Think Harder! Investigating the Effect of Password Strength on Cognitive Load during Password Creation

Yasmeen Abdrabou
yasmeen.essam@unibw.de
Bundeswehr University Munich
Germany

Yomna Abdelrahman
yomna.abdelrahman@unibw.de
Bundeswehr University Munich
Germany

Mohamed Khamis
mohamed.khamis@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Florian Alt
florian.alt@unibw.de
Bundeswehr University Munich
Germany

ABSTRACT

Strict password policies can frustrate users, reduce their productivity, and lead them to write their passwords down. This paper investigates the relation between password creation and cognitive load inferred from eye pupil diameter. We use a wearable eye tracker to monitor the user's pupil size while creating passwords with different strengths. To assess how creating passwords of different strength (namely weak and strong) influences users' cognitive load, we conducted a lab study ($N = 15$). We asked the participants to create and enter 6 weak and 6 strong passwords. The results showed that passwords with different strengths affect the pupil diameter, thereby giving an indication of the user's cognitive state. Our initial investigation shows the potential for new applications in the field of cognition-aware user interfaces. For example, future systems can use our results to determine whether the user created a strong password based on their gaze behavior, without the need to reveal the characteristics of the password.

CCS CONCEPTS

• Human-centered computing → Human computer interaction (HCI); **• Security and privacy → Human and societal aspects of security and privacy**.

KEYWORDS

Eye Tracking, Cognitive Load, Pupillometry, Cognition-Aware User Interfaces, Passwords Strength

ACM Reference Format:
Yasmeen Abdrabou, Yomna Abdelrahman, Mohamed Khamis, and Florian Alt. 2021. Think Harder! Investigating the Effect of Password Strength on Cognitive Load during Password Creation. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3411763.3451636

CHI '21 Extended Abstracts, May 8–13, 2021, Yokohama, Japan

1 INTRODUCTION

Passwords are the most popular authentication mechanism [25]. Ideally, a good password strikes a balance between being easy to remember and hard to guess [25]. Weak passwords might lead to unauthorized access to an organization's information assets. Thus, many organizations enforce frequent password changes to address passwords leakage [5]. At the same time, research showed that strict password policies decrease employees' productivity [27] and can even result in less security as employees work around rules to easily remember their passwords [40].

Password meters are used in many interfaces to help users create strong and secure passwords [40]. Ur et al. [38] found that participants had misconceptions about the impact of basing passwords on common phrases and including digits and keyboard patterns in their passwords. However, they also found that in most cases, users' perceptions of what characteristics make a strong secure password were consistent with password meter tools. The fact that users' perceptions of what characteristics make a strong password are accurate, motivated us to explore whether systems can learn about the strength of created passwords through the users rather than by examining the passwords themselves. Doing so has a security advantage: no third party applications would need to examine the created password to evaluate its strength. It also has a usability advantage: if we are able to determine password strength through the user's cognitive load (e.g., as estimated via an eye tracker), then users can consciously learn about their password's strength, even if the used interface does not measure the password's strength.

In this work, we contribute an investigation of the relationship between perceived password strength and cognitive load and how it affects the pupil diameter. We use a wearable eye tracker to monitor users' pupil size while creating passwords with different strengths. We found that the pupil dilates while creating strong passwords and contracts while creating weak passwords. To the best of our knowledge, we are the first to investigate the relation between password strength and cognitive load. Unlike password strength meters that estimate the password strength based on the password characters, our work allows systems to determine the perceived strength of a password without revealing its characteristics. Our findings allow for new applications in the field of cognition-aware interfaces, for example, suggesting verbal, visual or spatial cues to help the user creating unique, memorable passwords [3].

2 RELATED WORK

Our work builds on prior research on utilizing eye tracking for cognitive load state estimation and password strength.

2.1 Pupillometry and Cognitive Load

Three types of cognitive load measures were introduced in literature: subjective, physiological and performance measures [28]. Subjective measures reflect the user's subjective assessment of cognitive load. The NASA-TLX questionnaire [14] is a frequently used assessment tool for subjective cognitive load. However, such a tool cannot account for rapid changes in the cognitive load that may be the result of changes in the experiment. Physiological measures include pupil dilation, heart-rate variability, and galvanic skin response [6, 17, 19]. Changes in these measures have been shown to correlate with different levels of cognitive load [15, 41]. However, physiological measures depend on many factors, including other aspects of the user's cognitive state such as anxiety [7], arousal [21], the user's physical activity [33], and environmental variables such as light [32]. Hence, researchers should draw attention to the study conditions and user's state. Finally, performance measures captures how efficiently is the user performing a given task. The method is based on the standardization of raw scores for mental effort and task performance to z scores, which are displayed in a cross of axes [29]. In our work, we use the second measure "physiologically" as it is captured without requiring participants to reflect on their performance during password creation nor fill a questionnaire.

In the last decades, researchers have investigated the pupillary response for different types of tasks [8, 9, 16, 23]. Pupil dilation was found to be higher for more challenging tasks [11, 26]. Not only task demands have been found to influence the pupil diameter, but also factors like anxiety [7], stress [10], and fatigue [37]. A study done by Just and Carpenter [20], showcased that pupil responses can be an indicator of the effort to understand and process information. They conducted an experiment where participants were given two sentences of different complexities to read while they would measure their pupil diameters. They found that the pupillary dilation was larger while readers processed the sentence that was complicated and more subtle while reading the simpler one. It was also shown that pupil size correlates to the difficulty of a cognitive task [15]. Over the years, researchers have encountered some challenges in pupillometry such as luminance. One way to improve validity is to strictly control the luminance of the experimental stimuli, but this limits the potential of pupillometry. While cognitive load can be affected by a large number of factors, pupillometry offers a responsive signal that can potentially provide approximate real-time feedback of the users' arousal and potentially their cognitive load.

We expect that creating stronger passwords is more difficult and thus cognitively demanding. This motivated us to study the relation between cognitive load and password creation.

2.2 Password Strength

Passwords are the most popular authentication mechanism [25]. There are different types of attacks that passwords might be vulnerable to e.g., brute force and guessing attacks [31]. Hence, system administrators started employing password-composition policies to eliminate attacks [13, 39]. To help users create strong passwords, password meters are integrated to interfaces to give users an estimate of how strong their passwords are and hence, how easy it is to be cracked [13]. Researchers found that password meters design, color and feedback messages have an influence on the strength of the created passwords [12, 13, 34, 39]. Although prior work has shown that password-composition policies requiring more characters or more character classes can improve resistance to automated guessing attacks, many passwords that meet common policies remain vulnerable [22, 42]. Furthermore, strict policies can frustrate users, reduce their productivity, and lead users to write their passwords down [1, 18, 35].

Ur et al. [38] found that users are aware of what makes a password strong. This suggests that putting more effort in creating a password might be an indication that it is a strong one. This motivated us to study the relation between password strength and cognitive load during password creation. If such a connection exists, future systems can then determine the strength of a password based on the user's cognitive load, alleviating the need for systems to access the password characteristics.

Hence, the need to study the relation between creating passwords and cognitive load is a must. Therefore, in this paper, we introduce using pupillometry to detect users' cognitive load while creating weak and strong passwords.

3 CONCEPT AND METHODOLOGY

In this section, we describe our concept and approach of evaluating cognitive load from pupil diameter. Since the relation between pupil diameter and cognitive load has already been proven (see subsection 2.1). In this work, we look at how the users' cognitive load changes during weak and strong passwords creation (**RQ**). Bafna et al. [4] showed that there is increase in cognitive load when participants were asked to memorize and type difficult vs easy sentences. Inspired by them, we hypothesize that creating strong passwords will induce higher cognitive load compared to creating weak passwords.

For this we ran a lab study to answer our research question. In the following, we highlight how we analyzed the collected data. First, we analyzed the collected passwords' strength against the zxcvbn password meter [43] to see if participants' rating matches the system rating. Second, we extracted the pupil diameter variance between weak and strong passwords and tested their statistical significance. Third, we calculated the mean pupil diameter change (MPDC) as a mean to calculate the cognitive load while creating passwords of different strengths.

3.1 Password Strength Meter

We analyzed and compared user rated password strength against the zxcvbn password strength meter [43] (details in Section 5.2). In addition, we statistically analyzed the rated weak and strong passwords strength using repeated measures ANOVA and the generated entropy for weak and strong passwords by the zxcvbn meter. Finally, we further analyzed the post-study questions and reported their results. We used a cut off score of 2.5 for differentiating between weak and strong passwords where from 1 to 2.5 is considered as weak password and from more than 2.5 to 5 is considered as strong password.

3.2 Mean Pupil Diameter Change Calculation

We analyze the average pupil diameter and the commonly used mean pupil diameter change (MPDC) as a cognitive load metric [2, 24]. The MPDC calculation can be found in Equation 1 where MPD_p represents mean pupil diameter for a specific password and MPD_a represents mean pupil diameter for the participants while entering all passwords and N is the number of overall passwords in our case it is 12. The overall mean is subtracted from the password mean in order to compare results between subjects with different pupil sizes [30]. The MPDC has the advantage compared to MPD as it corrects the fluctuations in the baseline pupil diameter, and compensates for any structural temporal trends that might exist. Hence, the use of MPDC is appropriate as compared to other types of measures such as dilation percentage, as pointed out by Beatty et al. [6], "the pupillary dilation evoked by cognitive processing is independent of baseline pupillary diameter over a wide range of baseline values". On the other hand, the MPDC allows us to determine whether the baseline itself differed as a function of the password strength.

$$MPDC = \sum_{i=0}^{N} \frac{MPD_p - MPD_a}{N} \tag{1}$$

4 EVALUATION

We conducted a user study in which we recorded the participants' eye gaze data while creating weak and strong passwords on laptops.

4.1 Study Design

We applied a repeated-measures design, where all participants did all conditions. Overall, participants were asked to create 12 passwords (6 weak and 6 strong). The order of which password they should enter was counterbalanced using a Latin Square. Participants were advised not to reuse a password they already entered. We collected the entered passwords, passwords ratings and gaze data including pupil size as dependent variables. Passwords strength (weak vs strong) acted as an independent variable and the screen brightness, as well as the room light, was kept the same throughout the whole experiment.

4.2 Participants and Apparatus

We invited 15 participants (5 males), recruited via a University mailing list, to our lab. The age varied from 22 to 31 ($Mean = 24.27$; $SD = 2.91$). Participants had different backgrounds (CS, Engineering, Landscape Design), and different nationalities (Spain, China, Bangladesh, Pakistan, Egypt, Germany). Participants had basic to average experience with eye-tracking. Nobody wore glasses.

As shown in Figure 1, our experimental setup consisted of a Tobii Pro Glasses 2[1] with 120 fps running on Lenovo T440s[2] along with the Tobii glasses controller[3]. We implemented a simple web page interface where it shows the question and an empty field to write the password in.

[1]Tobii Pro Glasseshttps://www.tobiipro.com/product-listing/tobii-pro-glasses-2/
[2]Lenovo T440shttps://www.lenovo.com/gb/en/laptops/thinkpad/t-series/t440s/
[3]Tobii Glasses Controllerhttps://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/setup/installing-tobii-glasses-controller/

4.3 Procedure

After arriving in the lab, participants were asked to sign a consent form and received an explanation of the purpose of the study. After that, we calibrated the eye tracker using Tobii's one-point calibration[4]. We instructed the participants to change the keyboard style to the one they are using and to change the language as well if needed. We gave the participants the device and we asked them to create and enter a set of passwords (6 weak and 6 strong) one at a time in a randomized order. Participants were requested to enter passwords more than 8 characters but we did not give any hints on how to create strong password neither requested any requirements. After each password, we asked the participants to rate the password strength on a Likert-scale from 1 to 5 (very weak to very strong). At the end of the study, we asked the participants "What makes a strong password?" to understand whether they know the basic password policies. Overall the study lasted approximately 10 minutes and participants were rewarded with 5 EUR.

5 RESULTS

5.1 Data Cleaning and Reprocessing

In order to start analyzing the collected pupil size, we first removed the missing data. Then, we averaged both left and right eye pupil size to one value. After that, we plotted the data to check for outliers. The data of two participants were considered outliers due to excessive talking and asking questions during the study which highly affects the cognitive load [36]. Therefore, the following analysis is done only on 13 participants.

5.2 Rated Password Strength

To understand how participants perceived their passwords' strength, we compared their rated password strength to the zxcvbn password strength meter. Figure 2 shows the average rating for all the passwords entered per participant against the results from the zxcvbn meter. There is a variance between the passwords ratings. However, the difference is not statistically significant ($\chi^2(1) = 3.769$, $P = .0521$) as found by Friedman test. We also compared the entropy of the weak and strong passwords calculated by the zxcvbn meter and we found a significant difference between the entropy for the weak ($M = 14.45$; $SD = 3.59$) and the strong passwords ($M = 60.75$; $SD = 9.21$), ($F_{1,14} = 268.760$, $P < .001$) which assures that the entered passwords are valid to be used for further analysis [13] and that participants' perception of weak and strong passwords matches the password meter rating.

5.3 Post Study Question Analysis

At the end of the study, we asked the participants what makes a strong password. They mentioned special characters (22%), adding numbers (18%), upper/lower case letters (18%), increasing the length (14%), adding numbers (14%) and adding random characters (14%). While metrics like password length have a stronger positive impact on security than special characters [25], the responses still show that participants knew what makes passwords stronger.

[4]One Point Calibration: https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/run/running-a-monocular-calibration-with-the-Tobii-pro-spectrum/

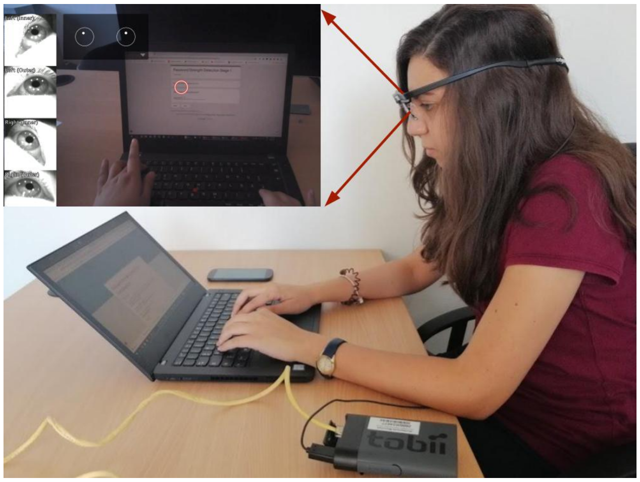


Figure 1: Experiment study setup consisting of a laptop and a wearable eye tracker Top Left: gaze monitoring while creating passwords viewed from Tobii pro glasses controller.

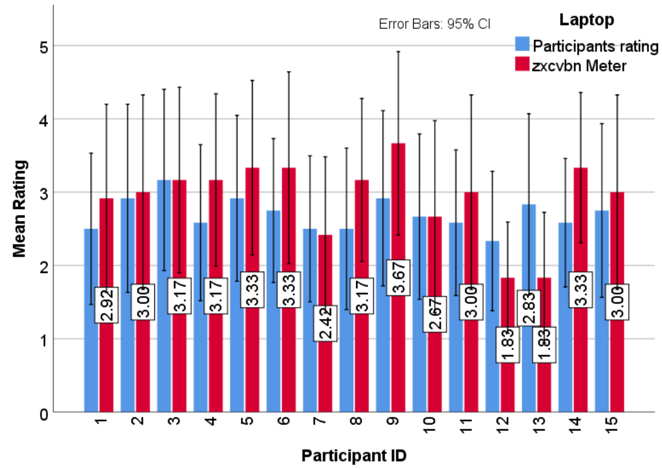
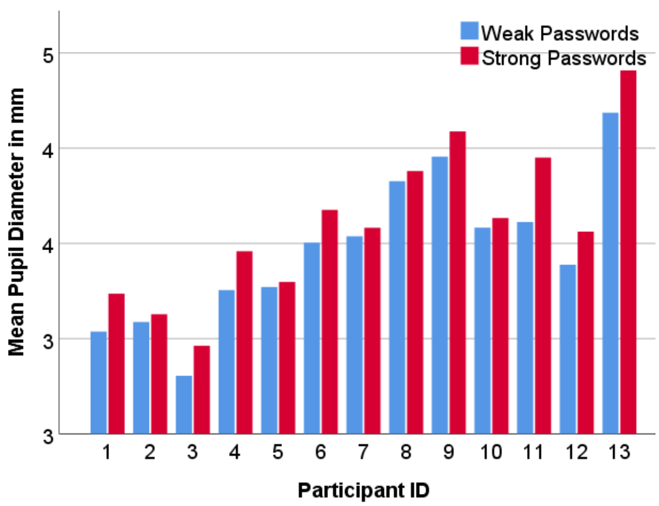


Figure 2: Password strength comparison between participants' rating and the zxcvbn password meter rating. Showing similar ratings between the zxcvbn meter and users ratings



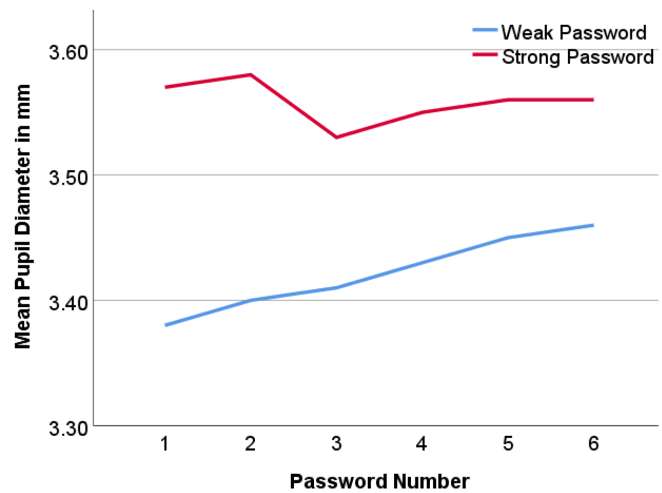


Figure 3: (Left) shows the MPD across the 13 participants. (Right) shows the MPD per created password

Table 1: MPD difference between creating strong and weak passwords for all participants

Pupil Diameter/ Participant ID	1	2	3	4	5	6	7	8	9	10	11	12	13
Strong Passwords	3.24	3.13	2.96	3.46	3.3	3.68	3.58	3.88	4.09	3.63	3.95	3.56	4.41
Weak Passwords	3.04	3.09	2.8	3.25	3.27	3.5	3.54	3.83	3.96	3.58	3.61	3.39	4.19
Difference	**0.2**	**0.04**	**0.16**	**0.2**	**0.03**	**0.17**	**0.04**	**0.05**	**0.13**	**0.05**	**0.34**	**0.17**	**0.22**

5.4 Pupil Diameter and Password Strength

Figure 3 left, shows the MPD across the 13 participants. As seen in the figure, the MPD dilates when creating strong passwords than weak passwords expect for participant 7 and 11. Repeated measures ANOVA showed statistical significant difference between the MPD for weak ($M = 3.47$, $SD = .4$) and strong passwords ($M = 3.60$, $SD = .41$), ($F_{1,12} = 29.497$, $P < .001$). This means that the password strength has a statistically significant effect on the MPD. Furthermore, We also looked into the MPD difference while creating strong and weak passwords for all participants(see Table

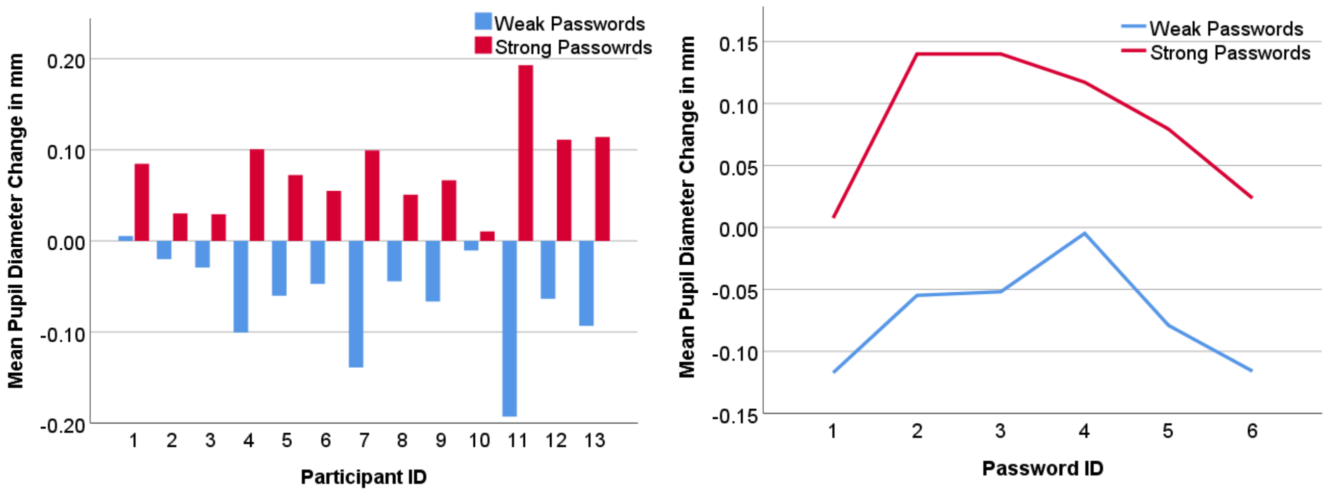


Figure 4: (Left) shows the MPDC across the 13 participants. (Right) shows the MPDC per created password

1) and we found that the mean difference is (M = .14, SD = .09) and the smallest difference is $M = 0.03mm$. Which means that even when we cannot draw a threshold due to different pupil size response across participants, the difference still exists indicating that strong passwords induce higher cognitive load.

Looking at the MPD per created password, we can see in Figure 3 right, that for all 6 passwords participants had wider pupil diameter which can indicate higher cognitive load while creating strong passwords than weak passwords. That was also highlighted by repeated-measures ANOVA where it showed a statistically significant effect of the password strength, weak (M = 3.42, SD = .03) and strong (M = 3.56, SD = .01) on the MPD throughout all repetitions ($F_{1,5}$ = 76.407, P < .001).

Since we did not have a baseline and each user has a different pupil size, we used the MPDC as another metric for reflecting on the cognitive load. The MPDC has the advantage of compensating for any structural temporal trends that might exist during the user task. Hence, the use of MPDC will give more insights into our case. Figure 3 left, shows the MPDC across all participants. The figure highlights the change rate of the mean pupil diameter while creating weak and strong passwords. From the figure, we can see that in most cases creating strong passwords leads to pupil dilation while creating weak passwords leads to contracting the pupil or far less dilation than when creating strong passwords. This was also highlighted statistically when using repeated measures ANOVA where it showed statistically significant difference between the MPDC while creating weak (M = −.07, SD = .05) and strong (M = .07, SD = .05) passwords, ($F_{1,12}$ = 28.245, P < .001).

Figure 4 (right) shows the change in MPDC across repetitions. The figure also shows that the trend of dilating the pupil while creating strong passwords and contracting the pupil while creating weak passwords is still valid across repetitions. This was also statistically highlighted as Repeated measures ANOVA showed statistically significant difference between the MPDC while creating weak (M = −.06, SD = .04) and strong (M = .08, SD = .06) passwords across repetitions, ($F_{1,5}$ = 139.283, P < .001).

6 DISCUSSION

Our results suggest that there is difference in the MPD when creating strong vs weak passwords. Even when we could not draw a threshold due to different pupil size response across participants, we found that the difference in pupil size still exists indicating that strong passwords induce higher cognitive load. For MPDC We noticed that after the third strong password, the pupil diameter started decreasing (see Figure 4 right). This might be due to participants finding a password strategy after their third trial and hence the cognitive load started decreasing. This answers our **RQ** where it is clear now with using different pupil diameter evaluation metrics across different repetitions, that creating stronger passwords leads to pupil dilation that is a sign of higher cognitive load than when creating weak passwords.

Our findings can be used to optimize user's workload for better productivity. It can be used to suggest alternative passwords to the user based on their pupil diameter. In addition, it can also be used to suggest verbal, visual or spatial cues to help the user creating unique memorable passwords [3]. Since we found that password strength is reflected in pupil diameter response, pupil diameter can be integrated in interfaces to assess password strength without revealing the actual password to the system.

7 LIMITATIONS AND FUTURE WORK

We acknowledge that we had a controlled setup, where the brightness of the surroundings was kept constant. However, more sophisticated approaches (e.g., using machine learning) could be used to consider the influence of the surrounding brightness change. For future work, it is valuable to investigate the effect of reusing passwords and whether it complies to our findings or not. In addition, we shall integrate pupil diameter as a password strength check policy and study gaze behavior as a metric to judge password strength. We will also investigate how would our approach distinguish between a low cognitive load due to a weak password and a low cognitive load due to the user adopting a password strategy.

8 CONCLUSION

In this work, we described our approach to infer users' cognitive load based on pupil diameter while creating passwords with different strengths. We hypothesized that creating strong and weak passwords will lead to change in pupil diameter reflecting the change of cognitive load. We found that creating passwords with different strength leads to changes in pupil diameter, hence, change in cognitive load. We found that creating strong passwords leads to pupil dilation while creating weak passwords leads to pupil contraction. This means that creating strong passwords induces more cognitive load than creating weak passwords. We believe that our findings will be a great addition to cognitive aware systems to better optimize user's productivity and performance. By presenting this work at CHI we hope to stimulate a discussion on which systems and contexts could benefit from our approach and how.

ACKNOWLEDGMENTS

This work was supported by the Royal Society of Edinburgh (RSE award number 65040), EPSRC New Investigator Award (EP/V008870/1), DFG under grant agreement no. 316457582 and 425869382 as well as by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr (Voice of Wisdom). Finally, the Studienstiftung des deutschen Volkes ("German Academic Scholarship Foundation").

REFERENCES

[1] Anne Adams, Martina Angela Sasse, and Peter Lunt. 1997. Making passwords secure and usable. In *People and Computers XII*. Springer, 1–19.

[2] Sylvia Kiosterud Ahern. 1979. Activation and intelligence: Pupillometric correlates of individual differences in cognitive abilities. (1979).

[3] Mahdi Nasrullah Al-Ameen, Matthew Wright, and Shannon Scielzo. 2015. Towards Making Random Passwords Memorable: Leveraging Users' Cognitive Ability Through Multiple Cues. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2315–2324. https://doi.org/10.1145/2702123.2702241

[4] Tanya Bafna, John Paulin Paulin Hansen, and Per Baekgaard. 2020. Cognitive Load during Eye-Typing. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 23, 8 pages. https://doi.org/10.1145/3379155.3391333

[5] Richard Baskerville and Mikko Siponen. 2002. An information security metapolicy for emergent organizations. *Logistics Information Management* (2002).

[6] J Beatty and B Lucero-Wagoner. 2000. The pupillary system In T. Cacioppo, L. Tassinary & G. Berntson (Eds.), Handbook of Psychophsyiology (pp. 142-162).

[7] I Chen, Chi-Cheng Chang, et al. 2009. Cognitive load theory: An empirical study of anxiety and task performance in language learning. (2009).

[8] Siyuan Chen and Julien Epps. 2014. Using Task-Induced Pupil Diameter and Blink Rate to Infer Cognitive Load. *Human–Computer Interaction* 29, 4 (2014), 390–413. https://doi.org/10.1080/07370024.2014.892428 arXiv:https://doi.org/10.1080/07370024.2014.892428

[9] Siyuan Chen, Julien Epps, and Fang Chen. 2013. Automatic and Continuous User Task Analysis via Eye Activity. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces* (Santa Monica, California, USA) *(IUI '13)*. Association for Computing Machinery, New York, NY, USA, 57–66. https://doi.org/10.1145/2449396.2449406

[10] Dan Conway, Ian Dick, Zhidong Li, Yang Wang, and Fang Chen. 2013. The Effect of Stress on Cognitive Load Measurement. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 659–666.

[11] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The Index of Pupillary Activity: Measuring Cognitive Load <i>Vis-à-Vis</i> Task Difficulty with Pupil Oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173856

[12] David Eargle, John Godfrey, Hsin Miao, Scott Stevenson, Richard Shay, Blase Ur, and Lorrie Cranor. 2015. You can do better—motivational statements in password-meter feedback. *Proc. SOUPS Posters* (2015).

[13] Serge Egelman, Andreas Sotirakopoulos, Ildar Muslukhov, Konstantin Beznosov, and Cormac Herley. 2013. Does My Password Go up to Eleven? The Impact of Password Meters on Password Selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2379–2388. https://doi.org/10.1145/2470654.2481329

[14] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[15] Eckhard H Hess and James M Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 3611 (1964), 1190–1192.

[16] Sazzad Hussain, Siyuan Chen, Rafael A Calvo, and Fang Chen. 2011. Classification of cognitive load from task performance & multichannel physiology during affective changes. In *Conference on Multimodal Interaction*. 1–4.

[17] Curtis S. Ikehara and M. Crosby. 2005. Assessing Cognitive Load with Physiological Sensors. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (2005), 295a–295a.

[18] Philip G. Inglesant and M. Angela Sasse. 2010. The True Cost of Unusable Password Policies: Password Use in the Wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 383–392. https://doi.org/10.1145/1753326.1753384

[19] Shamsi T. Iqbal, Piotr D. Adamczyk, Xianjun Sam Zheng, and Brian P. Bailey. 2005. Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 311–320. https://doi.org/10.1145/1054972.1055016

[20] Marcel A Just and Patricia A Carpenter. 1993. The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 47, 2 (1993), 310.

[21] Khaled Kassem, Jailan Salah, Yasmeen Abdrabou, Mahesty Morsy, Reem El-Gendy, Yomna Abdelrahman, and Slim Abdennadher. 2017. DiVA: Exploring the Usage of Pupil <u>Di</u>Ameter to Elicit <u>V</u>Alence and <u>A</u>Rousal. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (Stuttgart, Germany) *(MUM '17)*. Association for Computing Machinery, New York, NY, USA, 273–278. https://doi.org/10.1145/3152832.3152836

[22] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez. 2012. Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms. In *2012 IEEE Symposium on Security and Privacy*. 523–537. https://doi.org/10.1109/SP.2012.38

[23] Peter Kiefer, Ioannis Giannopoulos, Andrew Duchowski, and Martin Raubal. 2016. Measuring Cognitive Load for Map Tasks Through Pupil Diameter. In *Geographic Information Science*, Jennifer A. Miller, David O'Sullivan, and Nancy Wiegand (Eds.). Springer International Publishing, Cham, 323–337.

[24] Jeff Klingner. 2010. *Measuring Cognitive Load During Visual Tasks by Combining Pupillometry and Eye Tracking*. Ph.D. Dissertation. Stanford University, Department of Computer Science. http://purl.stanford.edu/mv271zd7591

[25] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. 2011. Of Passwords and People: Measuring the Effect of Password-Composition Policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2595–2604. https://doi.org/10.1145/1978942.1979321

[26] Thomas Kosch, Mariam Hassib, Daniel Buschek, and Albrecht Schmidt. 2018. Look into My Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188643

[27] Stephen Mujeye and Yair Levy. 2013. Complex passwords: How far is too far? The role of cognitive load on employee productivity. *Online Journal of Applied Knowledge Management (OJAKM)* 1, 1 (2013), 122–132.

[28] R. O'donnell and F. T. Eggemeier. 1986. Workload assessment methodology.

[29] Fred GWC Paas and Jeroen JG Van Merriënboer. 1993. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors* 35, 4 (1993), 737–743.

[30] Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 141–144. https://doi.org/10.1145/1743666.1743701

[31] Mudassar Raza, Muhammad Iqbal, Muhammad Sharif, and Waqas Haider. 2012. A survey of password attacks and comparative analysis on methods for secure authentication. *World Applied Sciences Journal* 19, 4 (2012), 439–444.

[32] Prentice Reeves. 1920. The response of the average pupil to various intensities of light. *JOSA* 4, 2 (1920), 35–43.

[33] Ramón Romance, Adriana Nielsen-Rodríguez, Javier Benítez-Porres, José Luis Chinchilla-Minguet, and Honorato Morente-Oria. 2018. Cognitive Effects and educational possibilities of physical activity in sustainable cities. *Sustainability* 10, 7 (2018), 2420.

[34] Richard Shay, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Alain Forget, Saranga Komanduri, Michelle L. Mazurek, William Melicher, Sean M. Segreti, and Blase Ur. 2015. A Spoonful of Sugar? The Impact of Guidance and Feedback on Password-Creation Behavior. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2903–2912. https://doi.org/10.1145/2702123.2702586

[35] Jeffrey M. Stanton, Kathryn R. Stam, Paul Mastrangelo, and Jeffrey Jolton. 2005. Analysis of End User Security Behaviors. *Comput. Secur.* 24, 2 (March 2005), 124–133. https://doi.org/10.1016/j.cose.2004.07.001

[36] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76.

[37] Masaaki Tanaka, Akira Ishii, and Yasuyoshi Watanabe. 2015. Effects of mental fatigue on brain activity and cognitive performance: a magnetoencephalography study. *Anat Physiol* 4 (2015), 1–5.

[38] Blase Ur, Jonathan Bees, Sean M. Segreti, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. Do Users' Perceptions of Password Security Match Reality?. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3748–3760. https://doi.org/10.1145/2858036.2858546

[39] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L. Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2012. How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation. In *21st USENIX Security Symposium (USENIX Security 12)*. USENIX Association, Bellevue, WA, 65–80. https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/ur

[40] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M. Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2015. "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 123–140. https://www.usenix.org/conference/soups2015/proceedings/presentation/ur

[41] Pauline van der Wel and Henk van Steenbergen. 2018. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic bulletin & review* 25, 6 (2018), 2005–2015.

[42] Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. 2010. Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords. In *Proceedings of the 17th ACM Conference on Computer and Communications Security* (Chicago, Illinois, USA) *(CCS '10)*. Association for Computing Machinery, New York, NY, USA, 162–175. https://doi.org/10.1145/1866307.1866327

[43] Daniel Lowe Wheeler. 2016. zxcvbn: Low-budget password strength estimation. In *25th {USENIX} Security Symposium ({USENIX} Security 16)*. 157–173.