularly use passwords, forget them, and store them in places that are not secure, etc. even after being aware of safe password setting practices.

Future Work and Limitations

Continuing the above section, we plan to expand this study in a more (psycho-)linguistic direction. We see that passphrases are influenced by randomness in some domain, which is ***not** heterogeneity of characters* but is more related to how the units of the phrase function with each other. There have been a handful of studies linking passphrases to **semantics** like the one on semantic noise (Lee & Ewe, 2007) or "guided word choices" (Blanchard et al., 2018) and **syntax** like the one on entropy vs. syntax (Panferov, 2016). Even fewer study about the cognitive aspects like the one done on augmented cognition and cognitive load (Loos et al., 2019). However, there is no current study on the association between these linguistic aspects of passphrases with metacognition/perceived memorability/perceived security. We plan to improve the work in this study and find out if such associations exist and influence password choice. Furthermore, studies have been conducted to discover other factors for passphrase utility, such as *pronounceability* (White et al., 2014) and whether *multilingual* passphrases can be strong as well (Maoneke et al., 2020). We intend to keep these options open to including in our next metacognition experiments as mentioned above.

We also aim to find solutions to possible limitations. One of them was that the survey was conducted with a majority of users who have learned English as a second language, however comfortable they might have been. As the majority of the population did not consider passphrases to be useful, this study was possible. However, a study that focuses exclusively on passphrases should be careful with this problem. There is

also the concern that the phrases used in the experiment could have a bias for some participants as they might have heard it before/used frequently in some scenario, thus perceiving it as more memorable vs. some user perceiving a common phrase as less memorable because the variant of English they use in their regions and societies might use a synonym for the same. Thus future work can include more randomness and a pilot to be sure that the phrases themselves are not biased to a subset of participants and a check could be done to see if the participants have similar linguistic and sociolinguistic backgrounds.

Finally, this experiment was based on Judgement as a major task to determine the correlation. A Generation task can lead to different results. Taking into account strong concerns about privacy and limiting the user through nudges (Renaud & Zimmermann, 2018) there is the possibility of creating a completely different experimental framework.

References

Blanchard, Nikola K. and Malaingre, Clément and Selker, Ted. (2018). Improving security and usability of passphrases with guided word choice. *Proceedings of the 34th Annual Computer Security Applications Conference*. doi: 10.1145/3274694.3274734

EnglishSpeak. (n.d.). *English Phrases.* Retrieved from www.englishspeak.com/en/english-phrases

Hearn, Mike and Wheeler, Dan. (n.d.). *Use a Passphrase.* Retrieved from www.useapassphrase.com/

Kaspersky. (2019). *Kaspersky: Secure Password Check.* Retrieved from www.password.kaspersky.com/

Kävrestad, Joakim and Lennartsson, Markus and Birath, Marcus and Nohlberg, Marcus. (2020). Constructing secure and memorable passwords. *Information & Computer Security*, *ahead-of-print*. doi: 10.1108/ics-07-2019-0077

Lee, Kok-Wah and Ewe, Hong-Tat. (2007). Passphrase with Semantic Noises and a Proof on Its Higher Information Rate. *2007 International Conference on Computational Intelligence and Security Workshops (CISW 2007)*. doi: 10.1109/cisw.2007.4425580

Loos, Lila A. and Ogawa, Michael-Brian and Crosby, Martha E. (2019). Impedances of Memorable Passphrase Design on Augmented Cognition. *Augmented Cognition*, 84-92. doi: 10.1007/978-3-030-22419-6_7

Loutfi, Ijlal and Jøsang, Audun. (2015). Passwords are not always stronger on the other side of the fence. *Workshop on Usable Security*. Retrieved 2021-02-01, from www.ndss-symposium.org/wp-content/uploads/2017/09/02_2_3.pdf

Luna, Karlos. (2019). If it is easy to remember, then it is not secure: Metacognitive beliefs affect password selection. *Applied Cognitive Psychology*, *33*, 744-758. doi: 10.1002/acp.3516

Maoneke, Pardon Blessings and Flowerday, Stephen and Isabirye, Naomi. (2020). Evaluating the strength of a multilingual passphrase policy. *Computers & Security*, *92*, 101746. doi: 10.1016/j.cose.2020.101746

My1Login. (2019). *Password Strength Test - My1Login.* Retrieved from www.my1login.com/resources/password-strength-test/

Nahum, Mor and Vleet, Thomas M. Van and Sohal, Vikaas S. and Mirzabekov, Julie J. and Rao, Vikram R. and Wallace, Deanna L. and Lee, Morgan B. and Dawes, Heather and Stark-Inbar, Alit and Jordan, Joshua Thomas and Biagianti, Bruno and Merzenich, Michael and Chang, Edward F. (2017). Immediate Mood Scaler: Tracking Symptoms of Depression and Anxiety Using a Novel Mobile Mood Scale. *JMIR mHealth and uHealth*, *5*, e44. Retrieved 2020-10-23, from www.mhealth.jmir.org/2017/4/e44/ doi: 10.2196/mhealth.6544

Nelson, Thomas O. (1990). Metamemory: A Theoretical Framework and New Findings. *Psychology of Learning and Motivation*, 125-173. Retrieved 2019-07-29, from www.sciencedirect.com/science/article/pii/S0079742108600535 doi: 10.1016/s0079-7421(08)60053-5

Nordpass. (2020). *Most common passwords of 2020.* Retrieved from www.nordpass.com/most-common-passwords-list/

Panferov, Eugene. (2016). An Observation About Passphrases: Syntax vs Entropy. *arXiv:1603.06133 [cs]*. Retrieved 2021-02-01, from www.arxiv.org/abs/1603.06133

Rastle, Kathleen and Harrington, Jonathan and Coltheart, Max. (2002). 358,534 nonwords: The ARC Nonword Database. *The Quarterly Journal of Experimental Psychology Section A*, *55*, 1339-1362. doi: 10.1080/02724980244000099

Renaud, Karen and Zimmerman, Verena. (2018). Nudging folks towards stronger password choices: Providing certainty is the key. *Behavioural Public Policy*, 1-31. doi: 10.1017/bpp.2018.3

Stoet, Gijsbert. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*, 1096-1104. doi: 10.3758/brm.42.4.1096

Stoet, Gijsbert. (2016). PsyToolkit. *Teaching of Psychology*, *44*, 24-31. doi: 10.1177/0098628316677643

White, Andrew M. and Shaw, Katherine and Monrose, Fabian and Moreton, Elliott. (2014). Isn't that Fantabulous. *Proceedings of the 2014 workshop on New Security Paradigms Workshop - NSPW '14*. doi: 10.1145/2683467.2683470

Wilson, Michael. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*, 6-10. doi: 10.3758/bf03202594

Woods, Naomi and Siponen, Mikko. (2019). Improving password memorability, while not inconveniencing the user. *International Journal of Human-Computer Studies*, *128*, 61-71. Retrieved 2019-11-10, from www.sciencedirect.com/science/article/pii/S1071581919300102 doi: 10.1016/j.ijhcs.2019.02.003