LEVERAGING EYE GAZE TO ENHANCE SECURITY MECHANISMS

YASMEEN ESSAM ABDRABOU MAHMOUD

Vollständiger Abdruck der von der Fakultät für Informatik der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation

Gutachter: Prof. Dr. Florian Alt
Gutachter: Prof. Dr. Mohamed Khamis

Die Dissertation wurde am 20.12.2022 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Informatik am 22.03.2023 angenommen. Die mündliche Prüfung fand am 23.03.2023 statt.

ABSTRACT

After more than six decades, passwords remain a ubiquitous approach for authentication. The main reason for this is that passwords currently provide a balance between usability, security, and administrability, meaning that no other mechanism offers an equally good trade-off between the effort required for implementation, ease of administration (e.g., reset/changing credentials), ease of use, and security. However, password memorability is nearly impossible due to the large number of accounts each user has. As a result, individuals are more likely to create weak, easily remembered passwords or reuse passwords. According to the literature, a single user has, on average, 80 accounts and over 3.5 passwords shared between them. Users expose themselves to guessing and brute-force assaults when they create weak passwords or a single point of failure if they reused passwords. On the other side, advances in computer vision made eye tracking ubiquitous. According to the literature, users' eye gaze movements can reveal their gender, age, ethnic group, sexual orientation, mental disease, physical illness, and more. Accordingly, in this thesis, we introduce eye gaze behavior to enhance security mechanisms, focusing on knowledge-based passwords as a use case. We first start by understanding users' gaze behavior during authentication and study the relation between password creation and cognitive load. Then we look at how this behavior can be modeled using different machine learning classifiers. After that, we provide a framework for employing gaze behavior in security systems. Finally, we discuss how gaze behavior can be used beyond authentication and reflect on different ethical and user privacy aspects of exploiting eye gaze behavior in different security mechanisms.

ZUSAMMENFASSUNG

Nach mehr als sechs Jahrzehnten sind Passwörter immer noch ein allgegenwärtiger Ansatz für die Authentifizierung. Der Hauptgrund dafür ist, dass Passwörter derzeit ein Gleichgewicht zwischen Benutzerfreundlichkeit, Sicherheit und Verwaltbarkeit bieten, was bedeutet, dass kein anderer Mechanismus einen gleich guten Kompromiss zwischen dem Aufwand für die Implementierung, der einfachen Verwaltung (z. B. Zurücksetzen/Ändern von Anmeldedaten), der Benutzerfreundlichkeit und der Sicherheit bietet. Allerdings ist es fast unmöglich, sich ein Passwort zu merken, da jeder Benutzer eine große Anzahl von Konten hat. Daher ist es wahrscheinlicher, dass Einzelpersonen schwache, leicht zu merkende Passwörter erstellen oder Passwörter wiederverwenden. Laut Literatur hat ein einzelner Benutzer im Durchschnitt 80 Konten wobei Passwörter für mehr als 3,5 Konten wiederverwendet werden. Die Benutzer setzen sich selbst dem Erraten und Brute-Force-Angriffen aus, wenn sie schwache Passwörter erstellen, oder einem Single Point of Failure, wenn sie Passwörter wiederverwenden. Auf der anderen Seite haben die Fortschritte in der Computer-Vision dazu geführt, dass die Blickverfolgung allgegenwärtig ist. Der Literatur zufolge können die Augenbewegungen der Benutzer Aufschluss über ihr Geschlecht, ihr Alter, ihre ethnische Zugehörigkeit, ihre sexuelle Orientierung, psychische und physische Krankheiten und vieles mehr geben. Dementsprechend stellen wir in dieser Arbeit die Nutzung des Blickverhaltens zur Verbesserung von Sicherheitsmechanismen vor, wobei wir uns auf wissensbasierte Passwörter als Anwendungsfall konzentrieren. Wir beginnen damit, das Blickverhalten der Benutzer während der Authentifizierung zu verstehen und untersuchen den Zusammenhang zwischen der Erstellung von Passwörtern und der kognitiven Belastung. Dann untersuchen wir, wie dieses Verhalten mit Hilfe verschiedener Klassifikatoren des maschinellen Lernens modelliert werden kann. Danach stellen wir ein Framework für den Einsatz von Blickverhalten in Sicherheitssystemen vor. Abschließend diskutieren wir, wie das Blickverhalten über die Authentifizierung hinaus genutzt werden kann und reflektieren verschiedene ethische und datenschutzrechtliche Aspekte der Nutzung des Blickverhaltens in verschiedenen Sicherheitsmechanismen.

PREFACE

This thesis is the result of the research I carried out at Universität der Bundeswehr München. The decisions taken in this thesis were influenced by several conversations and discussions with researchers and practitioners at conferences, workshops, and lab visits. This thesis was created in close collaboration with my second supervisor from the University of Glasgow who provided expert knowledge from their field.

Throughout this my work on this thesis, I supervised undergraduate student projects, Bachelor and Master theses that help to realize my ideas, prototypes, and evaluations. Many of these collaborations and theses resulted in publications that are a core part of this work. To emphasize these collaborations, I chose to write this thesis using the scientific plural ("we"). When applicable, all contributing authors of the resulting publication are cited at the beginning of each chapter, including the publication's reference.

STATEMENT OF COLLABORATION

In this thesis, I present research that has been conducted between 2019 and 2022. This thesis includes research made in conjunction with my professors, colleagues, and supervised students. To acknowledge these contributions, I utilize the scientific plural throughout this dissertation. Original work written exclusively for this thesis can be found in Chapters 1, 2, 8, 9, and 10. Chapters 3, 4, 5, 6, and 7 are based on co-authored peer-reviewed papers that were presented at international conferences. This statement elaborates these collaborations in detail.

Chapter 3 - Related Work: Gaze in Security and Privacy Applications This chapter is based on a CHI 2020 publication [181]. The idea of the paper was originated by my supervisor Mohamed Khamis and iterated collaboratively. The literature papers search was mainly done by me and then papers were divided equally among all coauthors to write their respective sections. The paper was presented as a video recording by me and my supervisor Mohamed Khamis and published in the ACM Digital library.

Chapter 4 - Gaze Vs. Touch for Knowledge-based Authentication This chapter is based on a publication in ETRA short papers 2021 [7]. I conducted the idea, implementation, and analysis of the study. As the study was between subjects, I conducted 20 participants and a student conducted the other 20 participants as part of his seminar tasks. I was the main author involved in writing up the paper and presenting it during the virtual conference, whereas my co-authors conducted several iterations of the paper, providing feedback. This paper was presented virtually by me.

Chapter 5 - Cognitive Load and Knowledge-based Authentication This chapter is based on a publication in CHI LBW 2021 [5]. The idea of this paper originated from me. I also did the implementation, study, and data analysis. I was the main author involved in writing up the paper and presenting it during the virtual conference, whereas my co-authors helped review the paper and provide feedback.

Chapter 6 - Detecting Password Strength from Gaze Behavior This chapter is based on a publication in ETRA 2021 Full Papers [10]. The idea of the paper was originated by me and further framed by my supervisors, Florian Alt and Mohamed Khamis. The implementation was done by me. The machine learning models were implemented by Ahmed Shams, Mohamed Mantawy, Anam Khan, and the statistical analysis conducted by me. I was the main author involved in writing up the paper and presenting it during the virtual conference, whereas my co-authors helped review the paper and provide feedback.

Chapter 7 - Detecting Password Reuse from Gaze Behavior This chapter is based on a CHI 2022 publication [9]. The idea for the paper originated from my supervisors and me. The implementation was done by Johannes Schütte as his bachelor thesis project, which I supervised with Ken Pfeuffer and Florian Alt.

Johannes conducted initial evaluations. However, further evaluations and statistical analysis were done by me. Ahmed Shams built the machine-learning models with Daniel Buschek's guidance and supervision. I was the main author involved in writing and presenting the paper during the conference. My supervisor Florian Alt and I wrote the paper, whereas our co-authors helped review the paper and provide feedback.

ACKNOWLEDGMENTS

In the past few years, I had the privilege of meeting and working with a number of outstanding people who guided, helped, and supported me throughout my journey. I take the opportunity in this section to point out all those great people, supervisors, colleagues, friends, and family who helped me in several ways, although it will not be enough.

After thanking Allah (SWT). It is with immense gratitude that I acknowledge with much appreciation the continuous help and support of my supervisor *Florian Alt*. Thank you for opening the door for me to conduct my Ph.D. in Munich, for your support throughout all my funding applications, scholarships and internship, for sharing your ideas and vision, for our weekly meetings where everything made sense afterward, and for inspiring me to pursue an academic career.

To my co-supervisor *Mohamed Khamis*, for putting me on the academic path, for introducing me to the usable security domain and eye tracking domains, for your support since early publications before my Ph.D., for believing in me, and for always listening.

To *Ken Pfeuffer*, for being an amazing co-author, and colleague, for teaching me a lot about eye tracking, and for your continuous support.

To *Yomna Abdelrahman* whom I was always waiting for her presentation every year in Cairo to get inspired by her HCI ideas. Thank you for showing me how fun HCI is and supporting my first-ever publications.

To *Radiah Rivu*, for being an amazing colleague, office mate, and Ph.D. partner, for all the yummy food you brought to me, for all the papers we did together early Ph.D. when we did not know which direction to take and for the inspiring office conversations.

To *Sarah Prange*, for the amazing art nights and crafts, we did to ease Ph.D. stress, for being a great travel partner, for the amazing beach holidays, and for all the small gestures you always do.

To *Mariam Hassib*, for your continuous support specifically during thesis writing, and for the brainstorming meetings in which life made more sense afterward.

To the entire USEC team, *Lukas Mecke*, *Michael Fröhlich*, *Felix Dietz*, *Ville Mäkelä*, *Simon von der Au*, *Oliver Hein*, *Pascal Knierem*, *Verena Distler*, and *Sarah Delgado*, for your support and laughter which made my PhD life easier and for the amazing kitchen conversations which helped me throughout different struggles. To *Heike Renner*, for always helping out with my paperwork, German translations, and my German class homework.

To *Ralf Biedert*, for your support throughout my internship and making it so fun, for teaching me everything I know about eye tracking in VR, and for our fun inspirational conversations.

To *Daniel Buschek*, for all our password-related conversations, for teaching me a lot about machine learning, and for being a great helpful co-author.

To *Anam Khan*, for teaching me how to work with raw gaze data and for teaching me about ML and eye tracking.

To *Ahmed Shams*, for responding to my ML additions shortly before the deadlines and for bearing with my changes and more figure requests.

To *Passant Elagroudy*, for listening to my nonsense, for hosting me in your place which is one of my comfort zones, for inspiring me to pursue a Ph.D., and for your stats notes since the HCI course that I still use until now. To *Alia Saad*, for our conversations, project ideas, and fun during CHI. To *Menna Essam*, for being a journey companion, for your continuous support, and for the fun during AfriCHI, SOUPS and CHI. To *Maha Elgarf*, *Sarah Falatous*, *Ahmed Tarek*, *Hana Medhat*, *Amr Kayid*, *Shaimaa Lazem*, *Ghada Hamdy*, *Heba Ayman* and *Amal Magdy* for always checking on me.

To *Habiba Farzand*, for your support during my stay in Glasgow, for our walks, culture talks, bearing out my crazy activities, and for always having and sharing ideas for the next project. To *Florian Mathis*, for being an inspiration for hard work, for your advises, for our career talks, for the avatar dancing hacks in virtual IEEEVR, for the fun during the conferences and for giving me different nicknames. To *Cristina Finai*, for all the coffee outings, and activities we had in Glasgow, for languages and culture talks, and for choosing me as an advisor. To *Melvin Abraham*, for being so helpful, for offering your desk when possible, for the food and culture talks, and for being a nice person. To *Omar Namnakani*, for all the eye-tracking conversations, the fun during the student master thesis supervision, the coffee breaks and CHI writing marathon.

To the entire SIRIUS team, *Noora AlSakar*, *Norah Alotaibi*, *Robin Bretin*, *Joseph O'Hagan*, *Karola Merky*, *Pejman Saeghe*, *Md Shafiqul Islam*, and *Shaun MacDonald*, for your support and all social activities and fun I had in Glasgow.

To *Ammar Eltaie*, for our office conversation, your advice and support, culture and food conversations, and for the fun during our CHI 2023 paper. To *Maha Alanquoudi*, for being an amazing friend, coworker, and writing buddy. For making my stay in Glasgow pleasant, for all our cultural conversations, coffee breaks, and fire alarms fun, you are such an inspiration.

To the MIMUC team and especially, *Linda Hirsch*, *Jingyi Li*, *Amy Yanhong Li*, *Yomng Ma*, *Fiona Draxler*, *Annika Kaltenhauser*, and *Hai Dang* for all the fun during the conferences and inspiring conversations during the IDCs.

To all my students and coauthors, for your help. To my scholarship providers *RSE* and *Studienstiftung des deutschen Volkes*.

To *MSI TUM*, for making my transition to Munich much more bearable and giving me the opportunity to meet such great friends. To *DAR e.v. family*, for being my family in Munich, for supporting me to grow and learn, and for believing in me.

To the *Lit Girls*, Hande, Warda, Ummara, Jannat, Sara, Fatime, and Sejla for being my support system, small family in Munich and for all the cooking and dancing nights. To my friends *Mariam, and Zeineb*, for your continuous support, for taking care of me, for your stimulating conversations, trips, movie nights, and cultural and language conversations. To *Hajar Ibrahim*, for always taking care of me, especially during my snowboarding accident, for being a sister, and for all our conversations, volunteering work, DAR e.v and coffee outings.

To *Omar Abdelwanis. Dina Galal, Farah ELdaour, Youssef Tarek, Mahmoud Yassin, Ahmed Abdelbadie, Dalia Maarek, Rana Saeed, Ahmed Darwish1, Ahmed Badran, Mohamed Abu Zahra, and Ahmed Darwish2*, for being amazing neighbors and friends, for the hiking trips, conversations, sharing food and for your continuous support.

To *Ghada Abdelwareth* for supporting me going our of my comfort zone and try new things. To *Sarah Seleem* and *Passant Sabri* for sharing my meme love, support through thesis writing and all Egyptian ladies outings.

To my Grandmother RIP, who's prayers are still accompanying me until this day. To my brother *Ali* and my sister *Reem*, for your continuous support, sushi and movie nights and for your support through writing this thesis, for your manuscript edits, figure inspirations, code debugging, and for bearing with my temper when I am stressed.

To my Mother *Alzahraa* and my father *Essam* without their support this thesis would not have been possible. Thank you for your continuous support, for listening to my complaints, and for your prayers. There are no words that can ever express how thankful I am. I know this journey was hard on all of us, but thank you for believing in me and supporting me throughout my entire life.

Finally, to myself for the hard work, sleepless nights, and for showing me that I can.

Thank You.

TABLE OF CONTENTS

List of Figures xix

List of Tables xxi

List of Acronyms xxiii

I INTRODUCTION AND BACKGROUND 1

1 Introduction 3

1.1 Research Questions 5
1.2 Methodology 5
1.3 Research Context 8
1.4 Research Contributions 9
1.5 Contributing Publications 10
1.6 Thesis Structure 10

2 Foundation and Background 13

2.1 Eye Gaze and Human-Computer Interaction 13
2.1.1 History of Eye Tracking 13
2.1.2 Eye Tracking Techniques 14
2.1.3 Gaze Behavior 14
2.1.4 Eye Movement Metrics 15
2.2 User Centered Security 16
2.2.1 Authentication and Knowledge-based Passwords 17
2.2.2 Password Heuristics and Strength Meter 18
2.2.3 Users' Passwords Habits 20

3 Related Work: Gaze in Security and Privacy Applications 23

3.1 Gaze-based Authentication 24
3.1.1 Explicit Gaze-based Authentication 25
3.1.2 Implicit Gaze-based Authentication 30
3.1.3 Gaze-supported Multi-factor authentication 33
3.2 Gaze-based Privacy Protection 34
3.2.1 Active Visual Privacy Protection 35

3.2.2 Raising Awareness of Shoulder Surfers in Real Time 36
3.3 Improving Security based on Gaze Behavior 36
3.4 Chapter Summary 37

II UNDERSTANDING GAZE BEHAVIOR DURING AUTHENTICATION 39

4 Gaze Vs. Touch for Knowledge-based Authentication 43
4.1 GazeLockPatterns: Implementation 45
4.2 Evaluation 45
4.2.1 Study Design 45
4.2.2 Apparatus and Participants 46
4.2.3 Procedure 46
4.3 Results 46
4.3.1 Pattern Length 46
4.3.2 Intersections 47
4.3.3 Overlaps 48
4.3.4 Knight Moves 48
4.3.5 Observation Risk 49
4.3.6 Memorability 49
4.3.7 Perceived Strength 49
4.3.8 Start and End Position 50
4.3.9 Additional Analysis: Gaze Calibration 50
4.3.10 Discussion 51
4.4 Chapter Summary 51

5 Cognitive Load and Knowledge-based Authentication 53
5.1 Concept and Methodology 55
5.1.1 Password Strength Meter 55
5.1.2 Mean Pupil Diameter Change Calculation 55
5.2 Evaluation 56
5.2.1 Study Design 56
5.2.2 Participants and Apparatus 56
5.2.3 Procedure 57
5.3 Results 58
5.3.1 Rated Password Strength 58
5.3.2 Post Study Question Analysis 59
5.3.3 Pupil Diameter and Password Strength 59

5.4 Discussion . . . 60
5.5 Chapter Summary . . . 61

III ENHANCING EXISTING AUTHENTICATION SYSTEMS 63

6 Detecting Password Strength from Gaze Behavior 67
6.1 Eye Tracking for Password Strength Classification . . . 68
6.1.1 Password Strength . . . 69
6.1.2 Perceived Password Strength . . . 69
6.2 Study . . . 70
6.2.1 Design . . . 70
6.2.2 Apparatus . . . 71
6.2.3 Recruiting and Procedure . . . 71
6.2.4 Limitations . . . 71
6.3 Methodology . . . 72
6.3.1 Statistical Analysis and Password Strength Estimation . . . 72
6.3.2 Feature Extraction . . . 72
6.3.3 Classification Approach . . . 73
6.4 Results . . . 74
6.4.1 Weak vs Strong Passwords . . . 74
6.4.2 Post Study Questions Analysis . . . 76
6.4.3 Gaze behavior Statistical Analysis . . . 76
6.4.4 Classifiers Performance . . . 77
6.5 Discussion . . . 79
6.5.1 Classification Performance . . . 80
6.5.2 Features Performance . . . 80
6.5.3 Input Modality Effect . . . 81
6.5.4 Influence on Security . . . 81
6.6 Chapter Summary . . . 81

7 Detecting Password Reuse from Gaze Behavior 83
7.1 Focus Group . . . 85
7.2 Concept and Research Questions . . . 86
7.2.1 Gaze Behavior Analysis . . . 87
7.2.2 Phases of Password Creation . . . 87
7.2.3 Research Questions . . . 88
7.3 Evaluation and Data Collection . . . 89
7.3.1 Study Design Considerations . . . 89

7.3.2 Study Design and Apparatus . . . 89
7.3.3 Study Setting, Procedure and Recruiting . . . 90
7.3.4 Limitations . . . 91
7.4 Feature Extraction and Classification . . . 91
7.4.1 Feature Extraction . . . 92
7.4.2 Classification Approach . . . 93
7.5 Results . . . 94
7.5.1 Participants . . . 94
7.5.2 Data Pre-Processing and Overview . . . 95
7.5.3 New vs. Reused Passwords . . . 95
7.5.4 Gaze Path . . . 97
7.5.5 Classifier Performance . . . 97
7.5.6 Effect of Data Sensitivity on User Behavior . . . 101
7.6 Discussion . . . 101
7.6.1 Gaze is More Informative than Typing . . . 102
7.6.2 Data Sensitivity Influences Accuracy of Password Reuse Prediction . . . 102
7.6.3 Dissecting Password Registration Process Enriches Modeling and Prediction . . . 103
7.7 Practical Implications for the Design of Password Systems . . . 103
7.7.1 Ubiquitous Eye Tracking . . . 103
7.7.2 Modeling . . . 104
7.7.3 User Privacy . . . 106
7.8 Chapter Summary . . . 106

IV DISCUSSION AND CONCLUSIONS 107

8 Framework 111
8.1 Understanding User Behavior . . . 111
8.1.1 Methodological Considerations . . . 112
8.1.2 Technical Considerations . . . 115
8.1.3 Empirical Considerations . . . 118
8.2 Modeling User Behavior . . . 119
8.2.1 Methodological Considerations . . . 119
8.2.2 Technical Considerations . . . 120
8.2.3 Empirical Considerations . . . 121
8.3 Providing Feedback . . . 121
8.3.1 Methodological Considerations . . . 121

8.3.2 Technical Considerations . . . 122
8.3.3 Empirical Considerations . . . 124
8.4 Ethical Considerations . . . 124
8.5 Chapter Summary . . . 124

9 Discussion 127
9.1 Discussion of Findings . . . 127
9.2 Gaze Behavior Beyond Authentication . . . 130
9.3 Ethics and User Privacy . . . 131
9.4 Towards Gaze-based Behavioral Security Systems . . . 132
9.5 Chapter Summary . . . 133

10 Conclusion 135
10.1 Summary of Contributions . . . 135
10.1.1 Systematization Of Knowledge on the Usage of Gaze in Security 135
10.1.2 Understanding Users' Gaze Behavior During Knowledge-based Authentication9n . . . 136
10.1.3 Eye Gaze Modeling to Enhance Knowledge-based Authentication . . . 136
10.1.4 Framework for Employing Gaze Behavior in Security Systems 136
10.2 Directions for Future Research . . . 137
10.2.1 Raising Users' Awareness and Implement Privacy Mitigation Techniques . . . 137
10.2.2 Exploring Different Physiological, Behavioral, and Cognitive Measurements . . . 137
10.2.3 Exploring Different Security Domains . . . 138
10.2.4 In-the-Wild Behavioral Gaze Data Collection . . . 138
10.3 Closing Remarks . . . 139

V BIBLIOGRAPHY 141

Bibliography 143

LIST OF FIGURES

1.1 Thesis Structure 11

3.1 Overview of the different application areas of eye tracking in security and privacy 24

4.1 Gaze distance relative to the 9 grid digits 47
4.2 The frequency of the users' chosen pattern lengths 47
4.3 Start and End Pattern Position for Gaze and Touch Conditions 48
4.4 Sample gaze path when entering lock patterns with gaze & touch . . . 48

5.1 Study Setup 57
5.2 Password strength comparison 57
5.3 MPD across allparticipants and per password 58
5.4 MPDC across all participants per password 59

6.1 Password strength comparison per input modelity 75
6.2 Password strength comparison per passwords strength 75
6.3 Confusion matrix for the user-independent, modality-dependent classifier 78
6.4 Feature Importance 79
6.5 Gaze plots during entering weak and strong passwords 79

7.1 Example of the Focus Group Responses 86
7.2 Phases of password registration 87
7.3 Study Setup 87
7.4 Registration interfaces 90
7.5 ML Classification Steps from data preparation until sending the data to the classifier [9]. 94
7.6 Visualization of selected users' gaze paths per interface 97
7.7 Results of the feature importance analysis 102
7.8 AUC comparison for multiple phases classifier 102

8.1 Comparison between three eye tracking glasses 117
8.2 Interplay of Actual User Behavior and System Prediction. 123

9.1 Password Choices branch addition to Kröger et al. [217] 133

LIST OF TABLES

1.1 Overview of the research questions addressed in the thesis with chapter number between brackets. 6
1.2 Research systems and prototypes developed in the thesis. 8

2.1 Fixation-derived metrics and Their Interpretations 16
2.2 Saccade-derived metrics and their Interpretations 17
2.3 Scanpath-derived metrics and their Interpretations 18

4.1 Comparison between gaze and touch modalities for the 3 situations . . 46

5.1 MPD difference between creating weak and strong passwords 59

6.1 Differences between actual and perceived password strength 70
6.2 Password Characteristics 74
6.3 ANOVA results for eye movements per password strength 77
6.4 ANOVA results for eye movements per input modality 77
6.5 AUC of the three Classifiers per input modality 78

7.1 Gaze Features 92
7.2 Keystroke Features 93
7.3 Number of new and reused passwords and task completion time . . . 95
7.4 Wilcoxon signed-rank tests for new and reused passwords 96
7.5 Wilcoxon signed-rank tests for keystroke dynamics 96
7.6 Wilcoxon signed-rank tests for gaze features 96
7.7 Interface-dependent Classifier per phase 98
7.8 Interface-independent Classifier per phase 99
7.9 Classification performance for the interface-dependent classifier for multiple phases 100
7.10 Eye movements for the different interfaces 100
7.11 Comparison of keystroke dynamics per interface 101

8.1 Framework for Employing Eye Gaze Behavior in Security Systems . . 111
8.2 Comparison between the most common Webcam-based eye trackers . 117

List of Acronyms

HCI	Human-Computer Interaction
ACM	Association for Computing Machinery
LBW	Late Breaking Work
API	Application Programming Interface
PIN	Personal Identifcation Number
AR	Augmented Reality
UI	User Interface
HMD	Head-Mounted Display
RQ	Research Question
CSV	Comma separated values
AOI	Area of Interest
TLX	NASA Task Load Index
IV	Independent Variable
DV	Dependent Variable
KNN	K-Nearest-Neighbor
SVM	Support Vector Machine
RF	Random Forest
LR	Liner Regression
DT	Decision Tree
AUC	Area Under the Curve
ML	Machine Learning
O Phase	Orientation Phase
ID Phase	Identification Phase

C Phase	Confirmation Phase
AWI	Aware User Interface

I

Introduction and Background

Chapter 1

Introduction

Nowadays, users are surrounded by different forms of ubiquitous devices including but not limited to smartphones, laptops, public displays, and smart glasses. As these devices enable access to sensitive information, security mechanisms are becoming crucial. At the same time, secure data access remains a challenge. One of the main reasons is that security methods are frequently not built to accommodate how people actually behave. Consider a password-based authentication system. This system, which requires a username and a password to authenticate a user, was created for settings where users had access to one computer and authenticated a few times each day during the mainframe era in the 1960s. Surprisingly, despite the fact that people now interact with computers in fundamentally different ways, this notion is still one of the most widely used security concepts today. For example, in 2007, users had around 25 accounts and roughly six different passwords on average [118]. Nowadays, people retain an average of 80 different online accounts, each with its own password [242].

This increase in the number of accounts and profiles by users amplified password issues where people started to pick weak passwords, write them down in insecure locations, reuse many of them, and share credentials with other people [345]. These behaviors are referred to as *password coping strategies*. A common solution is to reuse the same password on many different websites. This approach increases the potential damage if a password is stolen, or cracked, or if a service that has access to it is compromised since the attacker will be able to reuse it on all online services that share this password. In addition, to make accounts secure, users are required to change their passwords at periodic intervals which require physical and mental demand. Therefore, both researchers and practitioners investigated different solutions to solve this problem such as biometric authentications. However, many users are concerned about the leakage of their face ID or fingerprints and the transparency of sharing such information with 3rd parties. In addition, there is a possibility that existing fingerprint sensors could be hacked by a simple printed image[1]. Furthermore

[1] https://uk.pcmag.com/security/137235/hacking-fingerprints-is-actually-pretty-easy-and-cheap

all biometric systems have passwords as an alternative authentication technique in case the original one did not work. On the other side, behavioral biometrics also gained huge attention from the community and are already adopted in most smartphones such as using gait cycles to safely unlock the phone if the user is detected. Still the issue of collecting biometric data and its privacy concerns persevere, in addition to varying user states and size of the collected data [2].

Another approach is using password managers. Password managers help users to generate, store and manage their passwords [77, 158]. However, many users are hesitant to adopt password managers: according to a recent study conducted by PasswordManager.com and YouGov[3] among 1280 US citizens, nearly two-thirds of participants do not trust password managers. Furthermore, prior research has shown that password managers do not always solve the increase of passwords problem as a significant number of password manager users continue to reuse passwords [281]. In fact, the literature showed that password managers face different security and privacy challenges in addition to introducing new problems [400, 232, 132, 333, 346]. Such problems include and are not limited to 1) *single point of failure* where if it was attacked then all the other passwords are leaked, 2) not all password managers work across devices so this limits its usage, 3) some password managers are vulnerable to different types of attacks such as cross-site scripting (XSS) attacks, cross-site request forgery (CSRF), and network injection attacks, etc.

Although passwords are far from perfect as they demand physical and mental effort and are prone to several attacks [32, 52, 313] they remain essential [353]. The main reason is that currently, passwords present a balance between usability, security, and administrability [53] with no other mechanisms providing an equally good trade-off between the effort required for implementation, ease of administration (e.g., resetting/changing credentials), ease of use, and security. Therefore, until we get rid of passwords, we need to make them more secure.

On the other hand, eye-tracking technology is becoming more accurate, affordable, and ubiquitous. Eye gaze is part of humans' active sensing which is a key process in our interaction with the world as it allows our sensors to be directed to the environment in order to extract relevant information [390]. During interaction tasks, most of the time, our gaze behavior is regulated unconsciously by cognitive processes that are responsible for task performance. As a result, natural gaze behavior provides a window into the human mind and allows us to make conclusions about users' intentions and cognitive processes [36]. Hence, the security community was one of the first to employ eye-tracking technology. For example, eye gaze has been used in implicit [397, 96] and explicit authentications [220, 6]. In addition, gaze monitoring have been used to protect users from bystanders [190, 20, 293] and shoulder surfing attacks [58, 315, 402, 401]. Researchers also used eye tracking to study the effectiveness of the

[2] https://www.miteksystems.com/blog/advantages-and-disadvantages-of-biometrics?msclkid=82e2d880c1ac11ec9d9892c2e3484895

[3] Password Managers Survey: https://www.passwordmanager.com/password-manager-trust-survey/

security indicators on the websites [29], privacy policies [343] and email classification decisions [283, 259]. A recent literature survey showed that by monitoring users' gaze data, we can reveal different information about the users such as their age, gender, biometric identity, physical health, and many more [217]. Furthermore, researchers found that certain patterns in eye movement, pupil dilation, and eye blinking have been recognized as reliable indicators of mental workload [110, 253].

From previous points, we see monitoring natural gaze behavior to be particularly promising to enhance security mechanisms and passwords specifically. Unfortunately, there is a lack of understanding of how users' gaze behavior is while they create their passwords and which information is reflected in their eye gaze behavior. Consequently, it is essential to first understand users' gaze behavior in password creation in order to enhance existing systems and build novel ones. In this thesis, we combine the advantage of involuntary eye movements, with the ubiquity of eye trackers to enhance existing security mechanisms. Particularly, we will investigate the usage of eye gaze behavior in supporting users' password decisions as a novel approach to nudging and protecting users while creating passwords. Below, we identify the research questions.

1.1 Research Questions

Security decisions are built upon habits carried out frequently so they become automatic routines that people perform without thinking. Password creation and usage strategies are one of those routines that users developed over time as a coping strategy for having passwords for more than 80 accounts [242]. Our vision is to make passwords more secure by leveraging eye gaze behavior. First, we start with a holistic analysis of how eye gaze has been used in the past decades in security mechanisms (**RQ1**). From there, we narrow it down to monitoring users' eye gaze behavior to enhance authentication techniques. Then we dig deeper into knowledge-based authentication to understand users' gaze behavior and what influences it (**RQ2**). After understanding users' gaze behavior, we investigate how to quantify and model such behavior to enhance users' passwords (**RQ3**). Finally, we close the loop by applying our knowledge from the previous research questions to derive a set of implications in a framework for behavioral eye gaze security systems (**RQ4**). The questions are summarized in Table 1.1.

1.2 Methodology

We followed a user-centered design approach [162] as one of our principal methodologies. User-centered design is an iterative design process in which developers start with a general approach and focus on users' feedback in each iteration of the

Table 1.1: Overview of the research questions addressed in the thesis with chapter number between brackets.

CONCEPTUAL/ THEORETICAL	TECHNICAL	EMPIRICAL	METHODOLOGICAL
RQ1: How can security mechanisms benefit from users' eye gaze?			
Literature survey highlighting the research gaps and promising research directions [Ch 3]			
RQ2: What is the influence of knowledge-based authentication on users' gaze behavior?			
Concept of studying cognitive load and password creation [Ch 5]	Android pattern prototype to be used by either gaze or touch [Ch 4]	Lab study to collect touch and gaze dataset in addition to users' entered patterns and post-study questionnaires [Ch4]. Lab study to collect pupil diameter and questionnaire responses. Dataset of collected passwords and dataset of pupil diameter [Ch 5]	
RQ3: How can users' behavior during authentication be modeled?			
Concept of detecting password strength [Ch 6] and reuse [Ch7] from gaze behavior	Machine learning model to detect password strength [Ch 6] and password reuse [Ch 7] from users' gaze behavior	Lab study to collect users' gaze behavior while authenticating. Dataset of collected weak and strong password dataset of gaze data and movements [Ch 6]. Semi-controlled study to collect users' gaze data while authenticating, and another dataset of eye gaze data while creating new and reused passwords [Ch 7]	Conducting evaluation at participants' side (lunch cafeteria) [Ch 7]
RQ4: What are the considerations for employing gaze behavior in security systems?			
			Framework for employing eye gaze behavior in security systems [Ch 8]

continuous design process. We further apply design thinking methods, and co-design workshops to better shape the ideas. During the development process of our research probes, we evaluated these in user studies. With our empirical research practices, we foster a better understanding of users' gaze behavior during authentication and how it can be used to make systems more secure. Consequently, our findings provide knowledge that enables us to use users' gaze behavior as a means to assess passwords and add a behavioral aspect as part of password assessment to enhance password creation.

Systems and Prototypes

All the research systems and prototypes conducted in this thesis were developed in close collaboration with my colleagues and other researchers, with several undergraduate students, contributing to the development as part of their work. All prototypes were implemented with the objective of being deployed and evaluated in lab studies. A summary of the research prototypes that have been developed is shown in Table 1.2.

Controlled and Semi-controlled Studies

In our conducted evaluations, we used semi-controlled lab studies. As we are studying users' behavior, we did not have a fixed head position nor distance to the screens. However, as users' eyes are sensitive to the lights, in most cases we had controlled light systems to eliminate this factor.

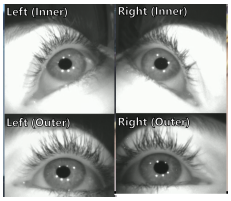
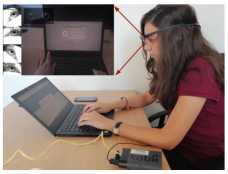
Evaluation

To evaluate our concepts and prototypes, we used several methodologies. We conducted co-design workshops to finalize the concepts. We also conducted controlled lab studies through which ecologically valid user data were collected. We used machine learning classifiers' accuracy as a tool to evaluate our concepts. We also used a mixed methods approach for data analysis such as different statistical analysis approaches and collecting subjective and objective feedback. This was done using a wide-ranging set of tools including focus groups, data logging, standardized and custom questionnaires, and participant observations.

Deception Studies

As we used the user's gaze behavior to understand more about the user, it was crucial in most cases that users do not know the main aim of the studies, hence, we conducted deception studies. Deception studies are used to eliminate user priming and at the same time elicit natural user behavior. It is worth mentioning that participants were debriefed with the true study aim at the end and they had the right to opt-out.

Table 1.2: Research prototypes developed in the course of this thesis.

SYSTEM/PROTOTYPE	DESCRIPTION	CHAPTER
	**GazeLockPatterns.** To understand users' gaze behavior while creating knowledge-based passwords and to highlight the reasons to improve existing authentication techniques, we monitored users' gaze while creating knowledge-based passwords. We monitored users' eye gaze while entering Android touch patterns and compared their behavior to entering the patterns using gaze. *Publication:* [7].	4
	**Think Harder.** To better understand the users while creating their knowledge-based passwords and to detect if password creation affects users' cognitive load. We envisioned a simple user interface where we monitored users' gaze movements with a wearable eye tracker and highlighted their areas of interest. *Publication:* [5]	5
	**GazeMeter.** There are two different ways where eye gaze data can enhance existing password creation methods either by enhancing existing password strength metrics or by creating new password strategy metrics. To enhance existing knowledge-based password strength metrics, and help users not to create weak passwords. We monitored users' eye movements while creating weak and strong passwords and built an ML classification model to predict password strength from users' eye gaze movement. *Publication:* [10]	6
	**Password Reuse.** To use eye gaze behavior to create new password strategy metrics. As password reuse is the second most critical password strategy failure. We monitored users' eye gaze movement to detect password reuse. We built ML classification models to predict password strength from users' eye gaze movement and from keystroke dynamics and we compared both. *Publication:* [9].	7

1.3 Research Context

The research leading to this thesis was carried out during the period from 2019 to 2022 at the Bundeswehr University Munich in the Usable Security and Privacy group (USEC) and at the University of Glasgow in the SIRIUS research group. Many of the projects presented in the thesis were in collaboration with experts from the field.

All of the projects presented in this thesis [181, 7, 5, 10, 9] are in collaboration with the University of Glasgow, colleagues from the USEC lab, and undergraduate students both bachelor and masters from Bundeswehr University Munich Germany, LMU Munich Germany and the German University in Cairo, Egypt.

Together with Christina Katsini and George Raptis from the Human Opsis Lab in Greece, we worked on a literature survey in 2019 [181] summarizing the literature work that used eye tracking in the usable security field. This joint work resulted in an honorable mention award at CHI 2020 conference. In addition, we followed this by organizing a workshop at ETRA 2021 on *EyeSec: Eye Gaze for Security Applications*.

In cooperation with Anam Khan from the University of Melbourne, we worked on investigating password strength from users' gaze behavior [10]. We also worked closely with Dr. Daniel Buschek from the University of Bayreuth an expert in machine learning and usable security and Dr. Ken Pfeuffer from Aarhus University an expert in eye tracking on investigating eye gaze behavior for password reuse detection [9].

In addition, we worked with Ahmed Shams and Mohamed Mantawy from the German University Cairo on password strength detection [10] and with Ahmed Shams again on password reuse detection [9].

1.4 Research Contributions

In answering our overarching research question of how can users' gaze be used to enhance existing security mechanisms, this thesis makes contributions in four main areas. First, we conduct a holistic understanding of how gaze has been used in the security and privacy domain. Then we use this knowledge to draw our concepts of using gaze behavior to enhance password creation **C1**. Second, we build different prototypes to evaluate our concepts **C2**. Third, we conduct empirical evaluations to collect and analyze users' behavior while creating passwords **C3**. Fourth, we present a framework for employing eye gaze behavior in security systems **C4**.

C1: Conceptual - We present in chapter 3 a holistic understanding of how gaze has been used in security and privacy applications. Moreover, we introduce the concept of using gaze behavior to enhance knowledge-based authentication which is reflected in different prototypes.

C2: Technical - In chapters 4, 5, 6, and 7, we present different systems and prototypes that allowed us to study our concepts and be able to collect empirical data. In the context of this thesis, we developed five research prototypes. A list of the prototypes can be seen in Table 1.2. Moreover, from each prototype, we collected a dataset for gaze movements.

C3: Empirical - Throughout the exploration conducted during this thesis, we performed different user studies testing our research prototypes and systems and collecting different data sets. The collected data enabled us to implement different classification models as well as draw different conclusions. The different studies with the evaluations introduced in this thesis are structured according to the relevant research question.

C4: Methodological - In chapter 7 we conducted a different type of study where we went to the participants in their environment and asked them to participate in the study. In chapter 8, we reflect on our findings from the thesis and present a framework for employing behavioral eye gaze in security systems.

1.5 Contributing Publications

This thesis's core contributions are based on research published at international conferences with a competitive peer-reviewing process. All publications benefited from close collaboration, ideation, and discussions with the stated co-authors. However, for the core publications that lead to this thesis, I took the lead during the development, study design, analysis, and writing as clarified in section .

1.6 Thesis Structure

This thesis is made up of ten chapters divided into four parts. Figure 1.1 depicts the interplay between the different structures of this thesis.

Part I - Introduction and Background This thesis part consists of 3 chapters. **Chapter 1**, reflects the introduction and motivation for the thesis. We introduce the topic at hand, the research objectives, and covered research questions, as well as illustrate the contributions of the thesis. In **chapter 2**, we provide an overview of how eye gaze has been used in the security domain in the past 25 years. The background and foundation chapters do not cover the detailed state-of-the-art related work to each of the empirical research probes later presented, but rather aim to give a historical and foundational background of the related topics to the core of the thesis and an overall view of how eye gaze has been used in the usable security domain. Finally, in **chapter 3**,

Part II - Understanding Gaze Behavior During Authentication This thesis part consists of two main chapters. In **chapter 4**, we present a comparison between Android's lock patterns for mobile devices (TouchLockPatterns) and an implementation of lock patterns that uses gaze input (GazeLockPatterns). We monitored users' eye gaze while entering Android touch patterns and compared their behavior to entering the patterns using gaze. We reflect on users' gaze positions while performing the touch patterns. In **chapter 5**, we investigate the relationship between password creation and cognitive load inferred from eye pupil diameter. We used a wearable eye tracker to monitor the user's pupil size while creating passwords with different strengths.

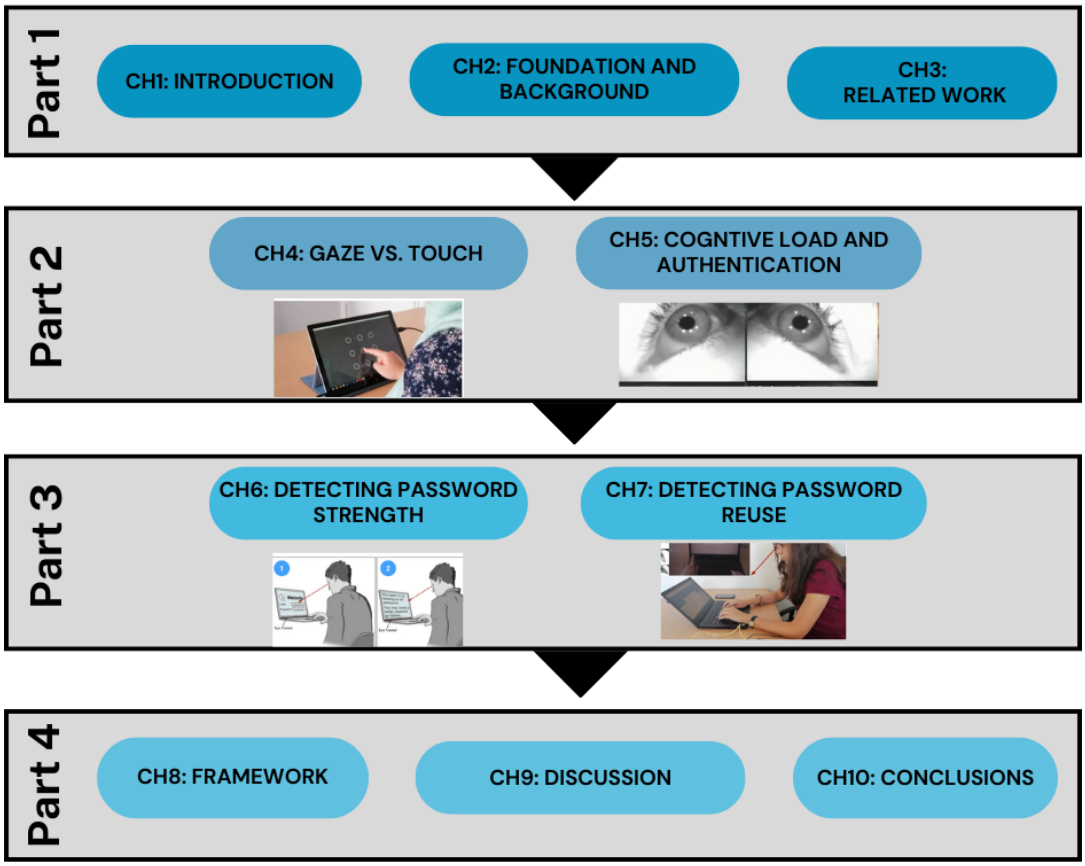


Figure 1.1: Thesis structure and interconnection between chapters

Part III - Enhancing Existing Authentication Systems This thesis part consists of two chapters. In **chapter 6**, we investigate the use of gaze behavior as a means to assess password strength as perceived by users. We demonstrate how eye tracking can enable this: by analyzing people's gaze behavior during password creation and running different machine-learning models. In **chapter 7**, we investigate and propose a novel approach to detect password reuse by leveraging eye gaze and comparing it to typing behavior, and studying its accuracy. We compare different ML models to detect password reuse from the collected data.

Part IV - Discussion and Conclusions This part consists of three main chapters. **Chapter 8** reflects on the synthesis and main findings from the thesis in a framework for employing behavioral eye gaze into security systems. **chapter 9** summarizes the results from the investigations of the research probes and empirical research conducted and presents different discussion aspects. Finally, in **chapter 10**, we summarize the contributions of this thesis and provide an outlook for future work.

Chapter 2

Foundation and Background

In this chapter, we discuss the foundations and background of several topics relevant to the overall thesis. The chapter is divided into two main sections. We reflect on the human eye and gaze metrics. Then we reflect on user-centered security with a focus on knowledge-based authentication.

2.1 Eye Gaze and Human-Computer Interaction

As William Shakespeare said, *"The Eyes are the window to your soul"*. Our eyes play a huge role in our communication and how we perceive our surroundings. From early stages, scientists have studied human eye anatomy and how can it be used in Human-computer Interaction. But first, how can we track users' eyes?

2.1.1 History of Eye Tracking

Since the development of eye-tracking equipment in research more than a century ago, numerous approaches have been utilized to monitor user eye movements [300]. For instance, electrooculographic methods used electrodes mounted on the skin around the eyes to measure variations in electric potentials and track eye movements. Other traditional methods involved wearing large contact lenses that covered the cornea and sclera (the visible white of the eye) and had a metal coil embedded around the edge of the lens. Eye movements were then detected by changes in an electromagnetic field as the metal coil moved in parallel with the subject's eyes [103]. The majority of contemporary eye-tracking systems currently employ video images of the eye to determine where a person is looking (i.e., their so-called "point-of-regard"). These techniques proved to be highly invasive. The iris-sclera boundary, apparent pupil

shape, and corneal reflections (also known as Purkinje images) are just a few of the unique characteristics of the eye that can be utilized to infer point-of-regard [103].

2.1.2 Eye Tracking Techniques

There are several methods available today for tracking users' eyes, some of which rely on infrared light and others on gaze estimation. Additionally, eye-trackers come in a variety of forms. One can employ wearable eye trackers, screen-attached eye trackers, or webcam/RGB cameras to estimate gaze.

The majority of currently available commercial eye-tracking systems measure point-of-regard using the "corneal-reflection/pupil-center" technique [291]. These types of trackers often comprise an infrared camera installed beneath (or next to) a display monitor on a regular desktop computer, together with image processing software to locate and identify the features of the eye utilized for tracking. In order to make the target eye features easier to track, infrared light from an embedded LED in the infrared camera is first directed into the eye (infrared light is used to avoid dazzling the user with visible light). The "bright pupil" effect occurs when light enters the retina and a significant portion of it is reflected back, giving the pupil the appearance of a bright, well-defined disc. The infrared light also produces the corneal reflection (or first Purkinje image), which appears as a tiny but distinct shine spot. Once the corneal reflection and the pupil center have been located by the image processing software, the vector between them is measured and the point of regard can then be determined using additional trigonometric calculations. Although the corneal reflection alone can provide an approximate point-of-regard, measuring both aspects allows eye movements to be effectively separated from head movements [167, 103].

A "calibration" process is required to adjust video-based eye trackers to the specifics of each person's eye movements. The system records the pupil-center/corneal-reflection relationship as corresponding to a specific x, y coordinate on the screen if the eye fixes on the dot for a time period greater than a predetermined threshold and within a predetermined area. To obtain a precise calibration throughout the entire screen, this is performed across a 9 to 13-point grid-pattern [291].

2.1.3 Gaze Behavior

From previous work, gaze behavior was defined as a *"fixated gaze on a location in the targeting environment or a shift in gaze from one environmental location to another"* [369]. As a shift in gaze is typically initiated by eye movements four different gaze behaviors were defined in the literature for the minimum duration of fixations, saccades, tracking movements, and blinks [395, 136]. In the literature, the minimum gaze duration of a fixation ranged from 80 to 150 millisecond [69], with the shorter

duration being observed when subjects performed highly focused activities. The minimum fixation period was established at 99.9 milliseconds (3 or more frames) and was defined as the stabilization of the gaze on a point because the subjects in this experiment were highly skilled. When a shift in gaze from one place to another was noticed, with a minimum duration of 133.2 milliseconds, the movement was categorized as a saccade (4 frames). When a moving item was followed by the subject's gaze for at least three frames or 99.9 milliseconds, the eye movement was categorized as tracking or smooth pursuit. Finally, when the subject's pupils dilated and covered the system's optics for at least three frames or 99.9 milliseconds, the blink signal was activated. In the following section, we describe each of the gaze metrics in detail.

2.1.4 Eye Movement Metrics

There are different eye movements that can be quantified and extracted from users' eyes. There are two main measures that are mostly used in research "fixations" (described in 2.1.2) and "saccades", which are quick eye movements occurring between two fixations. Moreover, there is also a wide range of derived metrics, such as "scanpath", "blink rate" and "pupil size". Below we describe each of them, what they mean, how can we interpret them, and whether can they be used. The descriptions below are accommodated by Poole et al. [287].

***Fixations*:** depending on the context, fixations can be interpreted differently. For example, more fixation on a specific interface area (e.g. on a website) can indicate greater interest. On the other side, if it was a reading task then it might indicate a more complicated or challenging task [167, 171]. Similarly, in a search task, more fixations can mean ambiguity in identifying a target item [167]. A higher number of single fixations, or clusters of fixations, are often an index of greater uncertainty in recognizing a target item [167]. Table 2.1 reflects the fixation metrics and their interpretations.

***Saccades*:** don't include any encoding, so they can't provide information about the complexity or importance of an object in the interface. Regressive saccades, or retracing eye movements, can, however, be used to evaluate how difficult a processing task is during encoding [300]. While most "regressions" (or "regressive saccades") in reading tasks are extremely short, only skipping back two or three letters, considerably greater phrase-length regressions may indicate difficulties in a higher-level processing of the text [300]. There should be an inverse relationship between the number of regressions and the importance of the phrase, so regressions could also be employed as a measure of recognition value. Table 2.2 describes saccade-derived measures.

***Scanpaths*:** A scanpath describes a sequence of saccade-fixate-saccade. A straight line with two short fixations on the desired target is considered the best scan path in a search task [130]. Table 2.3 describes qualitative measures to analyze scanpaths.

Table 2.1: Fixation-derived metrics and how they can be interpreted in the context of interface design and usability evaluation as reported by Poole et al. [287]. References are given to examples of studies that have used each metric.

Eye-Movement Metric	What it Measures	Reference
Number of fixations overall	More overall fixations indicate less efficient search (perhaps due to sub-optimal layout of the interface)	Goldberg & Kotval (1999) [130]
Fixations per area of interest	More fixations on a particular area indicate that it is more noticeable, or more important, to the viewer than other areas	Poole et al.(2004) [288]
Fixations per area of interest and adjusted for text length	If areas of interest are comprised of text only, the mean number of fixations per area of interest should be divided by the mean number of words in the text. This is necessary to separate out: (i) a higher fixation count simply because there are more words to read, from (ii) a higher fixation count because an item is actually harder to recognize.	Poole et al.(2004) [288]
Fixation duration	A longer fixation duration indicates difficulty in extracting information, or it means that the object is more engaging in some way.	Just & Carpenter (1976) [171]
Gaze (also referred to as "dwell, fixation cluster" and"fixation cycle")	Gaze is usually the sum of all fixation durations within a prescribed area. It is best used to compare attention distributed between targets. It can also be used as a measure of anticipation in situation awareness if longer gazes fall on an area of interest before a possible event occurring.	Mello-Thoms et al. (2004) [256], Hauland (2003) [145]
Fixation spatial density	Fixations concentrated in a small area indicate focused and efficient searching. Evenly spread fixations reflect widespread and inefficient search.	Cowen et al.(2002) [82]
Repeat fixations(also called "post-target fixations")	Higher numbers of fixations off-target after the target has been fixated indicate that it lacks meaning fulness or visibility.	Goldberg & Kotval (1999) [130]
Time to first fixation on-target	aster times to first-fixation on an object or area mean that it has better attention-getting properties.	Byrne et al. (1999) [64]
Percentage of participants fixating an area of interest	If a low proportion of participants is fixating an area that is important to the task, it may need to be highlighted or moved.	Albert (2002) [19]
On-target (all target fixations)	Fixations on-target divided by total number of fixations. A lower ratio indicates lower search efficiency	Goldberg & Kotval (1999) [130]

***Blink rate and pupil size*:** Pupil size and blink rate both serve as indicators of cognitive workload. It is considered that a higher workload is indicated by a lower blink rate and that a higher blink rate may signify tiredness [57, 59]. A greater cognitive effort may also be indicated by larger pupils [291, 349]. However, several other factors, such as surrounding luminance, can affect pupil size and blink rate, making them susceptible to contraction [291]. These factors make the use of pupil size and blink rate in eye-tracking studies less common.

2.2 User Centered Security

As the context of this thesis is to enhance existing security applications and mechanisms. First, let's define what we mean by security mechanisms. The term

Table 2.2: Saccade-derived metrics and how they can be interpreted in the context of interface design and usability evaluation as reported by Poole et al. [287]. References are given to examples of studies that have used each metric.

Eye-Movement Metric	What it Measures	Reference
Number of saccades	More saccades indicate more searching.	Goldberg & Kotval (1999) [130]
Saccade amplitude	Larger saccades indicate more meaningful cues, as attention is drawn from a distance	Goldberg et al. (2002) [131]
Regressive saccades (regressions)	Regressions indicate the presence of less meaningful cues.	Sibert et al.(2000) [331]
Saccades revealing marke addirectional shifts	Any saccade larger than 90 degrees from the saccade that preceded it shows a rapid change in direction. This could mean that the user's goals have changed or the interface layout does not match the user's expectations.	Cowen et al.(2002) [82]

User-Centered Security was first used by Zurko et al. [404] in 1997. It refers to security models, mechanisms, systems, and software where high usability is the main motivation or goal. There is a wide spectrum of areas under this umbrella such as *authentication*, which focuses on how mechanisms can be built that is secure but at the same time easy to use [97]. Much of the research in this domain is mostly focused on so-called *knowledge-based schemes*. However, there are different authentication methods as well such as implicit authentication and biometric authentication.

2.2.1 Authentication and Knowledge-based Passwords

Accessing different accounts and devices requires user authentication which is considered the first step in access control. Authentication is typically achieved by requesting a user to provide proof of one or more of the following [81]: Something you know (e.g. a password), Something you have (e.g. a token), Something you are (e.g. a fingerprint). In this thesis, we will only focus on *something you know* and we narrow it down to knowledge-based passwords. There are 31 different Knowledge-based authentication schemes grouped to 5 categories, Alphanumeric passwords, Graphical passwords, Haptic patterns, Semantic Knowledge Password and 2 factor authentication [107]. Different authentication methods, such as PINs (alphanumeric password), passwords (alphanumeric password), and lock patterns (graphical passwords), are likely to continue to be used in the future because they are simple to use and intuitive to users. Also, they can be used as a backup to other authentication methods (like fingerprints) or on devices without sensors that support biometric approaches. Passwords require unassisted recall, or the ability for the user to recall them precisely when asked, which is a requirement from the point of effort. Users who have to create and remember a large number of passwords tend to reuse them [118], write them down [14], or select passwords that are simple to

Table 2.3: Scanpath-derived metrics and how they can be interpreted in the context of interface design and usability evaluation as reported by Poole et al. [287]. References are given to examples of studies that used each metric

Eye-Movement Metric	What it Measures	Reference
Scanpath duration	A longer-lasting scanpath indicates less efficient scanning	Goldberg & Kotval (1999) [130]
Scanpath length	A longer scanpath indicates less efficient searching (perhaps due to a sub-optimal layout)	Goldberg et al. (2002) [131]
Spatial density	Smaller spatial density indicates more direct search.	Goldberg & Kotval (1999) [130]
Transition matrix	The transition matrix reveals the search order in terms of transitions from one area to another. Scanpaths with an identical spatial density and convex hull area can have completely different transition values – one is efficient and direct whilst the other goes back and forth between areas, indicating uncertainty.	Goldberg & Kotval (1999) [130], Hendrickson (1989 [146]
Scan path regularity	Once "cyclic scanning behavior" is defined, deviation from a "normal" scanpath can indicate search problems due to lack of user training or bad interface layout.	Goldberg & Kotval (1999) [130]
Spatial coverage calculated with convex hull area	Scanpath length plus convex hull area define scanning in a localized or larger area.	Goldberg & Kotval (1999) [130]
Scan path direction	This can determine a participant's search strategy with menus, lists and other interface elements (e.g.top-down vs. bottom-up scanpaths). "Sweep" denotes a scanpath progressing in the same direction.	Altonen et al.(1998) [82]
Saccade/fixation ratio	This compares time spent searching (saccades) to time spent processing (fixating). A higher ratio indicates more processing or less searching.	Goldberg & Kotval (1999) [130]

remember. Several studies showed that people tend to focus on a small number of predictable choices when choosing a password rather than using the full space of available options [52]. "Password" and "123456" are typically the most used passwords, according to a number of password database leaks (e.g.,'[52, 383]).

2.2.2 Password Heuristics and Strength Meter

To help users create strong passwords, password meters are integrated into interfaces to give users an estimate of how strong their passwords are and hence, how easy it is to be cracked [112] according to a set of heuristics. Such heuristics are and are not limited to minimum password length, password entropy, and the number of lower and upper case letters as well as digits and special characters. This is enforced on websites using password meters. The use of password strength meters was adopted a decade

ago [383]. Many studies investigated the effectiveness of using password meters on the security and memorability of passwords. For example, work by Ur et al. [94] showed that participants believe that adding an exclamation mark at the end of their passwords would make them stronger. Participants also believed that having keyboard patterns or adding their pet name in the password is an asset for a strong password. In 2013, Egelman et al. [112] examined whether the use of password meters influenced users' password strength or not. The authors asked participants to first change their real passwords according to the presence of the password meters, next to change an important account password, and finally to change an unimportant account password. They found that password meters significantly enhanced users' generated passwords for their real accounts and important accounts. However, for non-important accounts, password meters did not have an effect. The authors concluded that using password meters is only effective if the user is forced to change or create a password for an important account.

Further research by Ur et al. showed that also the appearance of the password meter affects the choice of passwords [364]. For example, meters without visual bars gave participants the impression that it is not important to enter a strong password and, hence, caused participants to put less effort into satisfying the meter's requirement. In contrast, participants who saw more lenient meters tried to fulfill the meter requirements and were reluctant to choose passwords a meter deemed as "bad" or "poor". Similarly, researchers found that password meters design, color, and feedback messages have an influence on the strength of the created passwords [112, 327, 108]. 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 [383, 186]. Furthermore, strict policies can frustrate users, reduce their productivity, and lead users to write their passwords down [15, 160, 340]. Ur et al. [363] found that users are aware of what makes a password strong. This suggests that putting more effort into creating a password might be an indication that it is a strong one.

Moreover, Shay et al. [328], studied the effect of password length on password strength. They found policies requiring longer passwords to reduce the percentage of easy-to-guess passwords. Also, enforcing combinations of certain requirements and increasing password length led to stronger passwords that was more usable compared to traditionally complex policies. Later, Shay et al. [327], studied the usability of feedback and guidance mechanisms for password meters. They found that service providers should present password requirements in combination with feedback to increase usability. However, feedback needs to be designed carefully, as the same requirements can have different security and usability effects depending on the way they are presented. In 2017, Ur et al. [362] proposed a 'Data-Driven Password Meters'. The meter communicates up to 3 ways to the user how the entered password can be enhanced. The results showed that data-driven meters with detailed feedback led users to create more secure, yet equally memorable passwords, compared to normal meters

with a strength bar indicator. Research by Dupuis et al. [106], studied the effect of changing the feedback on generated passwords' strength. Instead of indicating the actual password strength, they provided a comparison of the strength to the passwords of other users. For example, instead of showing *weak password*, they showed *weak compared to other users*. The authors report that by changing the feedback mechanism and comparing users' passwords to others, people generated stronger passwords.

2.2.3 Users' Passwords Habits

People have on average 80 accounts, protected with an average of 3.5 passwords. This makes password memorability challenging [139]. To cope with this, users employ different strategies. Popular ones are choosing easy-to-remember passwords (e.g., 'password' or '123456'), reusing passwords, and writing down passwords. According to a survey by Google, 65% of users reuse passwords for some or all of their accounts[4]. Hence, the community focused on better understanding user behavior regarding password reuse and concepts to mitigate such behavior.

Wash et al. have studied users' password reuse behavior [380]. The authors created a Web browser plugin to collect user passwords across frequently used websites. Their results showed that people reuse strong passwords more frequently across different websites. Pearman et al. conducted an in-situ study to understand users' password managing behavior [280]. The authors found that the larger the number of accounts a user has, the higher the chances are that they reuse parts or all of their passwords across their accounts. This was also confirmed by another study done in 2006 by Florencio et al. [118]. Here, the authors assessed the average number of passwords and account users have and conducted a large-scale study over 3 months to understand how many passwords users type per day, how often passwords are shared across sites, and how often users forget passwords. Findings show that on average participants have 6.5 passwords, each of which is shared across 3.9 different sites. In 2011, Campbell et al. [65] investigated the impact of imposing restrictive password composition rules on password choices made by users, such as requiring a minimum number of special or upper and lower case characters. They found that imposing password policies had a positive effect on password reuse, i.e. fewer people reused passwords if policies were enforced. The same was confirmed by Abbott et al., [1] in a study involving several US Universities. They found that stricter password policies led to a lower rate of reused passwords.

Researchers looked at users' behavior when registering and using passwords. Shay et al. [330] show that more than half of participants modify an old password or reuse a password when signing up. Von Zezschwitz et al. [376] found through user interviews that 45% of users reuse the exact passwords. Hanamsagar et al. [139] found that

[4] Google Survey: `https://services.google.com/fh/files/blogs/google_security_infographic.pdf`

after registration, participants 98% of the time reused the same passwords and in 2% of cases modified them. Data was collected using a Chrome extension, capturing passwords upon each attempt. Reusing passwords can become a considerable threat for users as attackers get access to the server on which the password or a hash thereof is stored. As a result, attackers may use this information to impersonate the user for getting access to another account [139]. Prior work has investigated approaches to address this from a system perspective. For example, Das et al. [91] show how client-side password hashing can be used to generate unique passwords for different websites, thus helping mitigate the risk of password reuse. In addition, some systems enforce that passwords are not used beyond a certain time span, require minimum password length, or do not accept a password containing a substring of a blacklisted password [329]. In the same direction, Seitz et al. suggested using dynamic password policies which adjust the password policy if a system detects a password that could be widely used [325]. Another countermeasure for password reuse is two- or multi-factor authentication. These solutions accept that passwords have weaknesses and try to mitigate this by requiring users to perform additional forms of authentication (e.g., entering a TAN). However, this comes at the expense of additional effort each time the user seeks to access an account.

Chapter 3

Related Work: Gaze in Security and Privacy Applications

In this chapter, we reflect on the overall use of users' eye gaze in the security domain. In the following chapters, specific related work will be discussed that directly links to the prototypes and studies presented. This chapter answers **RQ1:** How can security mechanisms benefit from users' eye gaze? by conducting a holistic literature review for eye gaze usage in security and privacy over the course of 25 years.

This chapter is partially based on the following publication:

- C. Katsini, Y. Abdrabou, G. E. Raptis, M. Khamis, and F. Alt. *The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions*, page 1–21. Association for Computing Machinery, New York, NY, USA, 2020

One of the significant application areas of eye tracking and gaze-based interaction is security. The security community is an early adopter of eye tracking. Security researchers have explored the use of eye tracking for biometric authentication [43, 179, 248, 247] and password entry [150, 220] since the early 2000s. Twenty years later, eye-tracking algorithms and technologies have matured significantly. Recent advances in visual computing, gaze estimation algorithms, cameras, and processing power of computing devices have led to eye tracking being no longer constrained to desktop computers but being also available on head-mounted displays [17, 122, 129, 149, 239], handheld mobile devices [187], and public displays [396]. Today, laptops such as Alienware 17 R4 and Acer Predator 21 X come with integrated eye trackers, and smartphones such as the iPhone X and Huawei Mate 30 Pro are equipped with front-facing depth cameras, capable of accurate gaze estimation. Eye trackers are also now used in driver monitoring systems in BMW [50] and Volvo cars [373].

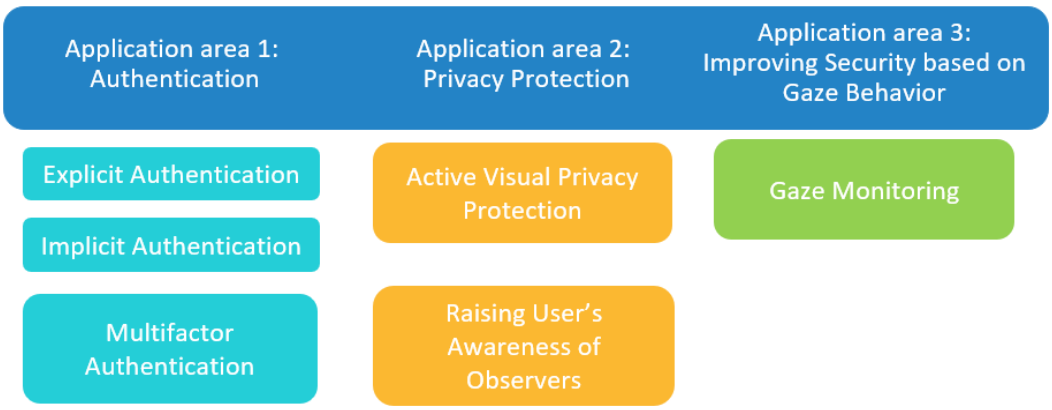


Figure 3.1: Overview of the Different application Areas of Eye Tracking In Security and privacy

These are important developments for security applications. The ubiquity of eye tracking means that researchers can finally take their gaze-based security applications to the real world. The benefits of large-scale eye tracking include: 1) higher adoption of gaze-based security applications (e.g., gaze-based privacy protection or gaze-based authentication), which in turn leads to a better understanding of their effectiveness and performance in daily scenarios, 2) allowing more gaze data to be collected that can be used to improve existing approaches (e.g., to improve the accuracy of biometric authentication), and 3) the ability to unobtrusively understand users through their gaze behavior during security critical tasks (e.g., understanding gaze behavior when subject to phishing attacks).

In this chapter, we 1) survey and organize knowledge in this field based on previous work in the area, and 2) cluster existing work into three main application areas: gaze-based authentication, gaze-based privacy protection, and gaze monitoring during security tasks. Figure 3.1 shows the different application areas of eye tracking in the security and privacy domain.

3.1 Gaze-based Authentication

Gaze has many advantages in the context of authentication. Namely, eye movements can be subtle and hard to notice, making gaze attractive for observation-resilient and high-entropy authentication. These reasons encouraged researchers to investigate ways to leverage gaze for explicit and implicit authentication. We summarize three lines of work: 1) explicit gaze-based authentication, 2) implicit gaze-based authentication, and 3) gaze-supported multi-factor authentication.

3.1.1 Explicit Gaze-based Authentication

Explicit gaze-based authentication refers to the use of eye movements to explicitly verify identity. In this type of authentication, the user has to first define a password that involves consciously performing certain eye movements (step 1: password creation). The user then authenticates by recalling these eye movements and providing them as input (step 2: password recall). The system detects the eye movements and compares them to the password defined in step 1 to verify the user's identity.

Researchers have explored a wide variety of eye movements that could be used for authentication. This includes fixations [220], gestures [98, 99], and smooth pursuit eye movements [22, 84, 101, 370]. There are two dimensions to consider in the use of gaze for explicit authentication: a) password type: legacy vs gaze-based password symbols, and b) used modalities: unimodal vs multimodal gaze-based authentication.

Legacy vs Gaze-based Password Symbols The first dimension refers to the type of password. Passwords in gaze-based explicit authentication can have two forms: 1) gaze can be used as a modality for entering *legacy passwords* (e.g., PINs, text passwords, or graphical passwords), or 2) gaze can be used to enter a *gaze-based password* where the password's symbols are made of eye movements (e.g., a password that involves gazing to the left, then gazing to the right).

Legacy Passwords Each password consists of a series of symbols. Traditional systems have used PINs and passwords (i.e., a series of digits and alphanumeric characters, respectively). Digits and alphanumeric characters are examples of legacy symbols that were argued to have been superseded, but it is difficult to replace them because of their wide use as they are easy to implement and easy to reset. Examples of systems that use legacy password symbols include banks and online websites. Gaze can support entering legacy symbols by providing a certain mapping between gaze behaviors and certain symbols. There are many examples of schemes employing gaze to enter legacy passwords. For example, Kumar et al. [220] proposed one of the first gaze-based authentication schemes where users fixated characters on an on-screen keyboard and then pressed the space button to select them. The same scheme was used on ATMs by Seetharama et al. [323]. Similar work was also done by Kasprowski et al. [175] who used gaze for pointing at PINs and confirmed selection by pressing a key. EyePassShapes uses eye movements to enter alphanumeric passwords [95].

Another body of work focused on using gaze to enter PINs. EyeDent [381] allows users to authenticate on desktops by entering 4-digit PINs using eye gaze. Users do not dwell or press triggers. Instead, the system automatically clusters the gaze points to estimate which targets the user intended to select. PathWord [22] is another system where users enter 4-digit PINs by performing smooth pursuit eye movements that follow the trajectory of the respective digits. GazeTouchPIN [192] allows users to enter 4-digit PINs on mobile devices using

touch and gaze input. Liu et al. [238] explored using gaze gestures to enter 4-digit PINs on smartphones. Best and Duchowski [45] proposed using gaze to enter PINs on a rotary dial interface.

Several works used smooth pursuit eye movements to allow users to enter 4-digit PINs on a public display [84, 196, 193]. Other researchers explored the augmentation of graphical password schemes by using gaze input. For example, in the work of Forget et al. [119, 120], Bulling et al. [63] and others [30, 204, 356], users fixated points on a shown image, using their gaze as an alternative to clicking with the mouse. Similarly, George et al. [129] used gaze to input 3D graphical passwords in VR. Another authentication scheme that has been extended using gaze input is PassFaces [301]; several works [105, 150] extended PassFaces to allow gaze-based selection based on fixations. In EyeSec [231], the authors propose using gaze for input on multiple existing systems, including PIN pads and Patterns.

The advantage of using gaze to enter legacy password symbols is that they can easily integrate with existing backends. For example, to employ EyePIN [98] at an ATM that accepts 4-digit PINs, all that is needed is a camera to track the user's eyes. However, the disadvantage is that the schemes might induce additional cognitive load on the user in order for them to understand the mapping between their gaze and the resulting symbol. Furthermore, most of these schemes are significantly slower to use compared to traditional, less secure alternatives. For example, GazeTouchPIN [192] and EyePassword [220] requires 10.8 and 9.2 seconds respectively to authenticate, while classical PINs require 1-1.5 seconds only [375].

Recently, several works were conducted to authenticate users in VR. For example, Khamis et al. [193] used smooth pursuit eye movements to allows users to authenticate on virtual ATM machine in a VR setup. Similarly, Mathis et al. [254] compared between using gaze, tap and head position for PIN entry in VR environments. Results show that although gaze was more secure it had the longest entry duration. Similar approach was also investigated by George et al. [128] in VR environments. Also work done by LaRubbio et al. [224] compared between dot authentication and gaze monitoring authentication in VR environments showing that gaze monitoring authentication had better authentication accuracy.

Gaze-based Passwords Gaze-based passwords are based on gaze behavior. These schemes transform the password space and, hence, are likely to have a different impact on memorability. Examples include GazeTouchPass [188] and GTmoPass [191], where gaze gestures constitute part of the password. Similarly, in EyePass [99] and another work by De Luca et al. [98], the password consists of a series of gaze gestures. In DyGazePass [295, 296], the user's input is a series of smooth pursuit movements that are supported by cues in the form of 2D geometric objects. In EGBP [319], users authenticate by gazing in one of

four directions and then performing a blink to confirm input. Similarly, Pushya et al. [85] used Morse Code blinks o authenticate users on desktop machines.

The advantage of gaze-based passwords is that they expand the password space by incorporating new sets of password symbols. Unlike legacy passwords where PINs, alphanumeric, and other widely used passwords are entered by gaze, gaze-based passwords might have a steeper learning curve as users are not accustomed to them, and require in-depth analysis of memorability as users would be required to learn and memorize unfamiliar password symbols which could be less intuitive.

Unimodal vs Multimodal Gaze-based Authentication Gaze has been used as the sole input method when authenticating. We refer to that type as unimodal gaze-based authentication. On the other hand, multimodal gaze-based authentication is when gaze is combined with other modalities.

Unimodal Gaze-based Authentication The advantage of unimodal gaze-based schemes is that they are handsfree. This makes them particularly useful for interfaces that are physically unreachable (e.g., displays behind glass windows [93]), for users with motor disabilities, and for touchless hygienic interactions. The disadvantage, however, is that unimodal gaze-based interfaces need a mechanism to distinguish input and perception. That is, the system needs to detect whether a user is gazing at a target to select it, or merely to perceive it. This has been traditionally addressed by introducing a dwell duration, i.e., the user has to gaze at the target continuously for a brief predefined period of time in order to select it [166]. This method has been adopted by several unimodal authentication schemes [63, 98, 120, 28, 389]. An alternative is to detect certain gaze behaviors that would indicate input, such as gaze gestures [102] as done in EyePassShapes [95], or smooth pursuit eye movements [84, 196, 295, 296, 370] or saccadic eye movements [47]

Multimodal Gaze-based Authentication In multimodal schemes, gaze is used a) as a pointing mechanism while another modality is used for selection (we refer to this as *gaze-supported multimodal authentication*); or b) alongside a second modality to improve resistance to observation attacks by *splitting the observer's attention*. The advantage of the former type is that, opposed to unimodal gaze input, the system can clearly distinguish input from perception as the user would use the second modality to confirm the intention to select the target being gazed at. Examples include EyePassword [220], GazeTouchPIN [192] and others [175], where users select each password symbol in two steps. Each step involves one of the modalities. In the latter type, gaze and other modalities are used together for improved protection against observation attacks. For example, users of GazeTouchPass [188], GazeGestureAuth [6] and GTmoPass [191] enter a series of gaze input alongside either touch input [188, 191] or mid-air gesture

[6]. This requires shoulder surfers to observe two elements which in turn reduces attack success rates. Recently, Kumar et al. [218] introduced *PassWalk* which uses users' gaze and lateral shifts for authentication in head mounted displays. They found that PassWalk entails a significantly smaller workload on the user than the current commercial methods. The use of gaze and foot was also used by Rajanna et al. [294], where they compared multimodal authentication using gaze and foot versus touch input where the results showed that touch outperformed the gaze condition. The general disadvantage of multimodal authentication is that it usually complicates password entry. This could influence the memorability of the password symbols.

How to Evaluate Explicit Gaze-based Authentication Users will, in general, not adopt a secure mechanism that is complicated to use and will find workarounds that reduce security (e.g., write down passwords). Therefore, it is important to make sure new schemes are both usable and secure.

Usability and Memorability Evaluation Usability studies often include participants entering passwords using the new scheme, comparing it to a baseline – usually a state-of-the-art scheme. The user's task is to enter a password provided by the experimenter. The password could be verbally communicated to the user [192] or read by a text-to-speech system [196]. Requiring the participant to read passwords from a piece of paper or a screen [195] is not recommended for gaze-based authentication schemes as it might impact the tracked gaze behavior. Error rates are often measured by detecting input failures. A less used approach, albeit equally important, is to present users with incorrect inputs and ask them to correct them [195]. This provides insights on how recognizable errors are, and how easy, fast, and accurately users can correct them. Memorability is often evaluated by querying participants after a period of time to understand whether they remember a) how to use the scheme, and b) their secret. Note that memorability is important to consider regardless of whether legacy or gaze-based password symbols are used. Even if users are entering the commonly used 4-digit PINs, the input method impacts the user's memory [95]. This underlines the importance of understanding memorability of proposed schemes.

Security Evaluation Security studies of gaze-based input often focused on side channel attacks that can be performed by bystanders or co-located adversaries. Examples of studied attacks include observation attacks (e.g., shoulder surfing) [113] and video attacks [392]. In security studies that investigate shoulder surfing resistance, the user is often recorded while authenticating from the best observation angle possible. The recorded videos are then presented to a different set of participants who have been trained to attack passwords entered via the respective authentication scheme. For example, to evaluate GazeTouchPass

[188], the experimenters recorded three videos: a) a video showing the user's eyes to simulate an attacker observing the user's eyes, b) a video showing the user's screen to simulate shoulder surfing the touch input, and finally c) a video from an angle that allows observing both the gaze and touch input. Guessing attacks, which help understanding how likely an attacker will find a password by random or smart guesses, are also important to investigate. However, the lack of knowledge on how users select their gaze-based passwords (Research Direction 1) means that there are no established strategies that attackers use to make informed guesses.

Evaluation Metrics Regarding the evaluation metrics, usability often entails efficiency (i.e., entry time) and efficacy (i.e., error rate due to system malfunction or user errors). Other aspects that can be considered include error tolerance (i.e., how likely is it that users perform errors), learnability (i.e., how easy/fast users can learn how to use the scheme), and user preference. In memorability studies, the metrics are often the recall rate and the recall accuracy. The latter is a measure of similarity of how close the user's recalled password is to the actual password.

Metrics that are commonly used in security studies are: a) attack success rate and b) attack accuracy. The former refers to whether or not attacks were successful while the latter measures how similar the attacker's guess is to the actual password. These values are measured under a threat model that simulates an attack scenario. Suitable threat models should be employed when evaluating the security of authentication schemes. For example, previous work evaluated multimodal authentication schemes against two observers [189] and by using video recordings from two cameras [188]. Commonly studied threats include shoulder surfing attacks [113], video attacks [95], thermal attacks [2], and smudge attacks [32]. While gaze input is not vulnerable to thermal or smudge attacks, multimodal authentication schemes involving touch or tangible input might leak heat or smudge traces that make the scheme vulnerable to said attacks. Studying the aforementioned threat models is important because new input methods do not impact backend security, but rather impact the possible side channel attacks. Further metrics that could be used to evaluate security include the number of guesses required until an attack is successful. Finally, the theoretical password space should also be computed for any new authentication scheme, while the practical password space can be computed through a longitudinal in the wild study to understand what kind of passwords users create.

3.1.2 Implicit Gaze-based Authentication

Implicit Gaze-based Authentication refers to the use of eye movements to implicitly verify identity; it does not require the user to remember a secret, but it is based on inherent unconscious gaze behavior and can occur actively throughout a session. It consists of two steps: the *enrollment* phase during which a digital representation of eye movements is acquired and stored as a template and the *recognition* phase during which the eye movement is tracked, processed and compared to the template to establish the identity of the individual.

Ideally, the tracked eye movements should possess the characteristics of an ideal biometric: universality, uniqueness, permanence and collectability. Research in the field mainly focused on assessing unique eye movements when performing activities with varying visual stimuli and type (e.g., time-dependent). Collectability mainly depends on the context of use (e.g., device capabilities, eye tracking metrics). The permanence of the eye movements has not been fully explored in this context.

Context of Use and Design Perspectives

Identification vs Verification In implicit authentication it is important to distinguish verification (verifying user's identity through a 1:1 comparison) from identification (i.e., discovering the user's identity through a 1:N search) as it affects the authentication performance [335]. In identification scenarios, the more users the system has, the more realistic and more difficult the problem is, and thus, the performance of the system heavily depends on the sample size, the classification models, the device capabilities, and the required resources. Several surveyed works (54) focus on identification (e.g., [41, 66, 83, 90, 177, 264, 267, 305, 316, 368, 152, 31, 234]) while relatively less (21) focus on verification (e.g., [11, 67, 153, 170, 335, 398]), which requires significantly less processing effort.

Eye Tracking Metrics In contrast to explicit gaze-based authentication, eye-tracking metrics for implicit authentication schemes are more diverse, such as gaze entropy [48], fixation density map [230], angular saccade velocity [244], and scan-paths [152]. Researchers, after acquiring gaze data, extend their data sets considering metrics that build upon the fundamental acquired data (e.g., fixation duration, angles velocity) and calculations on them (e.g., mean, maximum, minimum values). The more eye tracking metrics are available for building identification and verification models, the higher the likelihood for improved performance. Several toolkits (e.g., EMDAT, EALab), can be used to process eye-tracking data and generate larger, more inclusive data sets.

The eye-tracking metrics are often complemented with physiological eye metrics [43, 336, 11, 90, 211] or technology-based metrics, such as key-stroke sequences [332] and mouse dynamics [177, 312] aiming to improve the performance of

implicit gaze-based authentication. However, the integration with metrics from multiple and diverse sources introduces a higher level of interdependence and complexity, which could be a barrier when attempting to adopt such implicit authentication schemes in real-life scenarios and everyday tasks.

Device Capabilities The extracted eye-tracking metrics depend on the eye tracker's capabilities, as they are associated with device-dependent specifications, such as operating distance, frequency, and operating window. Considering that eye trackers vary from sophisticated systems to simple embedded cameras, they have varying characteristics that influence the universality, the acceptance, and the performance of gaze-based implicit authentication. Considering that people use multiple devices with diverse characteristics, that issue becomes more intense. Very few research teams have considered the equipment when analyzing and discussing the findings of their studies [109, 208, 371]. For example, Eberz et al. [109] used a down sampling approach to show that different sampling rates affected the quality of eye-tracking metrics. Similar work was reported by others [154, 174].The trade-off between equipment features, the effort of developing sophisticated algorithms for implementing authentication mechanisms depending on sampling data, and processing requirements for using such a scheme in the wild remain unexplored.

Continuous vs Controlled Visual Stimuli Two types of visual stimuli have been used in implicit gaze-based authentication: *controlled* and *continuous*. For controlled visual stimuli, people interrupt other tasks and focus their attention on this stimulus, thus being aware that they are going through an authentication task. Controlled visual stimuli can be either *static* or *dynamic*. Tasks that are based on static stimuli include text-based tasks [41, 43, 153, 290, 306, 332, 127], such as reading a passage excerpt, and static image-based tasks [43, 66, 76, 153, 244, 245, 267, 272, 306, 305, 336], such as the exploration of a photograph. The complexity of the images affects the accuracy of the scheme [394, 336]. Tasks that are based on dynamic stimuli elaborate the goal-oriented visual search approach of individuals as they typically are asked to track dynamic stimuli, such as moving target [17, 39, 43, 153, 170, 179, 176, 209, 212, 246, 270, 312, 334, 335, 336, 371, 398, 397, 399, 316, 127] or video recordings [70, 200, 306, 324, 307], or simiply look at a static target [393].

In contrast, when users authenticate through continuous stimuli, they may not be aware that they are being authenticated, as the stimuli are embedded to everyday tasks, such as reading emails and web browsing [109, 368, 397, 289]. While continuous visual stimuli are of major importance for HCI as they enable unobtrusive authentication, they typically present lower accuracy and it is under-researched field, in comparison to controlled visual stimuli. Research with controlled visual stimuli has focused on refining authentication (e.g., improve accuracy, reduce time) to make implicit gaze-based authentication practical. Understanding the interplay between the diverse factors that influence the controlled-based authentication (e.g., task type, eye

tracking metrics, minimum time) would help to move towards feasible and efficient continuous-based authentication.

Task vs Time as Authentication Factor Time is important for the acceptance of an authentication mechanisms [141]. Hence, implicit authentication should be performed fast. The majority of the surveyed papers is concerned with tasks as a whole (e.g., [51, 43, 67, 177, 272, 305, 92]), meaning that the authentication process starts after the user has performed one or more tasks. Time varies from a few seconds (e.g., less than 10 seconds [83, 179], 30 seconds [177], 40 seconds [305]) to a few minutes (e.g., 5 minutes [43], 25 minutes [200]), with tasks that are based on dynamic visual stimuli being faster. To improve performance, tasks can be repeated several times in the same session [123, 308, 339, 371, 399], resulting in longer duration. Very few works consider time as a factor (e.g., time-based analysis [109]). Five seconds seem to be a significant turning point for achieving good accuracy, with dynamic stimuli outperforming static ones [336, 334].

How to Evaluate Implicit Gaze-based Authentication Implicit gaze-based authentication has been evaluated towards *performance*, *security*, *usability*, and *resources consumption*. *Long-term* evaluation studies have also been conducted.

Performance Evaluation The majority of the works in implicit gaze-based authentication focus on evaluating the proposed schemes towards performance (i.e., efficiency of the classification mechanisms) and explore the efficiency of different features against identification accuracy [83]. Several metrics have been used towards this direction, such as equal error rate – EER (e.g., in [152, 153, 200, 334, 336, 397]), receiver operating characteristic curve – ROC curve (e.g., in [17, 41, 76, 305, 335]), false acceptance rate – FAR (e.g., in [152, 210, 209, 230, 305, 397]), and false rejection rate – FRR (e.g., in [152, 178, 209, 210, 305, 336, 397]). In the authentication domain, reporting only the accuracy of identification or verification algorithms could conceal critical information about the efficiency of the mechanism and raise serious privacy issues. For example, reporting a 90% accuracy suggests 90% of the attempted users were correctly matched, but does not explain whether the remaining 10% were granted access to the system or not. The used evaluation metrics should assess both the probability of false acceptance and that of false rejection. We underline the importance of adopting the respective ISO/IEC standard for biometric evaluation [163]. When evaluating the performance of identification and verification mechanisms, the sample size is key to ensure the reliability of the obtained results. While in literature several suggestions regarding the sample size of eye tracking studies have been made [116, 269], there are no specific guidelines that have been proposed regarding the implicit gaze-based authentication. The number of participants may vary between less that fifty [43, 336], a few hundreds [66, 88, 304], or even thousands [39]. Moreover, several works are based on publicly available

datasets [88, 176, 264, 267, 305, 339, 338, 12] to optimize the identification and verification algorithms. There is to date no gold standard for the sample size when evaluating implicit gaze-based authentication schemes.

Security Evaluation Security evaluation in implicit gaze-based authentication is most often concerned with impersonation [134, 170, 270, 334, 335, 397] and replay attacks [334, 335]. In impersonation attacks, an adversary tries to fraudulently impersonate a legitimate user. Sluganovic et al. [334] simulated this type of attack and showed that increasing the threshold above which a sample is considered legitimate, decreases the likelihood of falsely accepting but at the same time increases the likelihood of falsely rejecting a legitimate user. Success of such attacks is also related to the stimuli complexity [397] and the type of the attackers: internal attackers (i.e. attackers known to the system) and external attackers (i.e. attackers unknown to the system) [335]. Access to the legitimate user's calibration data also affects the success rate of impersonation attacks [270]. In replay attacks, a valid data transmission is maliciously or fraudulently repeated or delayed. To prevent this type of attacks the visual stimulus shown to the user should never be reused. For example, every time the user starts a new authentication session using a moving dot stimulus, the dot should move to different positions and in different order [334, 335].

Usability and Resources Consumption Evaluation Very few works focus on evaluating the implicit gaze-based authentication schemes towards usability, such as time efficiency [336] and user experience [336]. Likewise, resources consumption, such as CPU and memory footprint [336] and energy consumption [397], has received little research interest.

Long Term Evaluation While studying the long-term use is critical for authentication, very few works [17, 213, 270, 336, 398] report such studies. A degradation of accuracy is observed with time regardless of the type of visual stimuli used [213, 336]. To maintain high authentication accuracy, it is necessary to update regularly the owner template. Despite the fact that there is evidence that eye movements change with time [208], the aging factor in not well researched and understood.

3.1.3 Gaze-supported Multi-factor authentication

Multi-factor authentication schemes are those that involve two or more authentication factors. The three most common authentication factors are knowledge, possession, and biometric [274]. Eye gaze can be used for supporting either the knowledge factor, by requiring the user to explicitly move their eyes to demonstrate "knowledge" of the authentication pattern, or it can be used for the biometric factor, by processing the

user's implicit eye movements to verify their identity. The possession factor refers to authenticating a user by showing they "have" a token, a key, a card, or similar.

Knowledge and Biometric To outbalance the accuracy issues associated with the some implicit authentication schemes, two-factor authentication schemes were proposed which combine implicit and explicit mechanisms. In this case, an explicit mechanism is used to enter something the user knows, e.g., a PIN, using gaze input, and implicit metrics are collected and analyzed, e.g., angle kappa, to provide additional proof that the user is who they claim to be. Such examples are presented by [151, 290, 317].

In those examples, gaze was utilized for both the knowledge factor and the biometric factor. It is also possible to use gaze as a biometric factor while another modality is used for the knowledge factor. One example is to verify the user's identity through their gaze behavior while they enter a text password using a keyboard or a touchscreen.

It is also feasible to use gaze for the knowledge factor (e.g., enter a PIN by dwelling at digits on an on-screen keypad) while using a different modality for the biometric factor. In that case, the features used in the biometric factor should be passive ones such as the standing posture, facial features, or gait. Otherwise, requiring the user to authenticate via gaze and perform an additional task to collect biometric data could result in very long authentication times. The only work we are aware of that uses gaze input for the knowledge factor, and a non-gaze feature for the biometric factor is SAFE by Boehm et al. [51] where users gazed at a predefined target (knowledge factor) while facial recognition took place (biometric factor).

Knowledge and Possession Combining explicit gaze-based authentication with a possession factor would also provide an additional layer of security. For example, in GTmoPass [191], the user authenticates at a public display by entering a Gaze-Touch password on their mobile device. Here, the possession factor is the mobile device, while the knowledge factor is the Gaze-Touch password.

Possession and Biometric We are not aware of works that combined the possession factor and biometric gaze-based authentication. One way this could be done is by requiring the user to provide a physical key in addition to engaging them to a visual task and tracking their eye gaze (e.g., while showing a visual stimuli). This would be more secure than using either alone. For example, this could be used when accessing a door.

3.2 Gaze-based Privacy Protection

While authentication protects privacy indirectly by limiting access to confidential content, some approaches aim at directly protecting private content from attackers. Here, gaze can be leveraged in two ways: a) Actively protecting the user's privacy

by, for example, hiding content the user is not looking at, or b) raising the user's awareness of shoulder surfers by detecting the gaze direction of bystanders.

3.2.1 Active Visual Privacy Protection

Eiband et al. [113] showed that while most shoulder surfing resilience research focused on authentication, the vast majority of observed content is text, photos and videos. This means that we need methods to protect the visual privacy of users. Brudy et al. [58] proposed several methods to protect users of public displays from shoulder surfing. The gaze direction of the user and the bystanders were detected using a Kinect device. Privacy protection was done either by moving or hiding content, or by blacking out sensitive content such as personal emails. Ali et al. [20] proposed a slightly similar privacy protection application for detecting bystanders, but they only detected the presence of faces.

Similar systems, like EyeSpot [190] and Private Reader [293], were proposed for privacy protection on mobile devices. In EyeSpot [190], the content that the user is gazing at is visible to them, while the rest is masked either by a black filter overlay, a crystallized mask, or fake content. In the usability analysis of the different filter types, the authors found that the size of the visible spot impacts the reading speed significantly, and that the crystallized filter is more usable compared to the blackout one and fake text in terms of reading speed. The authors found no significant impact of the filters on neither the perceived mental workload nor text comprehension. However, participants favored the crystallized mask as it allowed them to see contextual information such as chat bubbles in chatting apps. Private Reader [293] similarly enhances privacy by rendering the portion of the text that is gazed at. The authors studied the impact on text comprehension and workload, and found that their method reduces comprehension and induces higher workload on attackers compared to the users.

While the aforementioned systems relied on eye gaze to selectively hide or render certain content, other works leveraged the inconspicuous nature of eye gaze to allow privacy-aware interactions. For example, iType [233] allows users to type text on mobile devices using their gaze. Another example is EyeVote [197], which allows users to anonymously vote on public displays without revealing their choices to bystanders. Several systems were proposed for transferring content from public devices to personal ones using eye gaze because it makes it more difficult for bystanders to know which content the user is interested in [250, 358, 359, 360]. While these systems were not built with the aim of privacy protection, privacy-aware interactions were a byproduct of using gaze input.

3.2.2 Raising Awareness of Shoulder Surfers in Real Time

In addition to active privacy protection, Brudy et al. [58] also proposed mechanisms for raising the awareness of public display users about bystanders who might be shoulder surfing them. They experimented with flashing the borders of the display when a bystander gazes at the display while it is in use by someone else, and visualizing the passerby's gaze direction and/or body orientation when it is in use.

Recently Bâce et al. [35] introduced PrivacyScout a novel method that predicts shoulder-surfing risk based on visual features extracted from the observer's face as captured by the front-facing camera. The authors also studied different distances and viewing angles affecting the success of the shoulder surfing attack.

Zhou et al. [401, 402] proposed multiple interfaces that raise the user's awareness of shoulder surfers through visual and auditory notifications. Similarly, Saad et al. [314] proposed different methods to communicate the presence of shoulder surfers to users by using face recognition. Despite not being based on gaze estimation, these works discuss this as a future step for improving the accuracy of detection and increasing the applicability of their concepts.

3.3 Improving Security based on Gaze Behavior

We discussed the use of gaze to support authentication and privacy protection. In addition, tracking the user's gaze behavior can help understand their attitude towards security and detect insecure behavior. A number of attempts to build mechanisms for supporting or improving security have been proposed based on the understanding gained from observing and analyzing user gaze behavior when performing security tasks.

Prior work used eye tracking to study the effectiveness of security indicators on web browsers and the ability of users to detect phishing websites. Here, gaze has not only been used to understand user behavior [29, 89] but also as a mean to improve behavior to prevent such attacks [259]. For example, Arianezhad et al. [29] reported correlations between security expertise and gaze durations at security indicators. While their results highlight the correlation, Miyamoto et al. [259] developed mechanisms to build on this knowledge by proposing a web browser extension preventing users from providing input in web forms until they gazed at the browser's address bar. Steinfeld et al. used eye tracking to explore users‘ attitudes towards privacy policies [343]. They revealed users‘ tendency to read the policy when presented by default, while when given the option to sign their agreement without reading the policy, they tend to skip it. Pfeffel et al. [283] used eye tracking to explore how users decide if an email is phishing email or real.

Rather than for input, some researchers analyzed gaze during authentication with the aim to nudge users towards adopting more secure behavior. Mihajlov et al. [257] explored how much time is spent by users in different registration fields. Similarly, Katsini et al. [182, 184] and Fidas et al. [117] explored where users‘ attention is drawn and how it is associated with graphical password choices. They used this knowledge to design mechanisms that nudge users towards better password decisions [184, 185]. In graphical authentication, eye tracking data has been used for building dictionaries of hot-spots [227], i.e., frequently selected – and thus insecure – positions and for creating cognition-based user models to provide personalized adaptations of authentication schemes [298, 183]. Another work done by Constantinides et al. [79] used gaze monitoring to estimate graphical passwords strength based on image familiarity and user eye movements.

A drawback of continuous gaze monitoring, even if done with the intention of improving security, is that the tracked gaze data can have negative privacy implications. For example, the widespread use of security applications that leverage gaze data can be a gateway for adversaries to spread malware exploiting the user's gaze data for profiling.

3.4 Chapter Summary

In this chapter, we conducted a Systematization of knowledge on gaze-based security applications and classify the utility of gaze into 1) authentication, 2) privacy protection, and 3) understanding gaze behavior in security tasks. We found that there is relatively little research in the area of gaze monitoring to enhance security systems. We also found that gaze behavior reflects cognitive processes, visual attention, and other user attributes [249] which can be used to identify vulnerabilities in security systems and design improved solutions. We also found very little work on gaze behavior and security indicating that this area is under-explored, especially in the field of passwords and social engineering. Moreover, we found that understanding gaze behavior can help improve security. For example, similar to previous work [29, 89, 259], eye tracking can be used to detect whether or not users examined an email's sender to deter users from accessing links in phishing emails. Another approach is to use gaze to detect fear or sense of urgency [3], which are among the emotions social engineers try to instill in phishing and voice phishing attacks [303]. Analyzing gaze behavior can also help improve usability and memorability. For example, the user's pupillary response can reflect if the cognitive load induced by recalling passwords is too high, indicating that the scheme's memorability can be improved. Similarly, frequent scanpaths might indicate confusion, which can in turn indicate that the usability of a system or a task (e.g., installing security updates) should be improved. Our findings from this chapter answer **RQ1:** *How can security mechanisms benefit from users' eye gaze?* and highlights the different research gaps that exist. It also shed the light on the scarce

work conducted on using gaze behavior to enhance security mechanisms and solidifies a good foundation to investigate **RQ2**.

II

Understanding Gaze Behavior During Authentication

Outline

Based on our assessment of the literature in chapter 3, we discovered that eye gaze monitoring is both, a promising, yet under-explored area. We encountered little research focusing on the use of gaze monitoring when users are performing a primary task to detect users' behavior which can help improve their security. To use users' gaze behavior to enhance the security of knowledge-based passwords, first, we need to understand their gaze behavior during authentication. This thesis part answers **RQ2**: *What is the influence of knowledge-based authentication on users' gaze behavior?* by taking a closer look at the following two aspects: 1) motivating the use of gaze by comparing gaze and touch for knowledge-based authentication and 2) investigating the relationship between cognitive load and knowledge-based passwords.

- **Chapter 4** - Gaze Vs. Touch for Knowledge-based Authentication. To understand how different using gaze versus touch to enter Android Lock Patterns and which information is reflected in users' gaze movements, this chapter presents a comparative study between entering Android Lock Patterns with gaze and touch. We look at the different pattern characteristics created by touch and gaze. We investigate where users look when they authenticate, and we highlight the strength of improving existing authentication techniques.
- **Chapter 5** - Cognitive Load and Knowledge-based Authentication. As password meters often require users to comply with password heuristics, this might induce cognitive load on the users. Because pupil diameter has been shown to be a promising indicator of cognitive load, this chapter investigates the effect of password creation on users' cognitive load as reflected in pupil diameter.

Chapter 4

Gaze Vs. Touch for Knowledge-based Authentication

Graphical passwords, such as lock patterns, are a popular means of authentication, especially among Android users [391]. In lock patterns, users authenticate by entering a pattern that connects up to 9 digits on a 3×3 grid. Lock patterns that are entered via touch (often called as TouchLockPatterns), have been extensively studied by the user-centered security community [377, 361, 111, 141]. This resulted in an understanding of how strong the TouchLockPatterns users create, common pitfalls, and areas of improvement. At the same time, advances in gaze estimation accuracy and eye tracking hardware led to gaze gaining popularity for authentication as a more natural and secure modality for entering passwords [181]. Gaze offers usability advantages over traditional modalities such as touch and pointing (e.g., mouse) and its subtleness makes it more secure against observation attacks.

Although gaze was employed for password entry, and showed promising results for entering graphical passwords [63, 95], there is a gap in understanding how users create lock patterns using gaze. For example, do users create stronger lock patterns when using gaze as opposed to touch? Or do they make the same mistakes rendering lock patterns entered via touch vulnerable (e.g., the majority of users create TouchLockPatterns starting on the top left corner, making them more predictable)? This knowledge is important to understand whether adapting lock patterns for gaze authentication is a meaningful approach. In addition, there is a lack of knowledge of where users look while authenticating.

Hence, in this chapter, We study how the use of gaze influences the creation of lock patterns. We provide the first comparison of GazeLockPatterns to the well-studied TouchLockPatterns to understand differences in usability and security on a tablet mobile device in addition to investigating where users look while authenticating.

This section is based on the following publication:

- Y. Abdrabou, K. Pfeuffer, M. Khamis, and F. Alt. Gazelockpatterns: Comparing authentication using gaze and touch for entering lock patterns. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA 2020 Short Papers, New York, NY, USA, 2020. Association for Computing Machinery

Previous work analyzed passwords created by users of Android TouchLockPatterns in the wild. Harbach et al. [141] conducted a month-long field study in which they logged locking-related events on smartphones. Almost half of the participants they surveyed had been using TouchLockPatterns. They found that the average pattern is of length of 5.9 cells, and only 8 of them had set up their device to make the strokes invisible. They found authentication times to increase on average by 147 ms for each additional cell in a pattern. Uellenbeck et al. [361] analyzed the guessability of patterns and showed that users are biased in their choices; users are biased towards starting their patterns from the top left corner and are biased against the center. Von Zezschwitz et al. [377] studied the influence of pattern length, line visibility, number of knight moves, number of overlaps, and number of intersections on observation resistance. They found that line visibility and length are the most important security factors, in addition to pattern complexity. Literature also showed that users often select patterns that are short and constitute simple strokes [361, 27]. Loge et al. [241] showed that age, gender, and experience in IT significantly influence the strength and length of chosen patterns.

The aforementioned work helped shape the user-centered security community's understanding of TouchLockPatterns' usability, and how the way people use them influences security. Work by Katsini and colleagues analyzed touch or mouse-based graphical authentication schemes where passwords consist of a series of pictures [183, 182, 184, 185]. However, none of these works discussed the use of gaze for entering lock patterns. We close this gap in understanding whether the same behaviors observed when creating and using lock patterns pertain to gaze, as opposed to touch.

A recent survey on eye gaze for security and privacy applications [181] showed that gaze is promising for password entry. In addition to usability benefits, gaze is subtle and hard to observe, and can be a powerful means to add biometrics as a layer on top of gaze-based passwords. Gaze can be used for implicit (biometric) authentication [11, 67, 153, 170, 285, 335, 398]) or explicit authentication, e.g., entering a password using gaze [6, 45, 219]. The latter is more relevant to our work: examples include EyePass [99] and others [319, 98] based on gaze gestures, and smooth pursuit eye movement based systems [196, 193, 295, 286, 296]. De Luca et al. [95] proposed EyePassShapes, that extend PassShapes [384], a mouse-based scheme similar to lock patterns. EyePassShapes was not compared with its touch-based counterparts. The slight differences between PassShapes and lock patterns suggest that users' behavior

might be different when using either, thus warranting the need to understand users' behavior for gaze lock patterns.

4.1 GazeLockPatterns: Implementation

We implemented a version of TouchLockPatterns that employs eye gaze for drawing the strokes between digits on a 3×3 grid. The system is based on existing application [78] and was implemented in C#. We used a radius of 34 pixels for the interface buttons and the entry pad area size was 0.4 of the screen width and 0.25 of the screen height. The pad was centered in the middle of the screen. For gaze tracking, we used the raw data stream of a Tobii eye tracker and mapped it to the screen size. The gaze trace feature of Tobii was enabled during the gaze condition as an indication of where the user is looking. The entry pad size and the gaze trace features were enabled as a result of a pilot test with 3 participants. The pattern pad implementation enables users to enter patterns of size 4-9, overlaps were enabled and closed shapes were disabled. To trigger gaze input, users touch on the screen, perform the gesture, and release. We used the same interface for TouchLockPatterns. The input was logged via touch rather than using the eye tracker.

4.2 Evaluation

In this chapter, we investigate *How different gaze authentication is from touch authentication?* and *Where do users look when they authenticate?*. By answering these two questions we better understand users' pattern choices and areas of interest on the screen during authentication.

4.2.1 Study Design

In order to answer our research questions, we designed a between-subjects experiment where half of the participants underwent the touch condition (TouchLockPatterns) while the other half underwent the gaze condition (GazeLockPatterns). There were security level *scenarios* inspired from literature [241]: The participants were asked to create a pattern for a smartphone, an account for online shopping (e.g. Amazon), and a new online banking profile. In this work, we had two independent variables – modality (Gaze vs Touch) and context (smartphone vs shopping vs bank) – and studied their impact on a set of 7 properties of lock patterns as used in literature [377] as dependent variables, *Intersections*, *Overlaps*, *Knight Moves*, *Observation risk*, *Start and End Positions*.

Table 4.1: Comparison between gaze and touch modalities for the 3 situations [7].

	Length		Intersections		Overlaps		Knight Moves		Observation Risk		Memorability		Perceived Strength	
	Touch	Gaze	Touch	Gaze	Touch	Gaze	Touch	Gaze	Touch	Gaze	Touch	Gaze	Touch	Gaze
Smartphone	5.7	5.9	1	3	1	1	2	2	80.25	79.48	4.4	4.2	2.9	2.8
Shopping	6.9	6.6	3	6	1	4	4	1	73.71	77.31	3.9	3.7	3.2	3.2
Bank	7.6	7.2	9	8	5	5	9	4	67.88	72.28	3.4	3.0	3.9	3.5

4.2.2 Apparatus and Participants

We invited 40 users (15F, 1 lefthanded, 2 contact lenses, 14 glasses) between 20 to 55 years (Mean=28.52, SD=9.5) to the experiment. A Microsoft Surface Pro 4 (2736×1824) was used with a Tobii 4C eye tracker.

4.2.3 Procedure

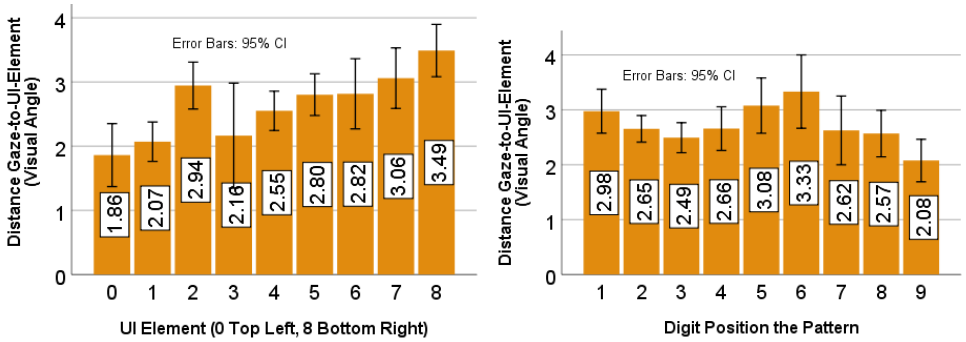
Upon arrival, participants were introduced to the study, filled in a consent form and a demographics questionnaire, then calibrated the eye tracker. Participants then went through three blocks, one per scenario. For each scenario, participants 1) created a lock pattern, then reentered it for confirmation, and 2) rated the strength and the memorability of the entered pattern on a scale from 1 to 5 (5=very strong; highly memorable). At the end participants filled out a questionnaire to share their experience with IT, IT security, lock patterns, the operating system of their smartphone, which authentication schemes they use, and their dominant hand. Those who experienced GazeLockPatterns additionally rated their experience of entering patterns using gaze and reflected on the technique.

4.3 Results

We analyzed the effect of input modality on the aforementioned lock pattern properties for 120 patterns. Results are presented in Table 4.1. Statistical analysis was performed with repeated measures ANOVA and posthoc pairwise comparisons with Bonferroni correction.

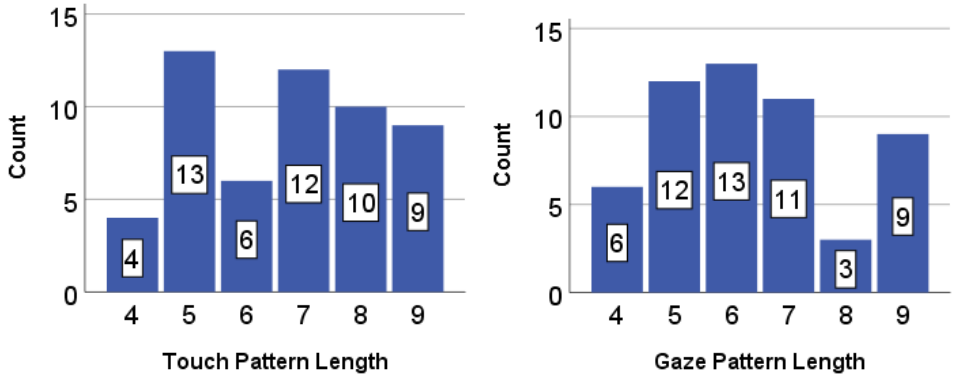
4.3.1 Pattern Length

We subsection analyze the length of the patterns. The distributions of pattern lengths are shown in Figures 4.2a and 4.2b. The mean pattern length across TouchLockPatterns and GazeLockPatterns for each scenario. As seen, there are almost no noticeable differences. No significant differences were found between input modalities ($F(1,53) = 1.308, P = .258$) of gaze ($M = 6.41; SD = 1.654$) and touch



(a) Gaze distance relative to UI 9 digits (b) Gaze distance relative to positions

Figure 4.1: Gaze distance relative to the 9 grid digits (a) and the pattern positions (b), showing how closely users follow their fingers [7].



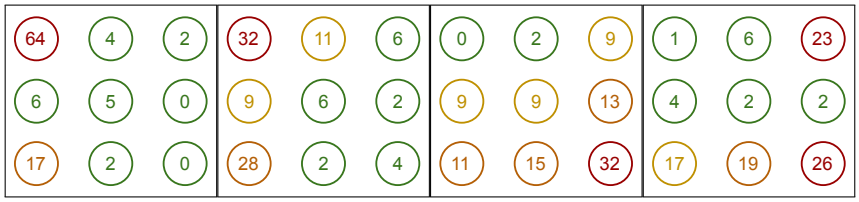
(a) Touch frequency (b) Gaze frequency

Figure 4.2: The frequency of the users' chosen pattern lengths [7].

($M = 6.80; SD = 1.784$). However, factor context revealed significant differences ($F(2,58) = 5.396, P = .014$). Posthoc tests showed the smartphone context (M=5.93; SD=1.337) had a significantly lower length than the bank context ($M = 7.33; SD =$ 2.07).

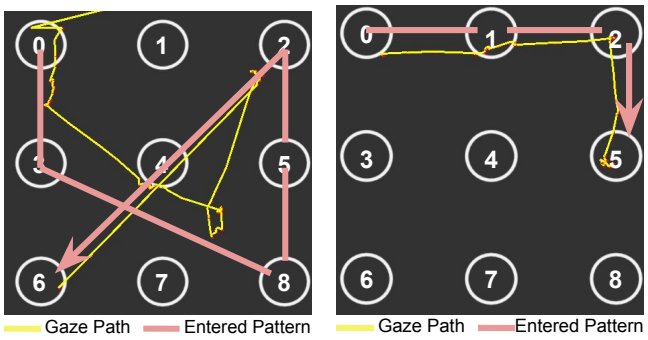
4.3.2 Intersections

Intersections are "strokes which cross already drawn strokes" [374]. Users tend to include more intersections when using gaze than touch. However, the analysis did not reveal a significant difference between gaze (M=.31; SD=.54) and touch ($M =$ $.24; SD = .51$) modality ($F(1,53) = .611, p = .438$). Factor context showed significant differences between smartphone (M=.10; SD=.31), shopping ($M = .27; SD = .450$) and bank ($M = .43; SD = .68$), at $F(2,58) = 3.718, P = .025$. Comparisons showed



(a) Gaze start points (b) Touch start points (c) Gaze end points (d) Touch end points

Figure 4.3: Start and end point distribution for gaze and touch patterns in percentages (a-d) showing top-left / bottom-right trends [7].



(a) Touch pattern (b) Gaze pattern

Figure 4.4: Examples of the gaze path when entering a lock pattern with touch (a) or with gaze (b) [7].

significantly fewer intersections with smartphones than bank context ($P = .047$), potentially as drawing intersecting patterns using gaze is easier than when using touch.

4.3.3 Overlaps

We use overlaps as “crossing over an already activated cell by connecting to a distant cell" [374]. No significant differences were found for the modality between gaze ($M = .19; SD = .39$) and touch ($M = .13; SD = .39$) at $F(1,53) = .525, P = .472$, neither for the context between smartphone ($M = .07; SD = .25$), shopping (M=.13; SD=.34) and bank ($M = 30; SD = .54$) at $F(2,58) = 3.222, P = .097$.

4.3.4 Knight Moves

A knight move “specifies the connection of two distant cells which are not directly neighbored.” [374]. Analysis did not reveal significant difference between gaze ($M = .13; SD = .34$) and touch ($M = .28; SD = .66$) modality ($F(1,53) = 2.180, P =$

.146), neither for the context ($F(2,58) = 1.661, P = .269$), between smartphone ($M = .10; SD = .31$), shopping ($M = .17; SD = .46$) and bank ($M = .33; SD = .71$). It was noticed that users include more knight moves in the bank situation, potentially due to the perceived sensitivity of the situation which requires a stronger password.

4.3.5 Observation Risk

Observation risk is a function of the number of cells, knight moves, overlaps, and intersections [374] that outputs a higher number with higher risk. We discard the parameter "visibility of the strokes" as the strokes were visible in both conditions. We find gaze has a slightly higher mean observation risk than touch, although no significant differences were found between gaze ($M = 76.49; SD = 9.89$) and touch ($M = 74.18; SD = 11.66$) modality ($F(1,53) = 1.345, P = .251$). A significant difference was revealed for smartphone ($M = 79.92; SD = 7.30$), shopping ($M = 74.89; SD = 9.21$) and bank ($M = 70.19; SD = 13.68$) context at $F(2,58) = 7.05, P = .005$. The posthoc test showed that the smartphone context resulted in significantly higher vulnerability than the bank context (p =.004), indicating touch patterns are slightly more secure to observation attacks. However, the nature of gaze input implies that it is robust to observation attacks, indicating that in future work the observation risk equation should be refined.

4.3.6 Memorability

Participants were asked to rate the memorability of the created patterns. Table 4.1 shows participants created similar patterns for both modalities. Analysis did not reveal significant difference between gaze (M=3.62; SD=1.44) and touch ($M = 3.98; SD = 1.15$) modality ($F(1,49) = 2.309, P = .135$). Significant differences were found for smartphone ($M = 4.30; SD = 1.06$), shopping ($M = 3.77; SD = 1.22$) and bank ($M = 3.03; SD = 1.38$), at $F(2,58) = 9.05, P = .001$. Posthoc tests showed, as expected, that smartphone patterns are significantly more memorable than bank ($P = .001$).

4.3.7 Perceived Strength

There is almost no difference between the rated strength of the patterns between gaze and touch (Table 4.1). There are minor differences for the 3 scenarios between gaze ($M = 3.14; SD = .990$) and touch (M=3.30; SD=1.129) modality ($F(1,49) = .635, P = .429$). A significant difference was found for smartphone ($M = 2.77; SD = 1.01$), shopping ($M = 3.20; SD = .89$), and bank ($M = 3.63; SD = 1.16$) context ($F(2,58) = 6.44, P = .002$), showing that bank patterns were perceived significantly stronger than smartphone patterns ($P = .002$).

4.3.8 Start and End Position

The majority of users tend to start their patterns from the top left corner [361]. We also analyzed start and end positions for our patterns. We found that the majority of users (64%) also tend to start their patterns from the top left corner while using gaze. However, for touch 32% of the patterns started from the top left corner and 28% of the patterns started from the bottom left corner. For the end position in the gaze condition, we found that the majority of users (32%) tend to end their patterns with the right-down position. However, for the touch end position 26% of the users ended their pattern with the right down position and 23% ended their patterns with the top right point.

4.3.9 Additional Analysis: Gaze Calibration

Authentication and calibration are both secondary tasks, in the user's way to achieve their goal (e.g., writing a message). Calibration is a tedious task, and conducting the calibration implicitly can improve usability [286]. We investigated whether the gaze data collected during TouchLockPattern authentication can be used to calibrate the eye tracker. The idea is that if the user's TouchLockPattern matches their gaze points, we could draw mappings between eye movements and positions on the screen, and use this data to calibrate the eye tracker.

We plotted the gaze path while entering patterns (Figure 4.4a-f). Users tend to look at the digits they are selecting when entering a TouchLockPattern. We calculated the visual angle of the gaze path for each point on the UI grid (the distance between the center of the UI element and the user's gaze at the moment the finger enters the digit), shown in Figure 4.1a. We found the gaze is close to the digit at the top left corner digits 0, 1, 3 (M=1.45 °), and further away at bottom right corner digits 5, 7, 8 (M=3.1 °).

We then analyzed the gaze path with respect to the digit position in the patterns (e.g., the distance between the gaze at the moment the finger enters the digit, and the center digit position. As the pattern length increases, the gaze data becomes more accurate from 3 to 2 ° (Figure 4.1b). A reason may be, as with longer patterns, difficulty increases, and users visually carefully inspect their touch. Thus, longer patterns can lead to more precise calibration data, which needs to be carefully designed as longer patterns reduce memorability (a Pearson correlation indicated a negative correlation between memorability and length of patterns ($r = -.34, n = 284, p < .001$)).

4.3.10 Discussion

Overall we found no significant difference between gaze and touch modalities, suggesting that findings from studies that investigated TouchLockPatterns are likely to match those on GazeLockPatterns. Users tend to use similar strategies (e.g. length and overlaps) while using the same interface with different modalities. We also found that users' eyes usually follow their hand movements and found that users' gaze while doing TouchLockPatterns in a 2 angles view opens a new design paradigm for future calibration methods which will save users'; time and effort. In addition to its known resistance to the common attack schemes, e.g., shoulder surfing [220], smudge attacks [33], and thermal attacks [2], we find gaze has promising potential as a secure and usable modality for entering lock patterns. Eye tracking is increasingly becoming more available in off-the-shelf devices (e.g., many smartphones come with front-facing depth cameras) [187], which makes this an even more promising time to adopt gaze for authentication. In addition, we found that users' gaze follows their hand movements while entering the TouchLockPatterns which acts as the primary task. We find this to be particularly interesting as it can be used to reflect users' internal state and maybe gives insights on the primary task which can be used for example, in attentive user interfaces. Most importantly, it shows that analyzing users' gaze while performing a primary task gives more insights about the users.

4.4 Chapter Summary

In this chapter, we investigated the difference between how users created touch and gaze patterns for authentication where we recorded gaze data for both conditions. We conducted a between-subjects evaluation ($N = 40$) where we asked the participants to create 3 patterns for 3 different scenarios. We found that as long as the interface is the same, people tend to use the same strategies. We also found that gaze is similar to touch while doing patterns, hence, it can be integrated into existing systems with no additional changes to the interface. We also found that users' gaze follows their touch patterns. By this, we highlight the strength of using gaze for authentication technique. In addition, we highlight that analyzing users' gaze data while they are performing a primary task can be used to reflect users' internal state and might give insights on users' cognitive load according to Kosch et al. [216].

Chapter 5

Cognitive Load and Knowledge-based Authentication

We discovered from the previous chapter that users' primary task affects their gaze behavior. Motivated by Kosch et al. [216], we indicate that analyzing users' gaze data while they are performing a primary task can be used to reflect users' internal state and might give insights into users' cognitive load according. In this chapter, we contribute an investigation of the relationship between perceived knowledge-based password strength and cognitive load and how it affects pupil diameter.

This section is based on the following publication:

- 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 Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21). Association for Computing Machinery.

After more than six decades, passwords remain a ubiquitous approach to authentication. While their end has been repeatedly predicted and other forms of authentication, such as fingerprint, facial recognition, and behavioral biometrics, have gained substantial popularity, we are far from getting rid of passwords anytime soon [26]. The main reason is that passwords currently present a Pareto equilibrium between usability, security, and administrability [54], i.e. there are no other mechanisms providing an equally good trade-off between the effort required for implementation, ease of administration (e.g., reset / changing credentials), ease of use, and security. However, passwords face different issues such as Weak or reused passwords. Bad password usage behavior might lead to unauthorized access to an organization's information

assets. Thus, many organizations enforce frequent password changes to address password leakage [40]. At the same time, research showed that strict password policies decrease employees' productivity [266] and can even result in less security as employees work around rules to easily remember their passwords [365]. This suggests that there is a relationship between cognitive load and password creation. However, this relationship was not studied in depth before. Our work builds on prior research on utilizing eye tracking for cognitive load state estimation and password strength.

There are three types of cognitive load measures introduced in literature: subjective, physiological, and performance measures [273]. Subjective measures reflect the user‘s subjective assessment of cognitive load. The NASA-TLX questionnaire [144] is a frequently used assessment tool for the 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 [42, 159, 161]. Changes in these measures have been shown to correlate with different levels of cognitive load [147, 367]. However, physiological measures depend on many factors, including other aspects of the user‘s cognitive state such as anxiety [71], arousal [180], the user‘s physical activity [311], and environmental variables such as light [302]. Hence, researchers should draw attention to the study conditions and user's state. Finally, performance measures capture 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 [275]. In this chapter, we use the second measure "physiologically" as it is captured without requiring participants to reflect on their performance during password creation nor fill out a questionnaire.

In the last decades, researchers have investigated the pupillary response for different types of tasks [73, 157, 74, 199]. Pupil dilation was found to be higher for more challenging tasks [214, 104]. Not only task demands have been found to influence the pupil diameter, but also factors like anxiety [71], stress [80], and fatigue [352]. A study done by Just and Carpenter [172], 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 [147]. 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. Hence, the need to study the relation between creating passwords and cognitive load is a must. Therefore, in this chapter, we introduce using pupillometry to detect users' cognitive load while creating weak and strong passwords.

5.1 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. In this chapter, we look at how the users' cognitive load changes during weak and strong password creation (**RQ**). Bafna et al. [37] showed that there is an 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 a 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 [386] 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.

5.1.1 Password Strength Meter

We analyzed and compared user-rated password strength against the zxcvbn password strength meter [386] (details in Section 6.4.1). 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 a weak password and from more than 2.5 to 5 is considered a strong password.

5.1.2 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 [203, 16]. The MPDC calculation can be found in Equation 5.1 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 [277]. 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. [42], "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.

$$MPDC = \sum_{i=0}^{N} \frac{MPD_p - MPD_a}{N} \tag{5.1}$$

5.2 Evaluation

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

5.2.1 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, password ratings and gaze data including pupil size as dependent variables. Password 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.

5.2.2 Participants and Apparatus

We invited 15 participants (5 males) to our lab by the university mailing list. The age varied from 22 to 31 ($Mean = 24.27$; $SD = 2.91$). Participants came from different backgrounds (Computer science, Engineering, Landscape Design), and different nationalities (Spain, China, Bangladesh, Pakistan, Egypt, and Germany). Participants had from basic to average knowledge of eye-tracking and none of them had glasses on.

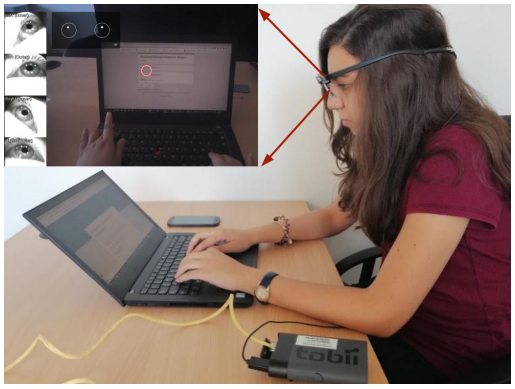


Figure 5.1: Experiment study setup consisting of a laptop and a wearable eye tracker Top Left: gaze monitoring while creating passwords viewed from Tobii pro glasses controller [5].

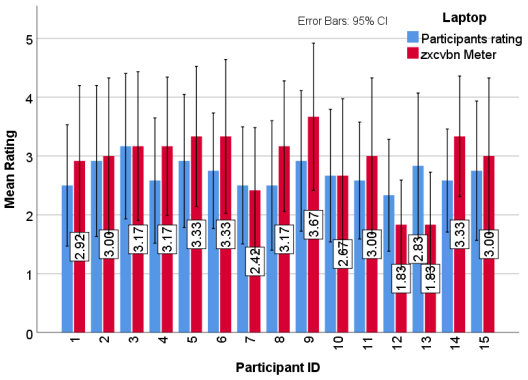


Figure 5.2: Password strength comparison between participants‘ rating and the zxcvbn password meter rating. Showing similar ratings between the zxcvbn meter and users ratings [5].

As shown in Figure 5.1, our experimental setup consisted of a Tobii Pro Glasses 2[5] with 120 fps running on Lenovo T440s[6] along with the Tobii glasses controller[7]. We implemented a simple web page interface where it shows the question and an empty field to write the password in.

5.2.3 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[8]. 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 of more than 8 characters but we did not give any hints on how to create strong passwords nor 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.

[5] Tobii Pro Glasseshttps://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

[6] Lenovo T440shttps://www.lenovo.com/gb/en/laptops/thinkpad/t-series/t440s/

[7] Tobii Glasses Controllerhttps://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/setup/installing-tobii-glasses-controller/

[8] 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/

Overall the study lasted approximately 10 minutes and participants were rewarded with 5 EUR.

5.3 Results

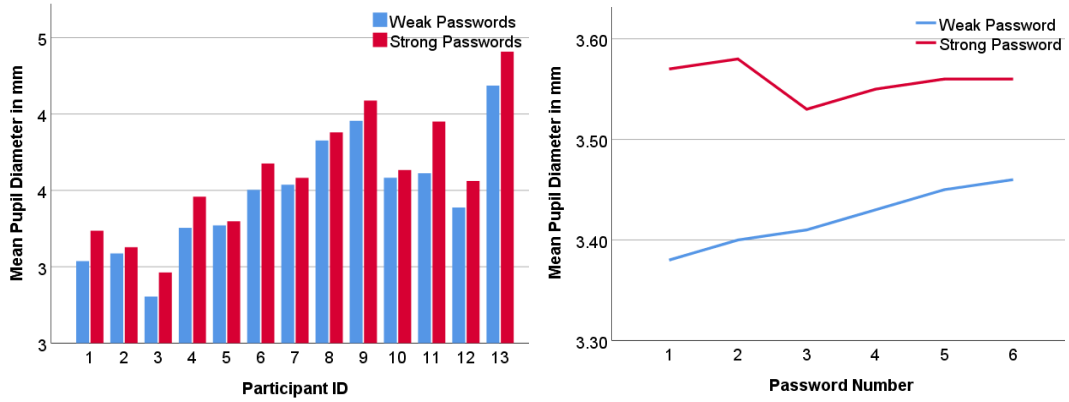


Figure 5.3: (Left) shows the MPD across the 13 participants. (Right) shows the MPD per created password [5].

Data Cleaning and Reprocessing 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 [350]. Therefore, the following analysis is done only on 13 participants.

5.3.1 Rated Password Strength

To get a better idea of how our participants perceived their passwords‘ strength, we compared their rated password strength to the zxcvbn meter password strength. Figure 5.2, shows the average rating for all the passwords entered per participant against the results from the zxcvbn meter. As seen, there is a variance between the passwords ratings, however, the difference between users’ rating and zxcvbn meter rating 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 [112] and that participants‘ perception of weak and strong passwords matches the password meter rating.

Table 5.1: MPD difference between creating strong and weak passwords for all participants [5].

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**

5.3.2 Post Study Question Analysis

At the end of the study, we asked the participants what makes a strong password. Special characters came in the first place (22%), then adding numbers (18%) and upper/lower cases (18%), and finally, 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 [207], the responses still show that participants knew what makes passwords stronger.

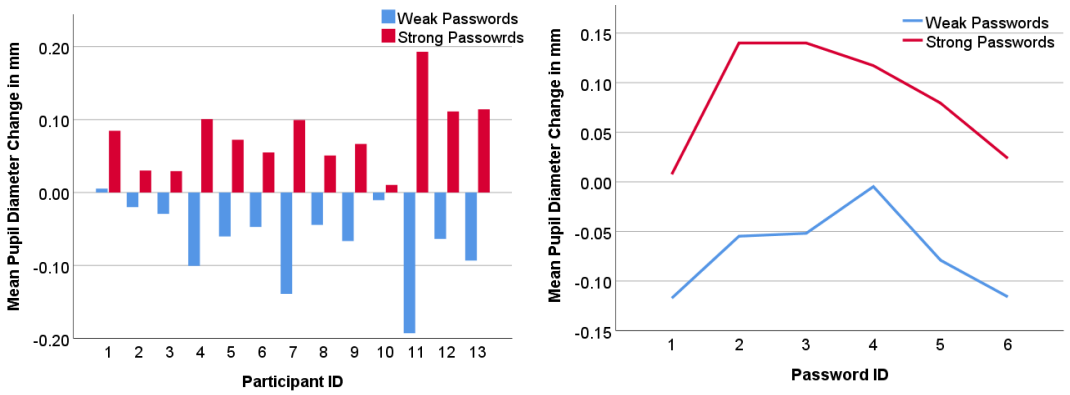


Figure 5.4: (Left) shows the MPDC across the 13 participants. (Right) shows the MPDC per created password [5].

5.3.3 Pupil Diameter and Password Strength

Figure 5.3 left, shows the MPD across the 13 participants. As seen in the figure, the MPD dilates when creating strong passwords than weak passwords except for participants 7 and 11. Repeated measures ANOVA showed a statistically 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 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 5.1) and we found that the mean difference is ($M = .14$, $SD = .09$) and the smallest difference is $M = 0.03mm$. This means that even when we cannot draw a

threshold due to different pupil size responses 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 5.3 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 which 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 5.3 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 a 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 5.4 (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 a statistically significant difference between the MPDC while creating weak ($M = -.06$, $SD = .04$) and strong ($M = .08$, $SD = .06$) passwords across repetitions, ($F_{1,5} = 139.283$, $P < .001$).

5.4 Discussion

Our results suggest that there is a difference in the MPD when creating strong vs weak passwords. Even when we could not draw a threshold due to different pupil size responses 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 5.4 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 which is a sign of higher cognitive load than when creating weak passwords.

5.5 Chapter Summary

In this chapter, we described our approach to infer users' cognitive load based on pupil diameter while creating knowledge-based passwords with different strengths. We hypothesized that creating strong and weak passwords will lead to a change in pupil diameter reflecting the change in cognitive load. We found that creating passwords with different strength leads to changes in pupil diameter, hence, a 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 users' productivity and performance.

In summary, our findings from both chapters illustrate the strength of employing eye gaze to reflect on user behavior and internal state. Furthermore, we revealed that constructing passwords in general imposes a cognitive load on users, particularly when creating strong passwords since users must comply with various password heuristics. This answers **RQ2:** *What is the influence of knowledge-based passwords on users' gaze behavior?*. The next step is to see if this behavior is consistent among users and if it can be modeled and utilized to improve users' security and encourage them to create better passwords.

III

Enhancing Existing Authentication Systems

Outline

Upon identifying where users look during login and how cognitively demanding knowledge-based password creation is, How can we apply this knowledge to enhance existing authentication systems? This thesis part answers **RQ3** *How can users' gaze behavior during authentication be modeled?*. To be able to answer this research question, we identified the two main weak points in password usage: weak and reuse passwords.
This part of the thesis takes a closer look at these two issues in depth:

- **Chapter 6** - Detecting Password Strength from Gaze Behavior. As we discovered that creating strong passwords causes cognitive load on users. In this chapter, we study if this behavior is constant across different participants and whether it can be predicted by using different machine learning classifiers.
- **Chapter 7** - Detecting Password Reuse from Gaze Behavior. This chapter studies how to utilize several machine learning classifiers to predict password reuse by modeling user gaze behavior.

Chapter 6

Detecting Password Strength from Gaze Behavior

From the last thesis part, we know that password creation induces cognitive load, which varies depending on the strength of the password. In this chapter, we compare several machine learning classifiers to see how we can utilize this information to model users' behavior.

This section is based on the following publication:

- Yasmeen Abdrabou, Ahmed Shams, Mohamed Omar Mantawy, Anam Ahmad Khan, Mohamed Khamis, Florian Alt, and Yomna Abdelrahman. 2021. GazeMeter: Exploring the Usage of Gaze Behavior to Enhance Password Assessments. In ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Full Papers). Association for Computing Machinery.

Our work draws from research on password strength meters and gaze behavior in the context of passwords.

As we discussed earlier, the literature showed that security mechanisms can benefit from eye gaze [181]. For example, eye gaze has been used for continuous verification [11, 67, 397] and implicit identification [371, 66, 41]. In 2018, Katsini et al. [184], investigated users' visual behavior and how it affects the strength of the created picture passwords. They used cognitive style theories to interpret their results. They found that users with different cognitive styles followed different patterns of visual behavior, which affected the strength of the created passwords. Furthermore, The authors introduced and studied adaptive characteristics of authentication mechanisms, aiming to assist user groups following different cognitive styles to create more secure

passwords. The results confirmed that adaptive mechanisms based on different cognitive and visual behavior enable new ways of improving password strength in graphical user authentication.

Other work by Katsini et al. [185], studied the feasibility of estimating the strength of user-created graphical passwords based on gaze behavior during password composition. The authors used unique fixations on each area of interest (AOI) and the total fixation duration per AOI. The authors also investigate whether gaze-based entropy is a credible predictor of password strength. Their results revealed a strong positive correlation between password strength and gaze-based entropy. This suggests that the proposed gaze-based metric enables the strength of the password to be predicted in an unobtrusive manner and, thus, help users create stronger passwords. We adopted a similar strategy for detecting password strength from users' gaze. In contrast to prior work we focus on text-based passwords (instead of graphical ones) and we assess password strength as perceived by users (as opposed to password strength as assessed by a system).

As discussed, throughout the years, password meters and heuristics have biased users' choice of passwords and forced them to adopt similar strategies for password creation. This yields a major security risk as most of the users create similar passwords which makes them more vulnerable to attacks. With the ubiquity of eye trackers and by proving that eye gaze behavior can act as a picture password strength meter, we propose adopting the same idea of using eye gaze behavior to estimate text-based passwords' strength. We hypothesize that users' behavior (reflected in the gaze data) while creating a strong password is different than while creating weak passwords and it can be used as a new behavioral aspect.

6.1 Eye Tracking for Password Strength Classification

Previous work showed that security mechanisms can generally strongly benefit from the use of eye gaze data. As previously mentioned, this becomes possible through eye trackers being increasingly available in situations in which security-related tasks (such as authentication) is performed. Hereby, a particular strength of eye tracking is that assistance during security-critical tasks can be provided in an unobtrusive, implicit manner, i.e. a system can make use of gaze data without the need for action from the user.

We investigate a novel application area of using gaze data in security-critical contexts, that is the implicit assessment of password strength as perceived by users. We focus on the distinction between weak and strong passwords with the ultimate goal of

supporting the design of future mechanisms using this knowledge for interventions making users choose stronger passwords.

6.1.1 Password Strength

Password strength can be assessed in different ways. Traditionally the theoretical password space was used to determine password strength, that is the overall number of possible passwords. However, it is today well understood that passwords are not uniformly distributed over the password space, as certain passwords are more likely to be chosen than others (for example, 'password' or '123456'. Hence, researchers today rather consider the practical password space, that is the number of actually used passwords. This password space is generally assessed through empirical studies.

Password strength estimators, such as *zxcvbn* are considering this fact. Strength is determined through the average number of guesses required to identify a password (a so-called guessing attack). The mentioned password estimator, which today serves as a standard way of estimating password strength, classifies passwords into 5 categories: (1) too guessable passwords can be identified through less than 10^3 guesses. (2) very guessable passwords, which protect from throttled online attacks, require about 10^6 guesses. (3) somewhat guessable passwords prevent unthrottled online attacks, requiring on average 10^8 guesses. (4) Safely unguessable passwords provide moderate protection from offline slow-hash attack scenarios (10^{10} guesses). (5) Finally, very unguessable passwords provide strong protection by requiring more than 10^{10} guesses.

We consider weak passwords any password that requires on average below 10^7 guesses, according to zxcvbn. Strong passwords require on average more than 10^7 guesses.

6.1.2 Perceived Password Strength

As laid out in the motivation of our work, a major challenge in usable security research is the mismatch between the password strength as determined by a strength estimator (we refer to this as the *actual password strength* and the strength as perceived by users *perceived password strength*). Figure 6.1 demonstrates this mismatch and its implications.

Optimally, the way users perceive the strength of their passwords would match the actual password strength (i.e., both strong – upper left, both weak – lower right). This would allow them to make a reasonable decision, whether or not their password is appropriate for the type of data they seek to protect. What is now interesting are cases in which the actual and perceived password strength do not match. In the case where the actual password is strong, but perceived as weak by users, no harm would be caused, but it might be worthwhile to explain to users their misconception. More

Table 6.1: Differences between actual password strength and perceived password strength and potential use cases [10].

ACTUAL STRENGTH/PERCEIVED STRENGTH	STRONG	WEAK
STRONG	No action is required	Need to clarify misconceptions/motivate users to choose a stronger password
WEAK	Opportunity to explain misconception to users	Opportunity to make users consider whether password strength is appropriate

problematic is the other case in which the password is perceived as strong by users but is actually weak. In this case, it might be useful to both explain this misconception (and the reasons for it) to the user but additionally support or even require creating stronger passwords.

Our work is meant to particularly identify cases where actual password strength and perceived password strength are at odds. In this way, we enable researchers to come up with interventions that address the respective cases.

6.2 Study

To demonstrate that it is possible to infer perceived password strength from gaze data, we conducted a proof of concept user study. We recorded participants' eye gaze data while creating weak and strong passwords on two input modalities: laptops and touchscreen smartphones. We chose to include different input devices to understand the influence on gaze movements, in particular, or eye movements between the keyboard and the screen.

6.2.1 Design

We applied a repeated-measures design, where all participants experienced all conditions. Participants were asked to enter 24 passwords (6 weak and 6 strong) on both laptop and smartphone. The order of the devices and the password they should create were counterbalanced using a Latin Square. Participants were advised to neither reuse passwords they were already using beforehand nor to reuse passwords they came up with for the study.

6.2.2 Apparatus

The experimental setup consisted of Lenovo T480[9] and Yotaphone[10] as input devices. For the eye tracker, we used the Tobii Pro Glasses[11], connected to a Lenovo T440s[12] using with the Tobii glasses controller[13]. We decided on a wearable eye tracker in order to use the same hardware across all conditions. Also, this allowed us to assess participants' pupil diameter. Deployed systems may rely on integrated cameras such as front facing depth cameras in smartphones [187]. We implemented a simple web page interface showing the task and login interface.

6.2.3 Recruiting and Procedure

We recruited 15 participants (5 males) via University mailing lists. 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). They had basic to average eye-tracking experience. Nobody wore glasses.

After arriving at the lab, participants signed a consent form. Then we explained the purpose of the study. After that, we calibrated the eye tracker using Tobii's one-point calibration. We then asked them to begin creating passwords. After each password, we asked the participants to rate the password's strength on a Likert scale (1=very weak; 5=very strong). After creating passwords on both devices, we interviewed them about what they thought characterizes a strong password. The study lasted approximately 20 minutes. Participants were compensated with 5 EUR.

6.2.4 Limitations

In our study, people did not create passwords to protect real data. Yet, prior research showed that people in such studies still create realistic passwords. We specifically focused on cases where people created new passwords. In reality, password reuse is a common strategy to cope with the issue of having to memorize too many passwords. The effect of this strategy on perceived password strength estimation could be subject to future work. We acknowledge that the sample for our proof-of-concept study was

[9] Lenovo T480: https://www.lenovo.com/us/en/laptops/thinkpad/thinkpad-t-series/ThinkPad-T480/p/22TP2TT4800

[10] Yotaphone: https://www.cect-shop.com/de/yota-yotaphone-3-plus.html

[11] Tobii Pro Glasseshttps://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

[12] Lenovo T440s: https://www.lenovo.com/gb/en/laptops/thinkpad/t-series/t440s/

[13] Tobii Glasses Controller: https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/setup/installing-tobii-glasses-controller/

rather small. At the same time, it is in line with prior studies including password creation tasks [121, 271, 309]. Also, asking people to create multiple passwords still allowed us to collect a data set appropriate for the employed machine learning techniques (cf. the confusion matrix in Figure 6.3).

6.3 Methodology

In this section, we describe the step-by-step process to derive perceived password strength from eye gaze.

6.3.1 Statistical Analysis and Password Strength Estimation

To validate the collected passwords, we analyzed and compared passwords entropy and user-rated password strength against the zxcvbn password strength estimator [386] (details can be found in Section 6.4.1 and 6.4.1). We normalized the zxcvbn password strength estimator score to the range of 1 to 5 and used it to classify passwords into weak and strong. We used a cut-off score of 2.5 for differentiating between weak and strong passwords, i.e. passwords with a score of 1 to 2.5 are considered weak, whereas passwords with a score of 2.5 to 5 are considered strong (cf. section 3).

We also investigated the effect of the input modality on password strength and gaze behavior. We used a repeated-measures ANOVA (with Greenhouse-Geisser correction if sphericity was violated). This was followed by post-hoc pairwise comparisons using Bonferroni-corrected t-tests. Finally, we analyzed the post-study questions.

6.3.2 Feature Extraction

To train the classifiers, we derived a set of seven features that best describe gaze behavior while entering passwords and are commonly used in literature [167, 298]. For extracting the features, we pre-processed the gaze data. First, we removed irrelevant data. As we used a wearable eye tracker, we also collected gaze data focusing on areas beyond the device screen and keyboard. We only considered gaze data inside the AOI (e.g., screen and keyboard) and removed the rest. We then identified fixations using the Dispersion-Threshold Identification algorithm [320]. It produces accurate results in real-time using only two parameters, which are dispersion and duration threshold (set to 25 and 100, respectively). We then extracted seven low-level gaze features from the defined two areas of interest (AOI), keyboard and screen.

Selected main features used for classification are: 1) avg fixation duration, 2) fixation duration, 3) avg saccadic duration, 4) avg left pupil diameter, 5) avg right pupil diameter, 6) screen fixations count, 7) keyboard fixations count.

In addition to those seven main features, we considered the duration spent while typing the password as well as the ratio between the fixations count on the screen and the fixation count on the keyboard. We used thresholding to split the gaze data points between the screen and the keyboard. Differences between AOI are not statistically significant. Hence, we did not take them into further consideration.

6.3.3 Classification Approach

The goal of our classifier is to map a feature vector computed from a window of data to one of the classes corresponding to the password strength (weak vs strong). We implemented two classifiers: a *user-independent*, modality-dependent classifier, trained on the data from different users but using the same input modality, and a *user-dependent*, modality-dependent classifier, again using the same input modality. As different classification models generate different levels of performance, we compared three classifiers with a leave-one-out classification approach: support vector machines (SVM), decision trees, and random forest.

User-independent, Modality-dependent Classifier

We created a user-independent, modality-dependent classifier by training the models on all users for both modalities available (laptop and smartphone). To ensure that the classifiers are performing well on all distributions of data, we split the data into 3 sets: testing, validation, and training. The test set consists of a participant who was not included in the training. The validation set was used for tuning the hyper-parameters of the employed machine learning model. It included data from one randomly selected participant with a specified seed for a participant who has been already included in the training set. The password used in the validation set is also not included in the training set. Finally, the training set included the data of all remaining participants. We used a "leave one participant out" cross-validation. For this purpose, we trained and evaluated the classifier for each modality 15 times and each time for a specific participant.

User-dependent, Modality-dependent Classifier

The goal of building user-dependent and modality-dependent classifiers was to determine if better accuracy could be achieved using a personalized model. The classifier was created once for each participant for each input modality. Again, we separated the data into the three sets mentioned above and we used "leave one observation out" cross-validation. For this, we trained and evaluated the classifier 15 times each, using all features, for each participant.

Table 6.2: Passwords' characteristics for weak and strong (laptop and smartphone). We compare the password length in characters, number of upper/lower case characters, number of digits, symbols, or special characters, and whether the password starts with an upper case letter or ends with a lower case letter. Finally, we show the zxcvbn strength estimator entropy score [10].

		Password Length	Number of upper case characters	Number of lower case characters	Number of digits in the password	Passwords count that start with upper case characters	Password count that ends with digit	Number of symbols in the password	zxcvbn Entropy score
Weak	**Mean**	7.25	0.41	5.02	1.74	0.21	0.39	0.07	14.61
	SD	3.87	0.92	4.05	2.47	0.41	0.49	0.33	3.59
Strong	**Mean**	15.32	2.25	7.4	3.45	0.49	0.37	1.96	60.76
	SD	6.67	2.22	5.47	2.69	0.5	0.49	2.67	9.20
Laptop	**Mean**	11.17	1.13	6.18	2.93	0.27	0.44	0.82	36.88
	SD	6.63	1.77	4.59	2.65	0.45	0.5	1.87	7.01
Phone	**Mean**	11.01	1.44	6.12	2.19	0.4	0.33	1.11	35.75
	SD	6.76	2	5.17	2.74	0.49	0.47	2.26	8.45

6.4 Results

6.4.1 Weak vs Strong Passwords

We collected 366 passwords from all participants in all conditions. In this section, we analyze the passwords collected and report the effect of the different passwords strength on the following:

Passwords Entropy

Table 6.2 shows the characteristics of the weak and strong passwords used for the comparison, as suggested by [112]. We found that passwords perceived as strong by participants were indeed characterized by a high entropy, i.e. they were indeed considered as actually strong by the password strength estimator.

This was also reflected in the statistical tests. An ANOVA reveals a statistically significant difference between the entropy of the weak ($M = 14.45$; $SD = 3.59$) and the strong passwords ($M = 60.75$; $SD = 9.21$), ($F_{1,14} = 268.760$, $P < .001$). An ANOVA did not show a statistically significant effect for the input device laptop ($M = 35.75$; $SD = 8.45$) and smartphone ($M = 36.89$; $SD = 7.02$) on the password entropy generated by the zxcvbn password strength estimator, $P > 0.05$. This suggests that the input modality did not affect the generated passwords' actual strength.

Rated Password Strength

To understand how participants perceive their passwords' strength, we compared the users' rated password strength to the strength as indicated by the zxcvbn strength estimator. Figure 6.1 and 6.2 compares the average rating for all the passwords

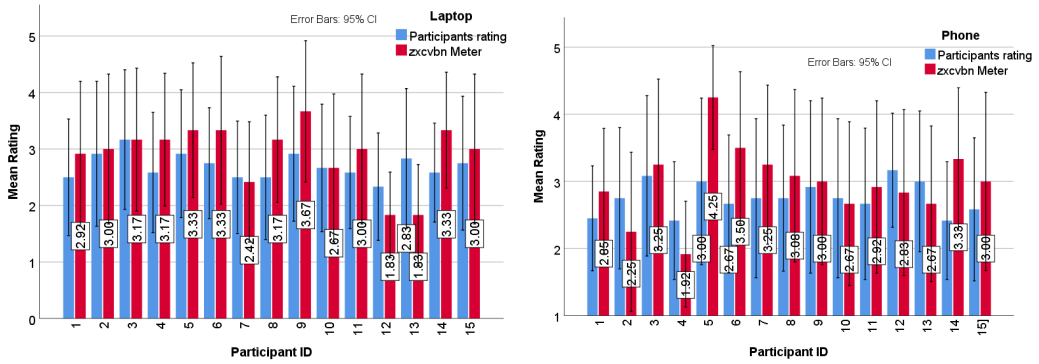


Figure 6.1: Laptop (left) and smartphone (right) strength comparison between participants' rating and the zxcvbn rating. User and zxcvbn meter ratings were similar [10].

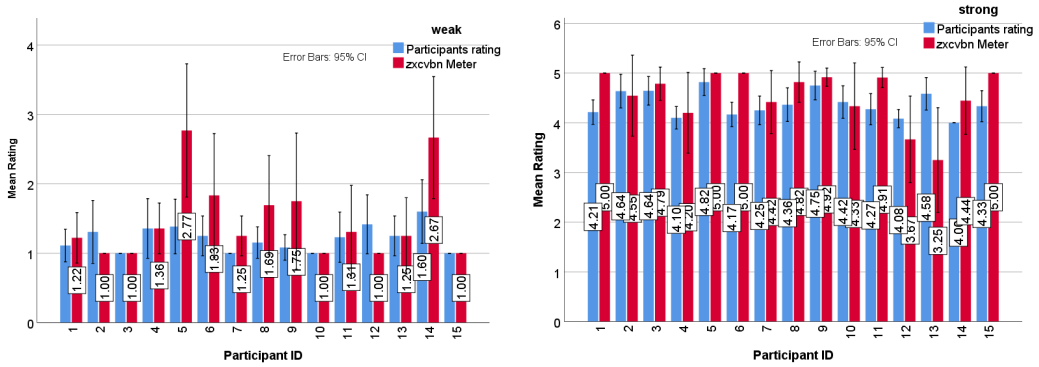


Figure 6.2: Weak (Left) and strong (Right) password strength comparison between participants' rating and the zxcvbn password meter rating. Showing similar ratings between the zxcvbn meter and users' ratings [10].

entered per participant against the results from the zxcvbn password meter. While there is a variance between passwords ratings, the difference between both ratings is not statistically significant – neither for laptop ($\chi^2(1) = 3.769, = .0521$) nor for smartphone ($\chi^2(1) = 1.66, P = .197$), as found by a Friedman test.

The Friedman test also did not reveal a significant effect of the modality on the strength of the weak ($\chi^2(1) = 3.6, P = .058$) and strong ($\chi^2(1) = 3.267, P = .071$) passwords. This suggests there might be no difference between perceived and actual password strength. It also shows that the input modality did not affect the strength of the entered password.

In summary, we found that the input modality did not affect the strength or the entropy of the password. This means that participants entered similar passwords on both input modalities. Additionally, we found a statistically significant difference between weak and strong passwords' entropy and strength which means participants were able

to create a password that was rated as weak and strong by the password strength estimator.

6.4.2 Post Study Questions Analysis

At the end of the study, we asked participants what characteristics make a strong password. Participants named special characters (22%), added numbers (18%) upper/lower case characters (18%), and, finally, increased the length (14%), adding numbers (14%) and adding random characters (14%).

6.4.3 Gaze behavior Statistical Analysis

To assess the relationship between password strength and gaze behavior, we conducted repeated-measures ANOVA.

Effect of Modality on Gaze behavior

We tested the effect of the input modality (laptop vs smartphone) on the gaze features (see Table 6.3). We found that for strong passwords, the input modality has a statistically significant effect on the average fixation duration, fixation duration, average saccadic duration, and keyboard as well as screen fixation count. This means that entering strong passwords on laptops induces shorter fixations, longer saccades, and more fixations on the keyboard as well as fewer fixations on the screen, compared to smartphone. Participants enter longer passwords on laptops than smartphones. In contrast, for weak passwords, the input modality did not have a strong impact on most gaze data, except for the left pupil diameter[14], screen and keyboard fixation count. This means that entering weak passwords on laptops induces less fixation on the screen and more fixations on the keyboard and also smaller pupil dilation than on smartphones.

Effect of Password Strength on Gaze behavior for Input Modalities

To understand the influence of password strength on the gaze features, we ran a repeated-measures ANOVA on the gaze features for both laptops and smartphones. We found that for laptops, entering passwords of different strengths has a significant effect on the average fixation duration, fixation duration, average saccadic duration, average left pupil diameter, screen and keyboard fixation count. In particular, entering weak passwords on laptops induces longer fixation duration, shorter saccadic length, smaller left pupil diameter, fewer fixations on the screen, and more fixations of the

[14] Possibly due to the *dominant eye effect*. We were not able to verify this as we did not assess the participants' dominant eye. We leave this for future work.

Table 6.3: ANOVA results for eye movements, comparing weak and strong passwords across modalities. (significant in bold) [10].

Eye gaze Feature	Strong Passwords	Pairwise Comp. (Bon.Corr.) (Mean; SD)		Weak Passwords	Pairwise Comp. (Bon.Corr.) (Mean; SD)	
	ANOVA ($F(1,14)$; P)	Laptop	Smartphone	ANOVA ($F(1,14)$; P)	Laptop	Smartphone
Avg fixation dur.	$F = 31.012$; $\mathbf{P < .001}$	$.94 \pm .017$	$.96 \pm .015$	$F = .330$; $P > .05$	$.95 \pm .015$	$.95 \pm .023$
Fixation dur.	$F = 31.012$; $\mathbf{P < .001}$	112.70 ± 2.09	114.61 ± 1.75	$F = .290$; $P > .05$	113.56 ± 1.80	113.81 ± 2.78
Avg saccadic dur.	$F = 31.012$; $\mathbf{P < .001}$	$.061 \pm .017$	$.045 \pm .015$	$F = .290$; $P > .05$	$.053 \pm .015$	$.052 \pm .023$
Avg L pupil diameter	$F = 4.039$; $P > .05$	$3.49 \pm .38$	$3.72 \pm .54$	$F = 12.071$; $\mathbf{P = .004}$	$3.39 \pm .38$	$3.69 \pm .55$
Avg R pupil diameter	$F = .095$; $P > .05$	$3.35 \pm .59$	$3.38 \pm .89$	$F = .625$; $P > .05$	$3.25 \pm .64$	$3.35 \pm .89$
Screen fixation count	$F = 6.377$; $\mathbf{P = .024}$	65.87 ± 22.31	87.27 ± 19.54	$F = 9.246$; $\mathbf{P = .009}$	55.37 ± 22.16	84.14 ± 27.67
Keyboard fixation count	$F = 6.377$; $\mathbf{P = .024}$	54.12 ± 22.31	32.72 ± 19.54	$F = 9.246$; $\mathbf{P = .009}$	64.63 ± 27.16	35.85 ± 27.67
Password duration	$F = 2.14$; $P = 165$	13.5 ± 8.8	11.02 ± 4.3	$F = .056$; $P = .817$	6.5 ± 2.9	6.7 ± 2.8

Table 6.4: ANOVA results for eye movements, comparing laptop and smartphone during creating weak and strong passwords [10].

Eye gaze Feature	Laptop	Pairwise Comp. (Bon.Corr.) (Mean; SD)		Smartphone	Pairwise Comp. (Bon.Corr.) (Mean; SD)	
	ANOVA ($F(1,14)$; P)	Strong	Weak	ANOVA ($F(1,14)$; P)	Strong	Weak
Avg fixation dur.	$F = 8.339$; $\mathbf{P = .012}$	$.94 \pm .017$	$.94 \pm .015$	$F = 3.182$; $P > .05$	$.96 \pm .015$	$.95 \pm .23$
Fixation dur.	$F = 8.401$; $\mathbf{P = .012}$	112.68 ± 2.09	113.57 ± 1.80	$F = 3.182$; $P > .05$	114.61 ± 1.75	113.82 ± 2.78
Avg saccadic dur.	$F = 8.401$; $\mathbf{P = .012}$	$.060 \pm .017$	$.053 \pm .015$	$F = 3.182$; $P > .05$	$.045 \pm .015$	$.051 \pm .023$
Avg L pupil diameter	$F = 4.984$; $\mathbf{P = .042}$	$3.50 \pm .38$	$3.39 \pm .37$	$F = .756$; $P > .05$	$3.72 \pm .54$	$3.69 \pm .55$
Avg R pupil diameter	$F = 1.497$; $P > .05$	$3.33 \pm .59$	$3.25 \pm .69$	$F = 1.970$; $P > .05$	$3.38 \pm .89$	$3.35 \pm .89$
Screen fixation count	$F = 6.453$; $\mathbf{P = .024}$	65.87 ± 22.31	55.37 ± 22.16	$F = .847$; $P > .05$	87.28 ± 19.54	84.14 ± 27.68
Keyboard fixation count	$F = 6.453$; $\mathbf{P = .024}$	54.13 ± 22.32	64.63 ± 22.16	$F = .847$; $P > .05$	32.72 ± 19.54	35.86 ± 27.68
Password duration	$F = 12.77$; $\mathbf{P = .003}$	13.5 ± 8.8	6.5 ± 2.9	$F = 25.28$; $\mathbf{P < .001}$	11.02 ± 4.3	6.7 ± 2.8

keyboard. We repeated the same analysis for the smartphone. We did not find a statistically significant effect of the password strength on the gaze behavior (see Table 6.4). A reason for this might be that for smartphones gaze is more strongly affected by the area around the device, which might have had an influence on gaze behavior.

6.4.4 Classifiers Performance

To measure classifier performance, we computed the Area Under the Curve (AUC), as proposed by Abdelrahman et al.[3]. It aggregates precision and recall into one metric. We also investigated the effect of using user-dependent and user-independent classifiers on the classification of passwords' strength for both modalities.

We first compared the performance of the classifiers on 3 models: decision trees, random forests, and SVMs. Each classifier was tuned with its relative hyper-parameters to achieve the best results.

As shown in Table 6.5, the three classifiers resulted in similar AUC, with SVM performing best in most cases. Hence, for the remainder of our analysis, we focus on the SVM results. We found that it is possible to differentiate between strong and weak passwords from users' gaze, independent of the user. The accuracy is 78% for laptops and 76% for smartphones. The user-dependent classifiers outperformed the user-independent for each modality. They achieve an accuracy of 86% on smartphone

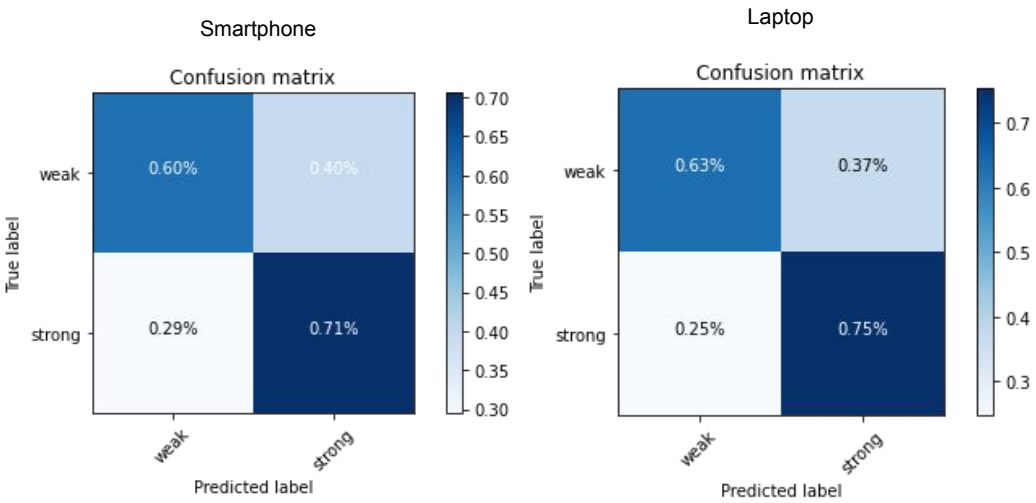


Figure 6.3: Confusion matrix for the user-independent, modality-dependent classifier for mobile (Left) and laptop (Right) [10].

Table 6.5: The AUC of the three Classifiers (Decision trees, Random Forests, and SVMs) for smartphone and laptop. The three classifiers have similar accuracy but SVM performs better in most of the results. The best results are highlighted in bold [10].

	SVM		**Random Forest**		**Decision Tree**	
	Phone	**Laptop**	**Phone**	**Laptop**	**Phone**	**Laptop**
User-indep., Modality-dep.	**.76 ± .19**	.78 ± .16	.76 ± .2	**.79 ± .15**	.70 ± .17	.71 ± .14
User-dep., Modality-dep.	**.86 ± .23**	**.80 ± .29**	.80 ± .25	.71 ± .29	.76 ± .28	.64 ± .29

and 80% on laptops. We report the true positive and true negative rates using the normalized confusion matrix over all participants for each of the user-independent classifier for both modalities in Figure 6.3.

Feature Importance

We used the SHAP [243] algorithm to investigate the importance of features on the performance of the model for classifying weak and strong passwords. The SHAP algorithm explains the output of any machine learning model by computing the contribution of each feature to its prediction. The feature importance graph for the SVM model is shown in Figure 6.4. We observed that for the smartphone modality the fixation count and fixation duration on the smartphone screen and the keyboard are significant in deciding the strength of the password entered by the user. Followed by this, the duration spent while typing the password plays a significant role in the model prediction. For laptops, we observed that the duration has the highest contribution on differentiating between weak and strong passwords followed by the pupil diameter and the fixation count on the screen and keyboard.

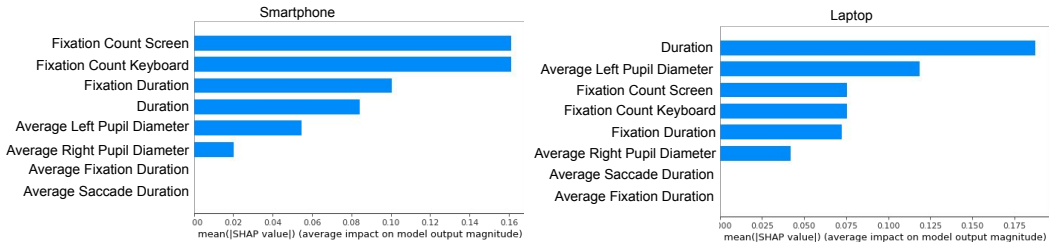


Figure 6.4: Features importance for the user-independent, modality-dependent classifier for smartphone (Left) and laptop (Right) [10].

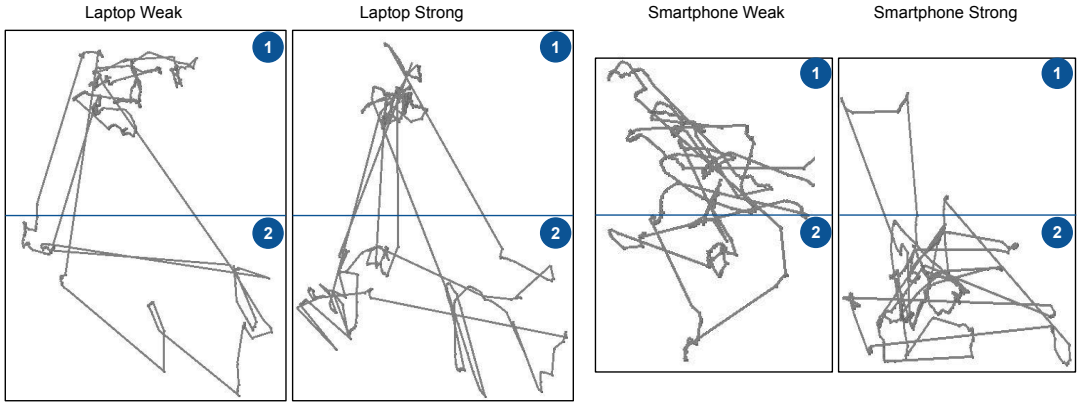


Figure 6.5: Gaze Plots, highlighting behavior while creating weak and strong passwords divided by the areas of interest (1) Screen and (2) Keyboard for both laptops (Left) and smartphones (Right) [10].

Scan Path

Figure 6.5 shows the different gaze plots for one Participant while creating weak and strong password on both modalities. For laptops strong passwords, participants had more fixations on the screen and keyboard compared to during creating weak passwords. For smartphones, participants had more fixations on the keyboard (area 2) in case of strong passwords compared to weak passwords where they had more fixations on the screen (area 1).

6.5 Discussion

Prior work showed that it is possible to assess graphical password strength based on eye gaze. We applied this idea to detect the strength of text-based passwords from users' gaze behavior and provided an in-depth investigation. Here, we summaries and discuss the results grouped by different observations.

6.5.1 Classification Performance

Our results show that password strength classification is feasible, achieving an accuracy of up to 86% when using user-dependent, modality-dependent classifiers. This result is promising as it paves the way for integrating gaze behavior in authentication where perceived password strength plays an important role, e.g., password strength meters.

When comparing the performance with a user-dependent classifier, we observed a decrease in the accuracy to 76% for smartphones and 74% for laptops. This performance might be sufficient for some applications and is substantially better than guessing. Yet, if high accuracy is crucial future systems might want to employ user-dependent classifiers.

Our results show that it is possible to distinguish the strength of text-based passwords by using gaze features and duration spent while typing the password for user-dependent classification. User-independent results were still strong, suggesting that by training the classifier on one specific task, the classification generalizes well to unseen users. This is also confirmed by the statistical analysis of the effect of the input modality and password strength on the gaze features.

It is important to highlight that password characteristics are likely to have an influence on gaze metrics. For example, passwords that include many upper and lower case characters are likely to influence features such as the fixation ratio between keyboard and screen. We only tested with a limited number of users and passwords, so it is likely that such cases are under-represented in our sample.

6.5.2 Features Performance

The feature importance graph for the SVM classifier shows that the fixation count substantially contributes to the classification accuracy. This is more pronounced for the laptop condition. One explanation for this is that people generally entered longer passwords on the laptop, resulting in this being a more suitable feature. We ran a Pearson correlation between the gaze features and password perceived strength. This, however, did not reveal any statistically significant effect of the password perceived length on any of the gaze features. We might simply not have had enough data to reveal such a correlation. Apart from this, for both laptops and smartphones pupil dilation is a strong feature. This can be explained through the increased cognitive load while creating strong passwords. In the literature, it has been proven that higher cognitive load leads to an increase in pupil dilation [104].

6.5.3 Input Modality Effect

According to our analysis, it was more difficult to estimate password strength from users' gaze behavior while entering passwords on smartphones compared to laptops. This can be due to the small screen size, as a result of which gaze movements might be more subtle. Also, the way each user holds the phone is different. Some users prefer to have the phone closer to their face than other. Besides, some participants held the phone in one hand and others held it with two hands. These factors can have affected classification accuracy. On laptops, the distance between the screen and keyboard is larger and, hence, gaze movements are easier to observe. Additionally, we found that participants generate significantly stronger passwords on laptops than on smartphones. This can be due to the different behaviors and reasons behind the use of input modalities. For example, participants might be more used to PINs and lock patterns on smartphones [378, 143]. In contrast, on laptops users are more likely to authenticate using text-based passwords [118].

6.5.4 Influence on Security

Finally, the question arises to which degree the presented approach has an influence on security in general. While our approach is primarily meant to be used by researchers and practitioners to design novel approaches that ultimately lead to stronger passwords, knowledge on password strength in the hand of an attacker might have an adverse effect. For example, if an attacker gets access to an eye tracker, they might find out which users employ weaker passwords of for which accounts they employ weaker passwords, making those a more likely target of an attack.

6.6 Chapter Summary

In this chapter, we introduced a novel approach of using gaze behavior as an additional metric to assess password strength. Our approach assesses users' gaze behavior while creating passwords. We hypothesized that the way in which users create strong and weak passwords is reflected in their gaze behavior. Our results confirmed our hypothesis and showed that it is possible to differentiate between weak and strong passwords with an accuracy of 86% for personalized classifiers on smartphone and 80% on laptops. Our findings pave the way for using gaze behavior in security interfaces, in particular interfaces that make people use stronger passwords. Our findings solve one core pitfall of password usage as an authentication technique, but what about password reuse, does the user's behavior also reflects whether they are reusing passwords or creating new one?. This will be discussed in the next chapter.

Chapter 7

Detecting Password Reuse from Gaze Behavior

Due to the exponential increase of users' online accounts, users have to remember too many and too complex passwords. This caused users to develop coping strategies of which many compromise security. A particularly problematic strategy is the reuse of passwords. One reason is that if a reused password is leaked, attackers can easily gain access to other accounts of the user for which the same password is being used [139].

Having recognized this issue, both researchers and practitioners worked towards solutions. One popular approach is password managers. However, a substantial number of users are hesitant to use such password managers: a recent survey[15] ran by PasswordManager.com and YouGov among 1280 US American citizens showed that almost two thirds of participants do not trust password managers. Furthermore, prior work also showed that password managers not necessarily solve the issue, as a substantial number of password manager users still reuse passwords [281].

Preventing people from reusing passwords is a challenging task for several reasons: First, it requires knowledge about whether or not a user is reusing a password. One approach is *comparing the just created password to a database of known, breached passwords*. Yet, this does not prevent cases in which users are reusing a password that has, so far, not been leaked. Another approach is *comparing all passwords in use by a person* – this becomes possible as people are using a service to centrally manage their passwords (e.g., the aforementioned browser-based or standalone password managers). Such analyses are offered, as part of Google's password checkup[16] or as features of common password managers, such as LastPass's Security Challenge[17].

[15] Password Manager Survey: `https://www.passwordmanager.com/password-manager-trust-survey/`

[16] Google Security Checkup: `https://passwords.google.com/`

[17] LastPass Security Challenge: `http://blog.lastpass.com/2016/06/protecting-lastpass-users-from-password-reuse/`

The drawback, again, is that a substantial number of people are not using password managers and post-hoc alerts on password breaches are often ignored by many users [355]. Furthermore, convincing people to post-hoc change their password is not easy. Prior work showed that even in cases where their passwords were verifiably breached, only 13% of users changed their passwords in the three months following the breach [46].

To overcome the aforementioned issues, we explore a novel approach to detect password reuse based on sensing physiological user information. In particular, we assess users' gaze to infer the reuse of passwords (a) independent of people's password history, (b) without access to the actual password, and (c) already during the password creation process. Our approach is based on the assumption that cognition and behavior are different when reusing or creating a new password. For instance, users might "think harder" about a new password (which would affect fixations) and be required to direct their gaze to the input device more often, due to not having developed a motor memory of the password as a result of frequent use (which would affect the gaze path).

This section is partially based on the following publication:

- Y. Abdrabou, J. Schütte, A. Shams, K. Pfeuffer, D. Buschek, M. Khamis, and F. Alt. "your eyes tell you have used this password before": Identifying password reuse from gaze and keystroke dynamics. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI 2022, New York, NY, USA, 2022. Association for Computing Machinery

Our work draws from prior work on users' password habits and work on typing and gaze behavior in security contexts. Prior research looked into how knowledge of users' behavior can serve to enhance security mechanisms. We will particularly review work on typing and gaze behavior.

Much of prior work on *typing behavior* was motivated by the endeavor of building new authentication mechanisms based on behavioral biometrics. An early example is the work of Monrose et al. [260]. The authors showed that the way people type on a keyboard can be used to identify them. In particular, the authors identified latency between keystrokes, keystroke pressing duration, finger position on the keyboard, and applied pressure on the keys as suitable features to build a classifier, based on which a user can be predicted. Buch et al. [60] looked at how users can be authenticated while writing longer texts, comparing copying text and entering free text. Similarly, Tappert et al. [354] built an authentication system based on free text entry, comparing different lengths entered on both laptop and desktop computers. The results suggest that the keyboard affects the classification accuracy. Hereby, typing on desktop keyboards led to a higher accuracy compared to laptops. Also the keyboard layout was shown to have a strong impact on typing behavior. Researchers compared different keyboards and languages [24, 137, 255, 25].

From prior work we learn that password reuse is still a major challenge in usable security research. There are several reasons for this. Firstly, detecting password reuse is difficult. If a system has access to users' passwords, reused passwords can be detected by comparing them to corpus of leaked passwords or to other passwords of the user. Secondly, when designing concepts for password reuse mitigation, the time of the intervention plays an important role as, when being asked at a later point in time, people are rather unwilling to change their password [139]. We conclude, that being able to know as early as possible that users are about to reuse a password can be valuable when designing mitigation concepts.

Of particular interest is prior research that tried to infer password reuse from keystroke dynamics [168], achieving an accuracy of up to 81.71%. At the same time, prior work showed that the keyboard layout has a considerable influence on accuracy, suggesting that using other modalities might further increase the accuracy and the time at which a reasonable prediction can be made as well as enable novel opportunities for interventions. In addition, prior work has shown that gaze behavior differs between weak and strong graphical and text-based passwords. This led us to assume that reusing passwords might equally be reflected in users' gaze behavior.

Next, we will lay out the concept for using gaze as a means to detect reuse of knoweldge-based passwords and discuss study design considerations. We then present a proof-of-concept implementation and evaluation. To compare our work to prior research, we included detection password reuse from keystroke dynamics as a baseline.

7.1 Focus Group

We conducted a focus group to explore how users' behavior may change depending on the password created, which data to gather, which scenarios to execute, where to conduct the study, and how to design the study interfaces. The focus group output highlighted our study considerations which we will discuss later. The focus group lasted for 90 minutes and it was conducted in the meeting room for the usable security and privacy team in the university of the bundeswehr. We provided colored sticky notes to document the ideas and feedback. The structure of the focus group consisted of an introduction followed by open discussion on the previously mentioned points. We conducted our focus group with 6 expert usable security researchers, we had 2 male and 4 female researchers ($Meanage = 29.3; SD = 6.08$). The focus group was recorded for further analysis and we got the researchers' consent before the recordings.

The focus group outcomes highlight a set of behavioral aspects that reflects users' behavior while creating different passwords (see Figure 7.1), such as cognitive load, entry time, keystroke count, etc. In addition, the focus group highlight that users' behavior might change according to the sensitivity of the protected data on the accounts, suggesting to study users' behavior on both. Furthermore, Last but not least,

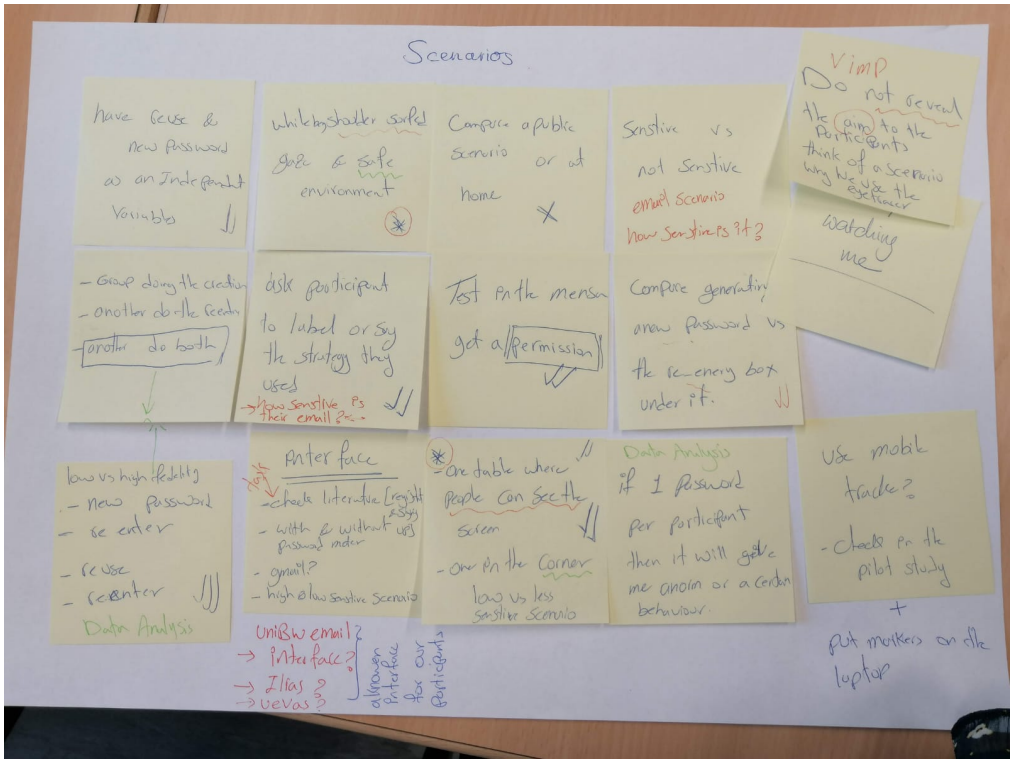


Figure 7.1: Example of the Focus Group Responses When Asked About the Scenarios.

the researchers highlighted the mimicking existing interfaces will increase the validity of the collected data and will induce natural user behavior. Finally, the researchers agreed that approaching users in their environments is better in terms of the induced behavior and quality of the collected passwords compared to lab studies where the controlled setup and the new environment might affect their password choice.

7.2 Concept and Research Questions

In this work, we explore the concept of identifying the reuse of text-based passwords from gaze and typing behavior. The objective of our work is (1) to improve state-of-the-art by showing that the use of gaze can enhance the prediction accuracy, (2) to investigate how the prediction accuracy changes across different phases of the password creation process, and (3) to understand how the sensitivity of the data being protected by the passwords influences the approach.

We first provide background information on eye gaze analysis. Then we explain the different steps of the password creation process. Finally, we present the main research questions driving our work.

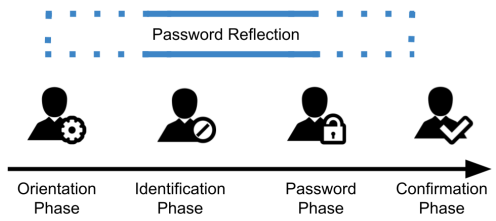


Figure 7.2: Phases of password registration: People first get familiar with the registration interface, then provide their ID and enter the password, and finally confirm their password. In parallel, they reflect on the password [9].

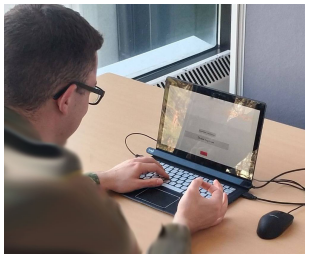


Figure 7.3: Study Setup: Participants were asked to register for two web services on a laptop. We logged keystroke dynamics and gaze using an eye tracker [9].

7.2.1 Gaze Behavior Analysis

Eye tracking research showed that from gaze, information can be derived on the user's state, intentions, and behavior. We explain how, based on different metrics, password reuse might be inferred.

Eye tracking provides information on where the user looks in the form of gaze points (fixations) and the transition between these (saccades). *Fixations* might provide valuable hints as to whether or not people are reusing passwords. The reason is that when reusing passwords, people can likely draw from motor memory (i.e. they know without looking how to enter the password). As a result, one can expect that people reusing a password fixate less on the input device (keyboard fixation count). Furthermore, the need to think about a new password is likely to result in a longer average fixation duration (fixation duration / average fixation duration) similar to literature where Katsini et al. found that users fixate longer while creating strong passwords [184]. Closely related is the *distribution of fixations*. We expect that users might while trying to come up with a new password, differently distribute their gaze on the screen, resulting in longer/shorter saccades (saccadic length / average saccadic length) and in more/less time spent on transitioning between fixations (saccadic duration / average saccadic duration). In addition, we define two areas of interest (AOI): the screen with the authentication interface and the input device (here a keyboard).

7.2.2 Phases of Password Creation

One important aspect of our work is *when* a system could predict password reuse based on gaze data. To investigate this, we decompose the password registration process:

Orientation Phase (O Phase) The authentication process begins with a phase of orientation, where the user is exposed to the authentication interface. During this phase, the user not only gets familiar with the interface, but might already start to think about the password they will use. This phase begins when the authentication interface is displayed, and ends when the user begins to enter their ID.

Identification Phase (ID Phase) In the second phase, the user enters their user ID, which can be a user name or email address. Users might still continue to think about their password while they are already entering their identification information. The phase begins with the first keystroke of the user, as they start entering their ID and ends as the cursor is moved to the password field.

Password Phase (P Phase) In this phase, the user enters the password they thought about. It begins as the cursor is moved into the password field and ends as the user moves the cursor to the password confirmation field.

Confirmation Phase (C Phase) In the final phase, the user re-enters the password. This phase begins as the cursor is moved to the password confirmation field and ends as the user moves the cursor to the register button.

Figure 7.2 depicts the process. Note that users might have different strategies of when they think about the password they want to use. Whereas some users might think about the password already during the orientation phase, others might do so only after they entered their ID. Also, this reflection might span across multiple phases and it could be that users even during the identification phase think about the password.

7.2.3 Research Questions

Previous work utilized keystroke dynamics to detect if a password entered is new or reused [168]. We hypothesize that physiological signals can give a better indication of password reuse. Hence, the first driving research question is: *How well can we predict the reuse passwords from gaze behavior, keystroke dynamics, or both (*RQ1*)?* To do so, we investigate the best gaze and typing features reflecting password reuse.

Second, we expect the sensitivity of data users want to protect to play a role, resulting in the second driving question: *Is password reuse behavior different as passwords protect data with a different degree of sensitivity (*RQ2*)?* Therefore, we compare behavior while creating a password for 1) a webmail client and 2) a customer account for a news website.

7.3 Evaluation and Data Collection

We conducted a data collection study in which we recorded users' gaze and typing behavior while creating passwords for two fictitious accounts, protecting data of different sensitivity. In this chapter, we want to investigate how well can we predict the reuse passwords from gaze behavior, keystroke dynamics, or both ? amd if password reuse behavior different as passwords protect data with a different degree of sensitivity.

7.3.1 Study Design Considerations

Our study design was driven by a number of considerations derived from our focus group, most importantly how to observe natural user behavior, how to preserve privacy by not storing users' passwords, and how to minimize influences from the hardware.

Observing Natural User Behavior Haque et al. [140] showed the sensitivity of the data being protected by a password to have an influence on password choice. Participants create shorter and less secure password when registering a password for a website protecting less sensitive data. As a result, we followed common practice from the literature [5, 10], investigating both cases where users were to chose passwords protecting a web mail account (more sensitive data) and a news website account (less sensitive data).

Password Privacy Our study had two objectives regarding password use: (a) ensuring users chose reasonable passwords they could remember and (b) not storing the actual passwords (which would be necessary for password verification). To address this we only store password characteristics. For example, as users chose A!3, we would store the following information <upper case letter><special character><digit>. We used this information later to verify whether the re-entered password matched those characteristics. The trade-off is that we could not exactly verify the password. However, as this was not the purpose of this approach, we prioritized privacy.

Influence of Hardware Prior work on keystroke dynamics showed that the keyboard hardware has an influence on user behavior [321]. Hence, we decided to collect data from all participants using the same hardware and setup.

7.3.2 Study Design and Apparatus

We designed a within-subjects study with one independent variable (authentication interface), resulting in two conditions: 1) *Webmail Client* – a web-based authentication

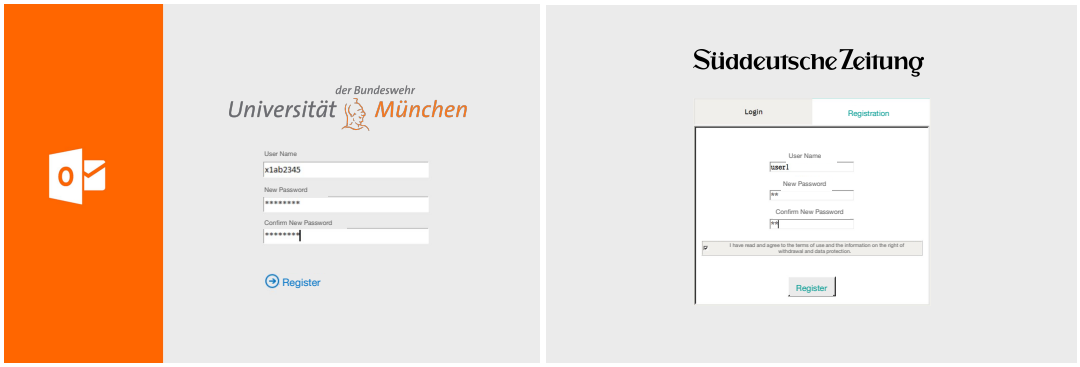


Figure 7.4: Registration interfaces: We rebuilt the webmail registration interface of the local University (left) and of a regional news website (right) to investigate differences in user behavior when creating passwords for accounts with more and less sensitive data [9].

interface, meant to protect sensitive, personal email data. The interface resembled the webmail client of our University. 2) *News Website* – a web-based authentication interface, protecting less sensitive data. The interface resembled the authentication interface of a popular regional news website (see Figure 7.4).

All participants experienced both conditions in a counter-balanced order. We measured 8 dependent variables: duration for the password registration process, gaze metrics, keyboard metrics, time spent on each form field, password characteristics, and perceived password memorability. We did not store the raw password, but instead its length and the characteristics of each character, i.e., whether it was lowercase, uppercase, a number, or a symbol). For the apparatus we used a Lenovo Yoga 900s 12ISK laptop with a 12,5" screen (3200 $\times$ 1800 pixels) and off-the-shelf Tobii 4C eye tracker with a framerate of 90 Hz. We also implemented a demographics questionnaire at the end of the study. The questionnaire had questions about, age, gender, background, profession, experience with eye tracking and experience with IT security.

7.3.3 Study Setting, Procedure and Recruiting

We setup a booth in a quiet area of one of our local university's cafeteria (Figure 7.3). We approached people on campus and asked them to participate in the study. When participants agreed, we went with them to the cafeteria and asked them to sit at the booth. Participants were facing the booth wall to eliminate the influence of people in the vicinity.

We first asked participants to fill in a brief demographic questionnaire and a consent form. They were then told that we conducted a usability test of a slightly updated version of the University's web mail's password registration interface. Hereby, we

specifically told them that the interface was not connected to the actual web mail system of the University. Furthermore, we explained that we compared it to the password registration interface of a regional news website. We also told them that we recorded gaze data to identify issues with the interface. After that, the eye tracker was calibrated using Tobii's 5 point calibration. Participants were asked to register an account for both websites. Participants were told that we did not store their passwords but that they had to remember them as they would be asked to later sign on with them. Participants were then shown the first registration page with three fields – one each for ID, password and password confirmation – and a register button (Figure 7.4). After participants had filled in the ID and passwords, they clicked the register button and were directed to the second interface, following the same procedure. Afterwards, participants were asked for each of the passwords how memorable they thought it was (5-Point Likert scale; 1=not memorable at all; 5=very memorable). Then, they were asked to log into both interfaces again in the order of registration. Finally, we wanted to know from participants whether they reused a password or created a new one. At the end of the study, we explained participants the true objective of the study and asked them to explain their strategy behind creating the passwords. On this occasion we were also able to clarify what password reuse means, if needed.

The experiment took around 10 minutes and participants were compensated with chocolates/treats. The study complied with our university's ethics requirements.

7.3.4 Limitations

We acknowledge the following limitation. Firstly, we cannot verify whether participants truthfully answered the questions regarding password reuse. Participants might have lied about non-compliant, insecure behavior. We tried to minimize any such influence by running the study in a completely anonymized way where no personal information was collected so as to establish trust. Furthermore, the percentage of reused passwords aligns with the literature, suggesting that participants mostly answered in a truthful way. Secondly, while the number of participants is in line with much similar prior work, we acknowledge the rather small size of our sample.

7.4 Feature Extraction and Classification

We describe our step-by-step process to evaluate eye gaze and keystroke dynamics for password reuse detection. First, we analyzed the collected passwords' characteristics and evaluated the effect of password type on password characteristics. Second, we extracted keystroke and gaze features required for classification and tested their statistical significance for the two types of passwords. Third, we built and tested

different classifiers based on these features. We distinguish two categories: new and reused passwords. All features below were extracted for both categories.

7.4.1 Feature Extraction

We extracted a feature set describing keystroke dynamics and gaze behavior from the collected data in addition to password characteristics. We also analyze perceived password memorability.

Password Characteristics

We extracted the following password characteristics: password length, number of upper-case letters, number of lower-case letters, number of digits, and number of symbols. We also tracked the study duration, i.e. time in seconds from when the UI was shown until the 'Register' button was pressed.

Gaze Features

From the collected raw gaze data (X and Y positions on the screen), we derived the following characteristic eye movement features [167, 298] shown in table 7.1. All features are computed and analyzed for each password phase, as well as for overall phases.

Keystroke Dynamics Features

We collected 5 keystroke dynamics, informed by the literature [263, 168, 126] shown in Table 7.2.

Table 7.1: Gaze Features

Gaze Feature	Description
Fixations Count	Number of fixations performed during the task.
Fixation Duration	Time for which users dwelled with their eyes on the laptop screen as well as on the keyboard.
Saccadic Length	Euclidian distance between two consecutive fixations with the eyes, determined in pixel.
Saccadic Duration	Duration between consecutive fixations.
Screen Fixation Count	Number of fixations on screen.
Keyboard Fixation Count	Number of fixations on keyboard.

Table 7.2: Keystroke Features

Keystroke Dynamics	Description
Total Duration	Duration for typing email and password in milliseconds (not considering password confirmation).
Password Typing Duration	Time taken by the participant to enter the password in milliseconds.
Password Keystrokes Count	Number of keystrokes needed to type the passwords (including insertion, deletion).
Flight Time	Average latency between key presses in ms.
Pre-input Time	Time from the moment the interface was shown until the first key was pressed in milliseconds.

7.4.2 Classification Approach

The goal of our classifier is to map a feature vector computed from a time window of data to one of the classes corresponding to the password type (new vs reused). We first built an interface-dependent classifier, accounting for data sensitivity (webmail client vs news website). The classifier is trained on the data from different users but on the same interface. We then built an interface-independent classifier, not accounting for data sensitivity.

We used 3 feature sets: 1) keystroke features + password characteristics, 2) gaze features, and 3) both features combined. Keyboard and gaze data were saved and synchronized using the timestamp.

We compared the performance of three classifiers: Support Vector Machines (SVM), decision trees, and random forest, as done by Abdrabou et al. when detecting password strength [10]. To optimize performance, hyper parameters for each classifier were empirically optimized on a small set of values.

Interface-Dependent Classifier: Webmail Client vs. News Website

To understand how generalizable our approach is across different interfaces, we created interface-dependent classifiers by training the models on all users' data for each of the two interfaces separately. For each of the previously mentioned phases, we created one classifier. We implemented a two-fold cross validation. Figure 7.5 shows the steps for creating the classifier. We start with cleaning the data by removing the data outside our areas of interest (i.e., the screen and keyboard). During the pre-processing we assign the label 'new' or 'reused' to each sample, according to the participants' responses. After that we calculate the features for both gaze and keystroke dynamics. The collected data is synchronized using the timestamp for the analysis. This is followed by assigning the data to the 2 folds and running the classification. These steps are repeated for each phase. At the end, we report the AUC (Area Under the

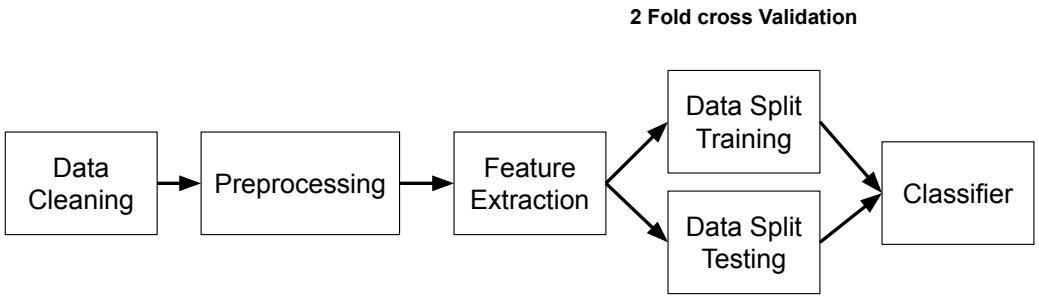


Figure 7.5: ML Classification Steps from data preparation until sending the data to the classifier [9].

Curve) score which measures the ability of a classifier to distinguish between the two classes ('new' and 'reused') and is used as a summary of the ROC curve[18].

Interface-Independent Classifier: Both Interfaces

To understand whether a classifier working for interfaces protecting data of different sensitivity could be built, we created models that were independent of the data to be protected – in our case the web mail and the news page data. To do so, the classifier is trained on the data of all users and both interfaces. We split the data similar to the interface-dependent classifier into a training set and a test set.

7.5 Results

In this section, we present and analyze the collected data.

7.5.1 Participants

A total of 52 participants (10 females) were recruited. The study ran over two weeks. Participants' age varied between 17 and 54 years ($M = 25.27$; $SD = 6.76$). 30 participants were students, 10 academic staff and the remaining 12 administrative staff. Most participants stated to be rather inexperienced with IT security (5-Point Likert scale; 1=no experience at all; 5=strong experience; $M = 2.23$; $SD = .35$). 23 participants wore glasses.

[18] AUC: https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/

Table 7.3: Number of new and reused passwords and task completion time [9].

	Webmail Client		News Website	
	New Passwords	Reused Passwords	New Passwords	Reused Passwords
Number of Passwords	35	14	31	18
Task Completion Time	52.28	37.89	42.07	25.99

7.5.2 Data Pre-Processing and Overview

We removed data from 2 participants due to poor calibration quality. We lost data from one participant due to technical issues while saving. Overall we collected 98 passwords, half of which were created on the news website interface and the other half on the webmail interface. Table 7.3 shows the number of the newly created and reused password for each interface. As can be seen, participants reuse more passwords for the news website than for the webmail client. Participants needed on average 52 seconds to create a new password for the webmail interface and 42 seconds for the news website. In contrast, for the reused passwords, participants needed on average 38 seconds for the webmail interface and 25 seconds for the news website. A Wilcoxon test, revealed statistically significant differences between the study duration for reused and new passwords for the news website ($Z = -2.85, P = .004$) but not for the webmail client ($P > .05$). For both gaze and keystroke data, we sampled data at 90 Hz from the eye tracker and from key input events. This led to an average of 3149 samples per password, resulting in overall 340K samples for all participants for both interfaces.

7.5.3 New vs. Reused Passwords

We analyzed and compared cases where password were newly created or re-used.

Regarding *password memorability*, we found a statistically significant difference between reused ($M = 4.8; SD = .6$) and new passwords' memorability ($M = 3.9; SD = 1.1$) for the webmail client, ($Z = -2.226$,$P = .026$). This show that reused passwords (at least those protecting sensitive data), are more memorable than newly generated ones. Table 7.4 presents characteristics of passwords obtained during the study, and their distribution over conditions.

No statistically significant differences were found between the two interfaces regarding *password characteristics* (password length, number of digits / special characters / upper-case letters).

Table 7.5 summarizes findings regarding *keystroke features*. Our results indicate that participants took more time to think about and type new passwords compared to when

Table 7.4: Wilcoxon signed-rank tests for both new and reuse password features on both interfaces. The results show that there is no statistically significant differences for the password characteristics between new and reuse passwords [9].

Password Characteristics	Email Interface			News Interface			Both Interfaces		
Feature	New Mean	Reuse Mean	Wilcoxon	New Mean	Reuse Mean	Wilcoxon	New Rank	Reuse Rank	Wilcoxon
Password Length	9.5	10.6	Z=-1.517, P>.05	9.5	10.3	Z=-.573, P>.05	10.4	10.3	Z=-1.154, P>.05
Upper-case Letters	1	1.1	Z=-.583, P>.05	.6	.6	Z=-.372, P>.05	0.8	0.8	Z=-.655, P>.05
Digits	3.3	3.2	Z=-.394, P>.05	2	3.3	Z=-1.800, P>.05	3.2	2.7	Z=-.892, P>.05
Symbols	.29	.71	Z=-1.403, P>.05	.3	.1	Z=-1.134, P>.05	0.4	0.3	Z=-.573, P>.05

Table 7.5: Wilcoxon signed-rank tests for keystroke features. For Webmail there is a significant effect of password type on the password typing duration. For the news website the password type had significant effects on flight time and thinking time [9].

Keystroke Feature	Webmail Client			News Website			Both Interfaces		
	New Mean	Reused Mean	Wilcoxon	New Mean	Reused Mean	Wilcoxon	New Mean	Reused Mean	Wilcoxon
Typing Duration	33.7	25.2	Z=-1.664, P >.05	27.5	16.9	Z=-1.764, P >.05	30.8	20.5	**Z=-2.711, P=.007**
Password Keystroke Count	16.5	13	Z=-.345, P >.05	13.6	12.3	Z=-.980, P >.05	15.1	12.6	Z=-.841, P >.05
Password Typing Duration	23	13.7	**Z=-2.103, P=.035**	15.8	10.2	Z=-1.851, P >.05	19.6	11.8	**Z=-3.048, P=.002**
Flight Time	1.7	1.1	Z=-1.852, P >.05	1.3	.9	**Z=-2.025, P=.043**	1.5	1	**Z=-3.160, P=.002**
Thinking Time	14.6	8.5	Z=-1.782, P >.05	7.4	4.2	**Z=-3.027, P=.002**	11.3	6	**Z=-3.586, P<.001**

reusing passwords. This includes shorter times when reusing passwords for pre-input time, typing duration and flight time.

Regarding *eye movement features*, we found several statistically significant differences between new and reused passwords (Table 7.6). The password type has a significant effect on several features for both interfaces. Furthermore, it shows that when considering both interfaces, for the reused passwords, users gaze was characterized by significantly shorter fixation times, shorter saccadic duration, less fixations, shorter saccades and less fixations on both the screen and keyboard. Overall, the many significant differences suggest eye movement features to be well suitable to accurately identify password reuse. We discuss practical implications in Section 8.

Table 7.6: Wilcoxon signed-rank tests for the gaze features. The results show that for both Webmail and the News Website, the password type had a significant effect on serveral gaze features [9].

Gaze Feature	Webmail Client			News Website			Both Interfaces		
	New Mean	Reused Mean	Wilcoxon	New Mean	Reused Mean	Wilcoxon	New Mean	Reused Mean	Wilcoxon
Fixation Duration	28041.9	15728.1	**Z=-2.542, P =.011**	20143.4	13631.6	**Z=-2.330, P =.020**	24497.3	14548.7953	**Z=-3.964, P <.001**
Avg. Fixation Duration	222.8	203.1	Z=-1.66, P >.05	210.9	208.8	Z=-.152, P >.05	219.9	206.2945	**Z=-2.375, P =.018**
Saccadic Duration	20850.4	18704.9	Z=-.471, P >.05	18896.4	10490.1	**Z=-2.199, P =.028**	19988.7	14084.1108	**Z=-2.001, P =.045**
Avg Saccadic Duration	174.6	257	**Z=-2.982, P =.003**	196.6	171.1	Z=-.370, P >.05	186.2	207.3771	**Z=-2.618, P =.009**
Fixation Count	2595.8	1458	**Z=-2.542, P =.011**	1862.1	1265.7	**Z=-2.330, P =.020**	2266.4	1349.7500	**Z=-3.927, P <.001**
Avg. Fixation Count	.6	.5	**Z=-2.982, P =.003**	.6	.6	Z=-1.067, P >.05	.6	.5327	**Z=-3.385, P =.001**
Saccadic Length	1677.9	1539	Z=-.282, P >.05	1436	960.3	**Z=-2.199, P =.028**	1574.5	1213.2813	**Z=-2.094, P =.036**
Avg. Saccadic Length	.4	.5	**Z=-2.982, P =.003**	.4	.4	Z=-1.067, P>.05	.4	.4673	**Z=-3.385, P =.001**
Screen Fixation Count	2149.5	1193.9	**Z=-2.668, P =.008**	1690.7	1122.7	**Z=-2.461, P =.014**	1947.7	1153.8437	**Z=-3.843, P <.001**
Keyboard Fixation Count	446.3	264	Z=-1.915, P >.05	173.2	142.9	Z=-1.918, P >.05	318.8	195.9063	**Z=-2.786, P =.005**

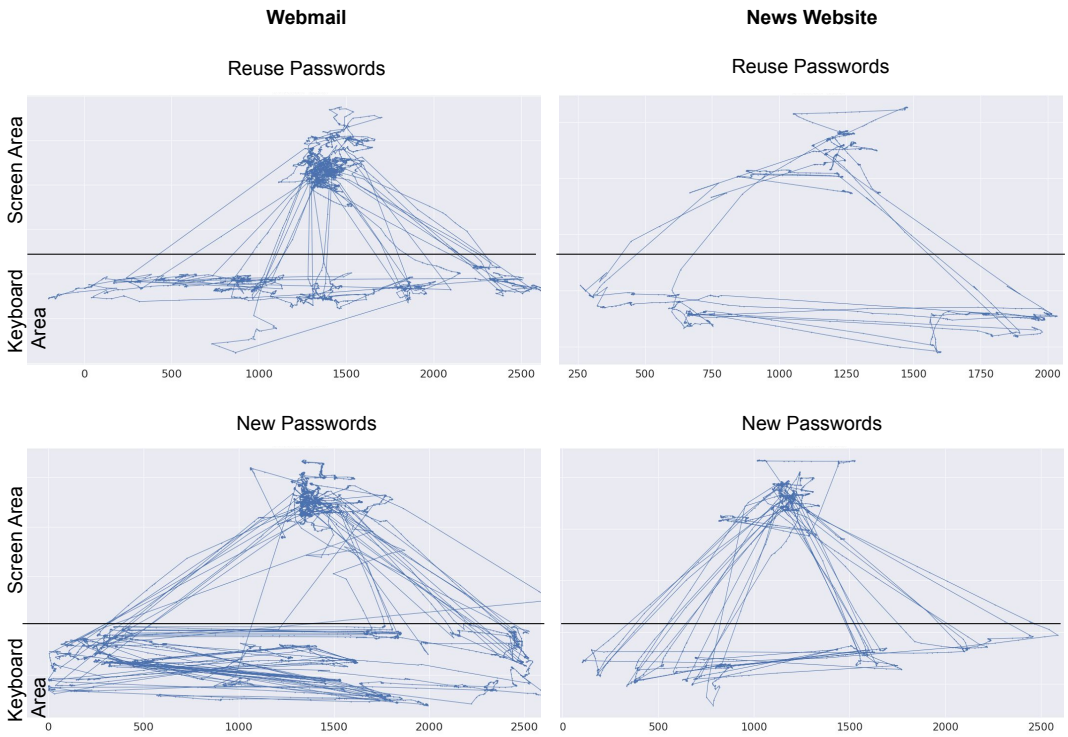


Figure 7.6: Visualization of selected users' gaze paths, both for the webmail (left) and news website (right) interface: In both cases, fixations are primarily focused on the input fields in the middle of the screen. Yet, for cases in which participants created new passwords, more transitions between screen and keyboard occur and more fixations are located in the keyboard area [9].

7.5.4 Gaze Path

As a complementary analysis, we visually inspected the eye movements in the form of the gaze path. Figure 7.6 shows some selected examples. We found that participants fixate more often on the screen (area 1) and keyboard (area 2) while creating new passwords, compared to when entering a reused password. This was independent of the interface on which passwords were created.

7.5.5 Classifier Performance

We compared the performance of three different models: SVM, random forest, and decision trees. We conducted two classifications: *phase-based classification* (i.e. per phase of the password registration) and *multiple phases classification*.

Table 7.7: Interface-dependent Classifier: Classification Performance per Phase for the Different Features (best AUC bold) [9].

Email Web-client		Orientation Phase (O Phase)		Identification Phase (ID Phase)		Password Phase (P Phase)		Confirmation Phase (C Phase)		All Phases	
		AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Gaze Features	**SVM**	64.79 ± 9.50%	55.63 ± 1.51%	**74.35 ± 2.12%**	62.77 ± 9.46%	70.22 ± 4.78%	48.61 ± 1.39%	80.81 ± 0.67%	61.46 ± 5.21%	**87.73 ± 0.23%**	77.08 ± 14.58%
	Random Forest	**72.18 ± 0.76%**	55.87 ± 1.75%	67.62 ± 0.03%	56.94 ± 6.94%	**81.67 ± 8.14%**	61.29 ± 10.93%	**81.92 ± 0.44%**	55.90 ± 0.35%	83.44 ± 0.35%	61.46 ± 5.21%
	Decision Tree	49.05 ± 0.95%	61.54 ± 10.36%	60.33 ± 0.78%	58.33 ± 8.33%	55.56 ± 5.56%	65.75 ± 17.59%	56.83 ± 0.58%	60.33 ± 0.78%	75.84 ± 1.94%	72.22 ± 2.78%
Keystroke Features	**SVM**	-	-	53.40 ± 2.48%	52.78 ± 2.78%	66.23 ± 1.42%	49.14 ± 7.48%	54.89 ± 0.26%	50.00 ± 0.00%	63.58 ± 4.02%	68.06 ± 6.94%
	Random Forest	-	-	**61.04 ± 7.34%**	50.27 ± 3.04%	**69.23 ± 2.10%**	66.24 ± 0.43%	**75.54 ± 2.40%**	50.18 ± 0.18%	**75.83 ± 0.10%**	68.75 ± 6.25%
	Decision Tree	-	-	47.64 ± 0.89%	42.91 ± 4.31%	69.23 ± 2.10%	70.75 ± 1.31%	61.83 ± 6.69%	50.90 ± 6.45%	72.16 ± 5.62%	63.11 ± 3.55%
Both Features	**SVM**	-	-	**76.85 ± 1.85%**	62.77 ± 9.46%	71.51 ± 5.34%	48.61 ± 1.39%	**81.37 ± 1.96%**	61.46 ± 5.21%	87.73 ± 0.23%	78.47 ± 15.97%
	Random Forest	-	-	70.11 ± 0.26%	67.97 ± 4.08%	**80.47 ± 8.42%**	56.70 ± 9.48%	75.83 ± 0.10%	67.36 ± 4.86%	**88.75 ± 0.14%**	62.77 ± 9.46%
	Decision Tree	-	-	55.82 ± 2.51%	58.60 ± 5.29%	56.93 ± 3.25%	57.39 ± 3.72%	54.51 ± 1.74%	57.29 ± 1.04%	74.92 ± 2.86%	72.22 ± 2.78%
News Website		Orientation Phase (O Phase)		Identification Phase (ID Phase)		Password Phase		Confirmation Phase (C Phase)		All Phases	
		AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Gaze Features	**SVM**	**67.35 ± 1.97%**	37.36 ± 9.80%	**84.56 ± 2.74%**	74.56 ± 0.44%	**77.82 ± 3.69%**	67.85 ± 7.36%	60.37 ± 2.33%	51.98 ± 1.98%	**77.49 ± 2.67%**	60.32 ± 10.32%
	Random Forest	48.00 ± 3.13%	52.56 ± 2.56%	78.82 ± 0.15%	63.28 ± 4.18%	75.23 ± 0.40%	60.54 ± 4.59%	**63.28 ± 4.55%**	61.09 ± 2.00%	73.94 ± 0.53%	55.16 ± 5.16%
	Decision Tree	43.07 ± 0.12%	44.99 ± 1.81%	60.85 ± 1.06%	60.05 ± 0.26%	62.99 ± 6.34%	69.53 ± 6.94%	48.77 ± 9.96%	49.14 ± 4.03%	63.19 ± 0.10%	55.16 ± 5.16%
Keystroke Features	**SVM**	-	-	**73.85 ± 3.92%**	73.92 ± 5.04%	54.36 ± 19.07%	76.35 ± 10.62%	**72.94 ± 4.68%**	56.24 ± 4.25%	74.65 ± 4.72%	66.16 ± 5.67%
	Random Forest	-	-	73.22 ± 0.21%	64.16 ± 0.52%	**67.53 ± 0.30%**	71.14 ± 9.95%	72.87 ± 0.14%	59.61 ± 5.07%	**80.97 ± 3.99%**	62.77 ± 0.87%
	Decision Tree	-	-	70.22 ± 3.55%	63.51 ± 6.77%	60.92 ± 0.43%	59.59 ± 1.60%	58.91 ± 0.18%	65.81 ± 6.02%	62.68 ± 2.36%	57.28 ± 2.51%
Both Features	**SVM**	-	-	**84.56 ± 2.74%**	74.56 ± 0.44%	**77.82 ± 3.69%**	66.38 ± 5.89%	61.61 ± 4.47%	51.98 ± 1.98%	76.70 ± 1.87%	60.32 ± 10.32%
	Random Forest	-	-	80.77 ± 1.40%	67.82 ± 0.36%	76.73 ± 2.26%	65.75 ± 5.26%	**78.21 ± 2.34%**	65.55 ± 1.91%	**77.96 ± 1.76%**	72.19 ± 4.00%
	Decision Tree	-	-	64.32 ± 3.14%	61.44 ± 1.65%	63.71 ± 2.37%	68.83 ± 7.64%	73.18 ± 3.74%	58.91 ± 0.18%	63.19 ± 0.10%	55.16 ± 5.16%

Phase-based Classification

We use data from the different registration phases (cf. Figure 7.2) to build the model. The phase-based model helped us understand how each phase contributes to the model. To understand which features are best for our classification task, we ran the classifier on gaze features only, keystroke features only, and both. Random forest and SVM classifiers resulted in a similar AUC (Area Under the Curve) score. However, SVM resulted in a better AUC score in most cases. Hence, the remainder of our analysis will focus on and report the SVM results.

For the *interface-dependent classifier*, Table 7.7 shows the overall performance of classification for each interface for all classifiers across the different phases. For webmail, the AUC is best when combining all phases. The highest AUC is 87.73% for gaze features and 88.75% for the combination of gaze and keystroke features. This means that users‘ behavior is more reflected in their gaze behavior features than in their typing behavior. Also, gaze features better reflect users‘ password behavior across the different phases. For the news website, similar to the webmail client, the best AUC is achieved when considering gaze features and the combination of gaze and keystroke features. The accuracy here is highest in the “identification phase” (84.56%). Our interpretation of this is that the password choice is primarily made during this phase. The keystroke features allow for an equally good prediction, but only when considering all phases. This means that for interfaces protecting sensitive content, password reuse is more accurately detected using gaze or both gaze/keystroke features during the identification phase.

For the *interface-independent classifier*, Table 7.8 shows the overall performance of the classifiers for all interfaces across the features. The highest AUC is achieved for gaze features and both features when combining all phases (71.87%).

Table 7.8: Interface-independent Classifier: Classification Performance Per Phase for the Different Features (best AUC bold) [9].

		Orientation Phase (O Phase)		Identification Phase (ID Phase)		Password Phase (P Phase)		Confirmation Phase (C Phase)		All Phases	
		AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Gaze Features	SVM	**66.27 ± 0.95%**	46.91 ± 1.91%	**58.12 ± 2.58%**	49.59 ± 1.51%	59.28 ± 5.78%	63.95 ± 3.62%	65.64 ± 3.56%	51.19 ± 4.69%	**71.87 ± 2.82%**	51.84 ± 1.84%
	Random Forest	62.40 ± 1.00%	48.65 ± 2.42%	56.81 ± 0.07%	61.28 ± 1.49%	**76.91 ± 1.77%**	63.42 ± 15.34%	**68.37 ± 1.18%**	56.94 ± 0.12%	68.22 ± 0.06%	58.50 ± 0.07%
	Decision Tree	52.22 ± 1.04%	52.15 ± 1.27%	52.79 ± 5.73%	56.26 ± 2.25%	59.81 ± 2.58%	56.00 ± 6.09%	52.61 ± 0.42%	53.74 ± 7.62%	61.21 ± 1.53%	57.20 ± 2.48%
Keystroke Features	SVM	-	-	56.70 ± 0.85%	53.49 ± 3.49%	54.05 ± 0.69%	52.78 ± 2.78%	**64.43 ± 1.90%**	52.16 ± 1.98%	60.34 ± 0.43%	53.40 ± 3.40%
	Random Forest	-	-	**62.66 ± 1.63%**	57.40 ± 2.16%	**57.68 ± 1.29%**	61.10 ± 2.76%	59.91 ± 1.82%	49.94 ± 0.06%	**69.22 ± 0.49%**	61.88 ± 4.35%
	Decision Tree	-	-	61.77 ± 3.87%	54.23 ± 13.79%	55.75 ± 2.25%	56.06 ± 4.40%	55.90 ± 4.19%	61.93 ± 4.99%	66.28 ± 1.47%	57.80 ± 2.34%
Both Features	SVM	-	-	58.30 ± 2.41%	51.10 ± 3.02%	59.74 ± 6.27%	63.95 ± 3.62%	65.96 ± 2.84%	51.19 ± 4.69%	**71.87 ± 2.82%**	51.10 ± 1.10%
	Random Forest	-	-	**61.15 ± 0.94%**	59.88 ± 7.25%	**66.89 ± 1.65%**	58.77 ± 3.40%	**69.13 ± 0.26%**	57.46 ± 3.52%	70.73 ± 0.08%	64.03 ± 3.54%
	Decision Tree	-	-	57.99 ± 4.67%	56.37 ± 4.28%	52.39 ± 1.89%	59.01 ± 5.54%	60.51 ± 5.02%	56.65 ± 8.88%	62.41 ± 3.63%	57.20 ± 2.48%

Multiple-Phase Classification

This model accumulates all information available on users‘ behavior, from the beginning of the registration process to a particular phase. The aim of this model is to understand which features are best for classification. We ran the classifier on gaze features only, keystroke features only, and both. Random forest and SVM classifiers resulted in a similar AUC score. However, SVM resulted in a better AUC score in most cases. Hence, in the following we will focus on and report the SVM results.

For the *interface-dependent classifier*, Table 7.9 shows the overall performance for the classification for each interface across all classifiers for the accumulated phases. For webmail, the AUC is best, when all phases are combined. The highest AUC is 87.73% for gaze features. However, the model shows a decrease of only 2% for considering only the *O + ID phase* as well as when the *O + ID + P phases* are considered. This means that our model can predict password reuse in the identification phase *before* the user start typing the actual password reasonably well. For the keystroke features, the best AUC is still the same as the phase-based classification. However, looking at the accuracy after each phases along the registration process, we found a difference in accuracy of 6% across the grouped phases. This means that by using the keystroke features only, the best accuracy is achieved when the user has clicked ‘register’. Finally, for both features combined, the picture was diverse. For webmail, accuracy continuously increased. Yet, for the news website, the highest accuracy was achieved in the identification phase. In subsequent phases, accuracy differed minimally.

For the *interface-independent classifier*, combining the phases did not yield a better accuracy compared to phase-based classification. This indicates that for the interface independent classifier any model will lead to a similar accuracy.

True Positive and True Negative Values

As multiple phase classification did not affect the true positive and true negative rate, we only report values for the phase-based classification for the gaze features models. The data set was unbalanced. The guessing baseline (i.e. trivial classifier always guessing majority class) is 71% for webmail and 63% for the news website. Our classifiers outperform the baseline (81.6% for webmail, 74.6% for news website).

Table 7.9: Classification performance for the interface-dependent classifier (multiple phases): Phases represented by O (orientation), ID (identification), and P (password entry). Best AUC in bold [9].

Email Web-client		O Phase		O + ID Phases		O + ID + P Phases		All Phases	
		AUC	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**
Gaze Features	**SVM**	64.79 ± 9.50%	52.06 ± 2.06%	77.11 ± 3.04%	73.61 ± 1.39%	85.04 ± 5.41%	54.17 ± 4.17%	**87.73 ± 0.23%**	71.53 ± 9.03%
	Random Forest	**72.18 ± 0.76%**	55.87 ± 1.75%	**85.68 ± 5.13%**	61.81 ± 0.69%	**85.16 ± 2.81%**	65.97 ± 3.47%	83.44 ± 0.35%	63.46 ± 2.35%
	Decision Tree	49.05 ± 0.95%	50.00 ± 0.00%	65.22 ± 0.52%	49.92 ± 2.86%	66.07 ± 6.15%	66.07 ± 6.15%	75.84 ± 1.94%	70.32 ± 7.46%
Keystroke Features	**SVM**	-	-	53.40 ± 2.48%	45.85 ± 1.37%	67.35 ± 1.17%	63.11 ± 3.55%	63.58 ± 4.02%	48.53 ± 1.47%
	Random Forest	-	-	**61.04 ± 7.34%**	52.78 ± 2.78%	65.36 ± 0.08%	54.17 ± 4.17%	**75.83 ± 0.10%**	61.81 ± 0.69%
	Decision Tree	-	-	47.64 ± 0.89%	40.30 ± 4.19%	**69.96 ± 7.82%**	68.85 ± 8.93%	72.16 ± 5.62%	72.16 ± 5.62%
Both Features	**SVM**	-	-	**77.60 ± 4.81%**	70.85 ± 1.37%	**85.04 ± 5.41%**	54.17 ± 4.17%	87.73 ± 0.23%	71.53 ± 9.03%
	Random Forest	-	-	77.40 ± 0.54%	61.38 ± 8.07%	84.50 ± 0.68%	65.62 ± 9.38%	**88.75 ± 0.14%**	63.11 ± 3.55%
	Decision Tree	-	-	65.22 ± 0.52%	49.92 ± 2.86%	66.07 ± 6.15%	66.07 ± 6.15%	74.92 ± 2.86%	70.32 ± 7.46%
News Website		**O Phase**		**O + ID Phases**		**O + ID + P Phases**		**All Phases**	
		AUC	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**
Gaze Features	**SVM**	**67.35 ± 1.97%**	46.45 ± 0.99%	**83.62 ± 0.29%**	72.98 ± 4.80%	**81.43 ± 0.31%**	69.76 ± 0.88%	**77.49 ± 2.67%**	67.30 ± 6.11%
	Random Forest	48.00 ± 3.13%	52.88 ± 2.88%	76.20 ± 2.77%	68.33 ± 4.69%	77.61 ± 1.42%	75.05 ± 2.33%	73.94 ± 0.53%	61.80 ± 7.25%
	Decision Tree	43.07 ± 0.12%	43.07 ± 0.12%	60.05 ± 0.26%	60.05 ± 0.26%	70.46 ± 0.18%	70.46 ± 0.18%	63.19 ± 0.10%	56.55 ± 6.55%
Keystroke Features	**SVM**	-	-	**73.85 ± 3.92%**	56.15 ± 6.15%	71.12 ± 1.89%	60.51 ± 4.57%	74.65 ± 4.72%	66.07 ± 10.12%
	Random Forest	-	-	73.22 ± 0.21%	59.61 ± 5.07%	**75.11 ± 0.29%**	63.28 ± 4.18%	**80.97 ± 3.99%**	67.14 ± 3.50%
	Decision Tree	-	-	70.22 ± 3.55%	67.48 ± 2.80%	68.69 ± 1.55%	65.36 ± 4.87%	62.68 ± 2.36%	62.68 ± 2.36%
Both Features	**SVM**	-	-	**84.42 ± 0.50%**	73.68 ± 4.10%	**81.43 ± 0.31%**	72.73 ± 2.10%	76.70 ± 1.87%	61.11 ± 11.11%
	Random Forest	-	-	77.00 ± 0.78%	63.28 ± 4.18%	76.21 ± 0.02%	62.39 ± 7.85%	**77.96 ± 1.76%**	58.82 ± 4.27%
	Decision Tree	-	-	58.91 ± 0.18%	58.91 ± 0.18%	71.65 ± 1.37%	71.65 ± 1.37%	63.19 ± 0.10%	56.55 ± 6.55%

Table 7.10: Comparison of eye movements for the webmail client/news website (only factors with statistically significant effects) [9].

Gaze Features	**Saccadic Duration**			**Avg. Fixation Duration**			**Saccadic Length**			**Keyboard Fixations**		
	Email Rank	News Rank	Wilcoxon	Webmail Rank	News Rank	Wilcoxon	Webmail Rank	News Rank	Wilcoxon	Webmail Rank	News Rank	Wilcoxon
Reuse Passwords	4.25	8.80	Z=-2.22, P=.026	5.25	8.40	Z=-1.97, P=.048	4.75	8.60	Z=-2.10, P=.035	7.50	7.50	Z=-2.35, P=.019

For *webmail* we found that 32 out of 35 new passwords were correctly classified as new. For the reused passwords, 8 out of the 14 reuse passwords were correctly classified. For the *news website* we found that out of the 31 newly generated passwords, 21 passwords were correctly classified as new. Out of the 18 reused passwords, 15 were correctly classified as reused. For the *interface independent classifier*, out of the 66 newly generated passwords, 56 were correctly classified as new. Out of the 32 reuse passwords, 12 were correctly classified as reuse. We reflect on these results in the discussion.

Feature Importance

We investigated which features mostly contribute to the accuracy of the classifiers. We found only small differences between both interfaces and here show the features for webmail only. We used SHAP [243], a tool that explain the output of a machine learning model by computing the contribution of each feature to its prediction. Figure 7.7 shows the feature importance.

We observed that for the gaze features, the fixation and registration duration are mostly contributing (.23 and .14 respectively). For the keystroke features, we observed that the overall registration duration and flight time contributed most to prediction of the password category (.09 and .06 respectively). For both features, we found that the

Table 7.11: Comparison of keystroke dynamics for the webmail client/news website (only factors with statistical significance) [9].

Keystroke Features	Typing Duration			Keystrokes Count			Thinking Time		
	Email Rank	News Rank	Wilcoxon	Email Rank	News Rank	Wilcoxon	Email Rank	News Rank	Wilcoxon
Reuse Passwords	4.50	8.70	Z=2.17, P=.03	6.67	7.73	Z=-2.04, P=.041	4	7.90	Z=-2.34, P=.019

gaze features have a stronger influence on the model‘s accuracy than the keystroke features.

Prediction Over Time

Figure 7.8 visualizes the *AUC* over time for the investigated conditions. Between interfaces, we can see that gaze leads to a higher accuracy much faster for webmail, i.e. when passwords are created to protect more sensitive data. The prediction accuracy for keystrokes is plateauing in the identification phase (i.e. after about 13 seconds for the news website and 22 seconds for webmail). Gaze enables predictions are possible from the beginning of the identification phase, providing a time advantage.

7.5.6 Effect of Data Sensitivity on User Behavior

To study the effect of content sensitivity on user behavior, we ran a Wilcoxon signed-rank test on users' gaze features and keystroke features. We didn't find a statistically significant effect of data sensitivity, neither on gaze behavior nor on keystroke dynamics. However, for reused passwords, we found significant effects of data sensitivity on behavior.

Table 7.10 and 7.11 show the statistical significant features. For users' *gaze behavior*, we found significant differences for the saccadic duration, average fixation duration, saccadic length, and number of keyboard fixations between the webmail client (more sensitive) and the news website (less sensitive). For users' *keystroke dynamics*, we found statistical differences for users' typing duration, keystrokes count, and thinking time. The results show differences in users' behavior between interfaces protecting data with different sensitivity, but only when registering reused passwords.

7.6 Discussion

We presented an investigation of eye movement behavior and keystroke dynamics to identify whether people reuse passwords, specifically during the password registration phase. I

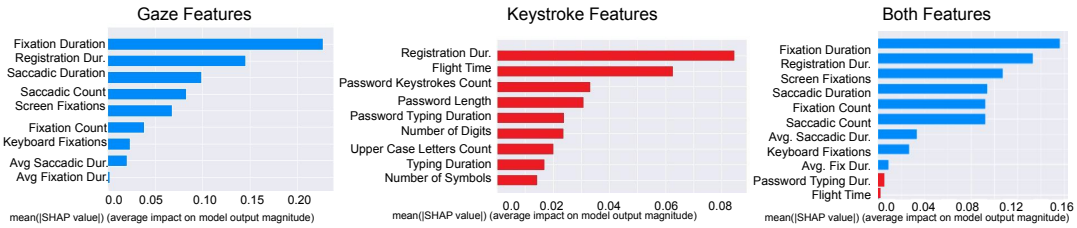
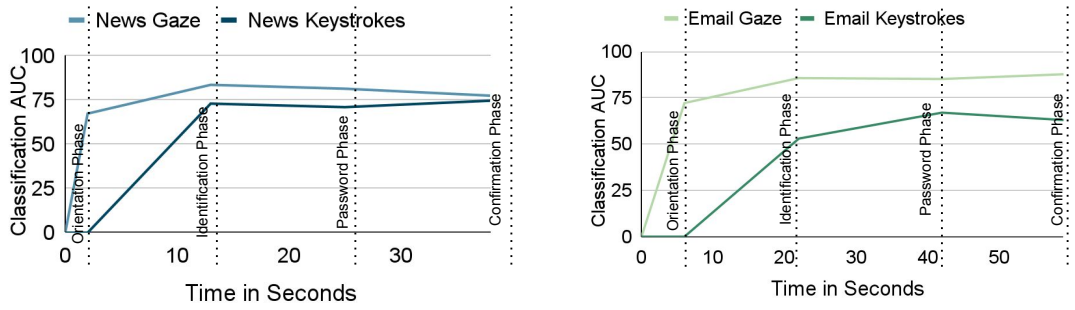


Figure 7.7: Results of the feature importance analysis across the tested feature groups for the email client [9].



(a)News Gaze and Keystroke AUC for cumulative multiple phases

(b)Email Gaze and Keystroke AUC for cumulative multiple phases

Figure 7.8: AUC comparison for multiple phases classifier between gaze and keystroke features for new and reuse passwords across interfaces. It shows that our addition of using gaze outperformed using keystrokes [9].

7.6.1 Gaze is More Informative than Typing

We found that a classifier based on gaze-related features (88% AUC for the interface-dependent classifier) outperforms a classifiers based on typing behavior only (80% AUC). Note that the results for typing behavior are in line with prior work [168]. Furthermore, the accuracy can be improved by *combining* typing and gaze features in some cases. Prediction accuracy for keystroke features is higher only at a later stage – namely after users have typed the password. These findings answer our raised questions earlier (section 7.3). More specifically they show that it is not only possible to detect password reuse from these features but to also obtain rich additional information.

7.6.2 Data Sensitivity Influences Accuracy of Password Reuse Prediction

We found that the sensitivity of the protected data affects the characteristics of the chosen password and whether it is a new one or a reused one is reflected in the user's

gaze data. More participants reused passwords for the news website than for the webmail client. This suggests that the more sensitive the protected information is, the more effort people put into their password and the less often they reuse passwords. This also leads to users' behavior getting more distinguishable. This is revealed by the statistical analysis where, for the webmail client, most features (gaze and typing) were significantly different between reused and new passwords. In contrast for the news website we could not find significant difference in our collected data.

7.6.3 Dissecting Password Registration Process Enriches Modeling and Prediction

Contributing to the literature, we dissected the observation of password creation behavior into multiple phases.

For the webmail client, we found that considering users' behavior during the whole password generation (all phases combined) to detect password reuse leads to the best accuracy. In contrast, for the news website, we found that the identification phase better reflects users' behavior to detect password reuse. This suggests that people think about passwords during different phases of the registration process and that this thinking takes longer when protecting more sensitive data. We ran a Wilcoxon test to see whether the duration of the identification phase differed for new ($MeanRank = 10.27$) and reused passwords ($MeanRank = 8.29$) for the news website. We did not find statistically significant differences ($Z = -1.98$, $p > .05$). This motivates a future study, striving to obtain a deeper understanding of when and how much people 'think ahead' when creating passwords.

7.7 Practical Implications for the Design of Password Systems

Being able to identify password reuse before the end of the registration process, we envision interfaces to implement interventions ultimately leading to better passwords. We reflect on the role of eye tracking, the design of interventions, the implications of user and interface characteristics on modeling, and on privacy implications.

7.7.1 Ubiquitous Eye Tracking

We believe the vision sketched in this work to be timely as eye tracking is about to become ubiquitously available and to, in particular, gain relevance in usable security

[181]. Access to gaze data is possible today in different ways. Firstly, there is laptop and desktop computers being equipped with dedicated eye tracking hardware. The fact that Apple bought SMI, one of the world's leading manufacturers of eye tracking hardware suggests, that one of the next generations of Macbooks might come with integrated eye tracking. Secondly, advances in computer vision made it possible to perform appearance-based gaze estimation simply by means of analyzing the video feed of a web cam or smartphone cam [187]. Thirdly, eye wear (such as augmented reality glasses and head-worn devices) are envisioned to use gaze as a communication medium for everyday interactions [284], and thus could open doors for security use cases.

Our approach could be implemented in various forms. Providers wishing to support users in choosing better passwords could integrate the approach with their password registration interface (e.g., by accessing the webcam on a PC, by a smartphone app accessing the front-facing camera, or the built-in eye tracker of head-worn devices). A provider-independent solution would be a browser plugin that accesses the camera and assesses users' gaze data as they enter a website requiring the registration of a password. Finally, the approach could run as a service in eye wear, that activates when users are about to register a password and then assesses their physiological data.

7.7.2 Modeling

Different factors can influence the classification modeling.

Ground Truth: Determining Password Reuse

The first step to building predictive models is to *collect behavioral data* during the authentication process. The challenge during this data collection is to *obtain a ground truth*, i.e. whether or not users are creating new password or reusing an existing ones. Several alternatives exist. Firstly, users could be asked to provide this information. Yet, this creates an overhead for the user. Secondly, the created password could be compared to (the hashes of) passwords other users created for the data or service the mechanism is protecting. Third, the created password could be compared to databases of leaked passwords. Afterwards, a model can be trained based on the labeled set of behavioral data, following the approach outlined in this wsork.

Influence of Typing Proficiency

In our study, we sampled among a University population where people were likely to have a rather high typing proficiency. However, this might be different for other samples. Typing behavior is mainly a result of how long people type daily. In addition, typing and keystroke dynamics are influenced by cognition, which differs when typing routine words (i.e. password reuse) as opposed to non-routine words (i.e. new

passwords) [168]. Dhakal et al. [100] analyzed typing behavior in an online survey and they clustered typists into eight groups based on their typing performance, accuracy, rollover, and hand usage. Given all this, we learn that user's typing proficiency plays a role to affect keystroke behavior and, hence, the accuracy of a classifier predicting password reuse. A user-dependent model is more suitable to capture individual characteristics and can enhance accuracy.

Influence of Screen Properties

Users might access the same password registration interface on devices with different screen properties (e.g., a laptop vs. a large external monitor). While we maintained the same screen in our study for data consistency, other display types might be worth considering. In our analysis, we inspected the degree of influence the features have on prediction accuracy. Fixation and registration durations are among the most prominent features. We expect the influence of the screen properties on such relative features to be low. However, to further enhance the classification accuracy and take into account device-dependent features such as saccadic duration and path, it might be useful to consider screen-optimised classifiers.

Influence of Layout

Ideally, a model would make highly accurate predictions independent of the password registration interface layout. In our study, we investigated two examples from the real world that we believe are representative for many of the layouts in use. However, other registration interfaces might look different and ask the user, for example, to provide information beyond credentials on the same page, such as an address or payment information. One might speculate whether users already display behavior related to password composition before working on the respective part of the form. If so, this would be interesting, as it gives a system employing our concept more time for an intervention and also more typing and gaze data. At the same time this would require a new model to be trained.

Future work could investigate, how exactly the registration interface, in particular, the requested information and the layout (e.g., at which part of the registration interface the password is composed) influence prediction accuracy.

Influence of Interaction Modality

We hypothesize that different interaction modalities will likely affect the typing behavior, because input devices vary across systems (e.g., using a mechanical vs. a touch-sensitive keyboard). The same is potentially true for gaze as different forms of eye trackers might be employed with different systems and typing behavior might influence gaze behavior in a different way. At the same time, it is plausible that the implicit nature of eye movements could represent a more constant predictor of

password reuse across systems. This interesting question should be pursued by future research.

7.7.3 User Privacy

Note that it is important to consider the potential privacy implications of using gaze data. There is an ongoing discussion on the need to use gaze data carefully. From gaze, information beyond password reuse can be inferred, including but not limited to the users' interest, attention, fatigue, or sexual orientation (see Steil et al. [342] for an in-depth assessment of this topic). One could assume that users might be willing to share gaze data if it was to their benefit, in particular, in a security context. Yet, consent to collect and assess gaze data should not only be obtained by the provider of a password reuse identification system but be limited to this authentication procedure.

7.8 Chapter Summary

We presented a novel approach for predicting password reuse. We separated password registration into different phases, namely the 1) orientation phase, 2) identification phase, 3) password typing phase, and 4) confirmation phase. We then looked at how well password reuse can be detected in the different phases (separately and accumulated) based on gaze, keystroke dynamics, and both. In addition, we compared two interfaces, meant to protect more and less sensitive data. Beyond showing that our approach improves the accuracy of prior work, we additionally demonstrated that prediction becomes now feasible throughout the entire password registration process.

In addition to the preceding chapter, the findings from this work provide significant contributions to the usable security community by introducing a new dimension of assessing passwords. Our approach strengthens passwords by adding a behavioral layer to password assessment tools or creating new tools that are independent of the interface and do not require access to the passwords. Findings from both chapters answer our **RQ3:** *How can users' behavior during authentication be modeled?*.

IV

Discussion and Conclusions

Outline

So far, we've demonstrated how tracking and analyzing users' gaze movements may be utilized to improve existing authentication systems. In this thesis part, we synthesize our findings into a framework for employing eye gaze behavior in security systems. We also discuss how existing interfaces can benefit from our findings and whether our novel idea of employing gaze monitoring to improve existing systems be utilized for purposes other than authentication. We also reflect on the thesis findings, provide implications and recommendations for deploying behavioral gaze systems, and finally, conclude with some future directions. This thesis part answers our **RQ4**: What are the considerations for employing gaze behavior in security systems?

This part of the thesis takes a closer look at the following aspect:

- **Chapter 8** - Synthesis. This chapter combines the findings from all the user studies presented in the earlier chapters. We synthesize our findings into a framework for behavioral eye-based security systems.
- **Chapter 9** - Discussion. This chapter reflects on the main findings of the thesis, how can our findings be used beyond authentication, and reflects on users' privacy.
- **Chapter 10** - Conclusion. This chapter concludes the thesis by reflecting on the different research contributions presented, and providing a set of future work directions.

Chapter 8

Framework

In this chapter, we combine the findings from all the publications and the user studies presented in the earlier chapters. We synthesize the findings from the conducted data collection studies, interviews, and surveys into a framework for conducting behavioral eye-tracking studies. We highlight three main stages for employing gaze behavior in security systems, 1) understanding, 2) modeling, and 3) feedback. We also reflect on methodological, technical, empirical, and ethical considerations for each stage. Table 8.1 shows the overview of the framework. This chapter answers **RQ4**: What are the considerations for employing gaze behavior in security systems?

8.1 Understanding User Behavior

The first step in employing eye gaze behavior in security systems is to understand users' behavior in this context. To do that, there are a set of considerations to be tackled on different levels.

Table 8.1: Framework for Employing Eye Gaze Behavior in Security Systems

	Considerations	**Understanding User Behavior**	**Modeling User Behavior**	**Feedback**
Ethical	**Methodological**	Study Setup Study Type Study Design Measures	Classifiers Areas of Interest Features	Nudging Interventions
	Technical	Eye Tracker Storage	Data Cleaning Processing	Feedback Channel Feedback Timing
	Empirical	Recruitment and Participants	Amount of Data	Communicating data collection

8.1.1 Methodological Considerations

The first one is the methodological considerations. There are different aspects for the methodological consideration, 1) study setup, type, design, and measurements. Below we discuss each of them.

Study Setup

With new advancements in eye-tracking manufacturing, eye-tracking data collection can be conducted in different setups. Such as in a remote eye-tracking setup, in the lab, in the wild, and in virtual reality (VR) environments. All setups are valid and could be used to collect behavioral eye gaze data, however, it depends on the nature of the task and the data needed.

For example, remote setups enable the collection of more ecologically valid data. This is due to environment familiarity where participants are in their comfort zone removing the effect of unfamiliar environments which can add cognitive load on participants affecting data collection. However, such uncontrolled environments can induce noise to the data due to interruptions which study designers should tackle ways to detect and deal with interruptions. Also, experimenters can offer help remotely if needed, or such studies could be conducted using any video conferencing tool if needed. Such studies are not conducted very often in the eye-tracking community due to several aspects that could affect users' gaze. On the other side, data collection in lab setup is the most common way in the literature. Although, data collection in the lab enables more control on the user position, and interruption, it might not induce natural user behavior due to being in an unfamiliar environment where the experimenter can be in the same room noticing their behavior.

Conducting in-the-wild studies is critical, although it provides ecologically valid data from users' day-to-day life, it also introduces a lot of uncertainty about the reasons behind the change in behavior. hence, it is challenging to analyze and draw conclusions from. Such studies are very scarce in the literature as we showed in chapter 3. Finally, with new head-mounted displays integrating eye trackers, it opened a new environment for collecting eye gaze data. Using VR as a research method enables studying users' behavior in scenarios that are hard to study in the real world such as studying shoulder surfing [8] and to have more control over the environment and stimuli [292].

Study Type

Another aspect when collecting behavioral eye gaze data is the study type, is it a deception study or not? Most behavioral eye gaze data studies tend to conduct deception studies to eliminate priming users with the study aim. However, in such cases, experimenters should debrief the participants at the end of the study and retake their consent to analyze the data with the new aim revealed and enable opting out to preserve users' privacy.

Study Design

We reflect on different study design aspects affecting the accuracy of the data collection.

Study Task The study task affects users' perception, which affects their gaze behavior. This includes task duration, which can cause eye fatigue and boredom. For example, we attempted to recruit a large number of participants, each undertaking a single task that took 5 minutes, and recruiting a smaller number of participants with more repetitions that took 45 minutes. Each serves a distinct goal of feeding user-dependent or independent classifiers, but the shorter the task and the fewer the repetitions, the more inclined users are to participate in the study, especially if the task requires high mental effort. It is also worth mentioning that in most behavioral eye-tracking studies, gaze monitoring is conducted in the background while users are performing a primary task. Hence, the primary task is important to either elicit the unique behavior to be studied or worst case scenario not affect the actual aim of the study.

Interface Design and Layout Another major aspect of the study design is the interface design and layout. The interface components, such as the implicit and explicit cues included in the interface, influence users' gaze behavior. For example, if the interfaces contain moving or animated advertisements/objects, users' gaze tends to follow movement [370], which influences their overall behavior, especially if the movement is dynamic. Furthermore, the information density and location of areas of interest (AOI) influence user gaze behavior. Hence, rigorous research should be conducted to identify the positions and density of the AOI to elicit users' unique gaze behavior. As our gaze guides our perception of the surroundings, it indicates that the interface design can guide or influence users to look at certain AOI. Following this, the size of interface elements influences user perception, which in turn influences user gaze behavior. Adding vertical or horizontal lines, for example, directs users' perception, influencing their gaze behavior.

Adaptive Interfaces Gaze data might contain inattentive users' behavior, which could be considered noise in different scenarios. As a result, gaze-based interfaces should account for such noise. The interfaces should be adaptable to identify and eliminate these motions without impacting the data obtained from the task. Such interruptions include, and are not limited to, Internet connection failure, pets, such as a cat, moving in close proximity to the users, on-screen notifications, phone rings, door knocks, etc.Furthermore, changing light conditions affect users' gaze' thus, user interfaces should keep track of the light conditions in the user's environment and issue warnings if direct light sources or insufficient light are detected, as this affects the quality of data collection and the precision of eye tracking. This can be addressed by, for example, inverting the screen colors to account for varied lighting conditions, disabling on-screen notifications if necessary, suspending the interface if no direct gaze is detected, etc.

Interaction Modality and Screen Properties In chapter 6, we found that the different interaction modalities affect users' behavior. This means that the device and screen size have an impact on users' behavior, for example, larger displays might contain more AOIs. Moreover, various devices are utilized for different use cases, such as employing the device on the move, e.g. tablet or smartphone, or in a static setup, such as a desktop computer, therefore the environment setting can also have an influence on users' gaze behavior. As a result, the various devices have distinct interaction modalities, such as on-screen interactions like on a smartphone or extended interactions like utilizing an external keyboard and mouse, for example, in desktop contexts, or devices that can be used either way, like tablets. These various modalities influence users' gaze behavior as they are influenced by the shift between the screen and the input modality. In addition, it affects the display configuration, for example, in the case of smartphones, the interface shifts up to create room for the touch keyboard. This is quite different in the case of head-mounted displays, as the interactions are frequently in the air or haptic.

Measures

Finally, the last aspect of the methodological considerations for understanding user behavior is measurements. Which reflects the different aspects that are intended to be investigated. These could be cognitive, physiological, and behavioral measures. Below we discuss some aspects that could be detected from users' eyes and help understand their behavioral data.

Mental Effort According to Kirschner et al. [202], mental effort refers to how hard a person works to actively process the information that is offered. The mental effort, which is more commonly referred to as a cognitive load, is made up of three separate factors: cognitive, perceptual, and volitional [201]. Eye tracking has demonstrated significant promise for identifying mental effort. For instance, Chen et al. [75] demonstrated that three classes of eye features are capable of differentiating between various levels of mental effort. Similar to this, task-evoked pupillary responses (TEPR), blink frequency, and gaze speed were employed by Mosaly et al. [262] to identify various task difficulty levels. Moreover, earlier research indicated that Heart Period Variability (HPV) might be used as a gauge of genuine mental exertion [165].

Mental Stress Mental stress affects users' thinking and impairs decision-making [276, 261]. This affects their ability to take security decisions such as accepting cookies or clicking on a phishing link. On the other side, such attacks put users under the stress of taking quick decisions, e.g. expiring links in 24 hours, or accepting cookies to continue, etc. Therefore, we believe that stress can be a good indicator of being under attack or we can use it to prevent users from taking decisions while they are stressed, e.g. delay sending a response to a phishing email in the morning. From psychological response, stress can be detected using heart rate [351], eye tracking [173, 382, 379], facial expressions [21], mouse movements [38, 379], and keystroke dynamics [372].

Personality Traits Personality traits and profiles affect users' on their tendiness to be hacked. To identify underlying elements that increase susceptibility to scams, a few research explored the relationship between personality traits and scams. For example, Agreeableness, characterized by warmth and friendliness, was shown to be a significant predictor of phishing susceptibility such that those high in agreeableness were more susceptible [23]. Similarly, Extraversion, or level of outgoingness, was a significant predictor of phishing susceptibility such that more extraverted users were more susceptible [23, 225]. Ss personality traits could be detected from users' gaze [72, 44], and mouse and keystroke dynamics [198], we believe that it can be used to better understand users' behavior.

Emotions There is an emotional component incorporated in each of the measurements discussed earlier. Users' actions and attitudes are primarily motivated by their emotions [279, 240]. As demonstrated by Alseadoon et al. [23] and Halevi et al. [138], emotional stability, for instance, was a highly significant predictor of phishing vulnerability. The inability to manage one's emotions is another reason why individuals continue to fall for phishing. Researchers at the University of Exeter [226] discovered that scam victims claimed to be unable to resist giving in to persuasion and being random in their responses. Similarly, Hirschberg et al. [148] observed that individuals with high levels of neuroticism had a considerably smaller chance of spotting lies, maybe because they are more likely to become upset when told a lie and prefer to believe that people are generally honest (to avoid emotional pain). There are various examples relating users' emotions to their security profiles and attack propensity. We believe that since emotions are detectable and quantifiable in a wide range of ways, including facial expressions [215], heart rate [297], wearable devices [223], and eye tracking [299] similar to personality traits, it could be used to understand and reflect on users' behavior.

8.1.2 Technical Considerations

The second aspect of the understanding phase is the technical considerations. Below we discuss the eye tracker choice and data storage.

Eye Tracker Choice

Advances in computer vision algorithms have improved the accuracy and reliability of gaze estimate approaches, paving the way for conducting eye tracking studies in different setups such as remote, in the lab, semi-controlled studies, etc. Each method can be conducted using different hardware and each hardware will have a different impact on the collected gaze data.

For example, smartphone cameras and webcams can now act as eye trackers and be used in-the-wild studies [187]. Simultaneously, eye trackers are becoming more

affordable and ubiquitous, and can be purchased for less than 200 euros. In addition, some laptops and smartphones are now equipped with eye trackers and soon enough Macbook devices will be equipped with integrated eye trackers. Additionally, the addition of eye trackers to the newest versions of Augmented Reality (AR) glasses and the expectation that AR glasses will become a common part of our daily life expand the range of devices that can be used.

The various types of eye-tracking technology, however, can impact data collection and quality. In addition, the use case influences the hardware to be used, which affects data analysis. A wearable eye tracker, for example, provides very accurate gaze data at a high framerate, allowing for more precise data processing. However, because it is obtrusive, others can see it, which influences users' natural behavior. Furthermore, if individuals are not used to wearing glasses, this may affect their gaze data. Furthermore, the changing nature of the environment makes data processing difficult. When it comes to gaze-behavior analysis, we are usually interested in specific regions of interest in the environments that were identified to trigger users' unique behavior. With users' continuous movements, annotating dynamic areas of interest is becoming more challenging and should be carefully considered when choosing eye-tracking hardware.

Our recent work [8] demonstrated the feasibility of employing head-mounted displays in Virtual Reality (VR) environments as a testbed for eye-tracking investigations, allowing for high-quality data collecting as well as manipulating and annotating the environment, making data processing much easier. Our approach of analyzing users' gaze behavior to nudge them to make better security decisions is implemented in various forms. It can be used by service providers, for example, email clients to help users not to fall for phishing, or social media websites to automatically label and flag fake news, etc. This can be done by accessing the webcams on the PCs or front cameras on the smartphones. It can also be used provider independently for example as a chrome plug-in or on eyewear like AR glasses or VR headsets especially since VR headset shipments increase 241.6% after the announcement of the metaverse leading to spending longer duration in VR environments in the near future.

In summary, the choice of eye tracker has a huge influence on users' perception, consequently their gaze behavior. On one side wearable eye trackers can be more obtrusive and might not be socially accepted as others can notice it which can affect users' behavior, however, on the other side, they provide high accuracy with high framerate. We show a comparison between different eye trackers in figure 8.1 highlighting front camera position. Contrary to wearable eye trackers, using webcams enable remote studies and collecting ecologically valid data with the cost of low accuracy and frame rate which make data analysis less accurate as it loses a lot of data. We show a comparison between different APIs for using webcams as eye trackers in table 8.2 and we reflect on their pros, cons, calibration, license, and data retrieval. Another issue is the eye tracking accuracy which is affected by the calibration step. Some eye trackers have one-point calibrations like Tobii Glasses 2 and some have

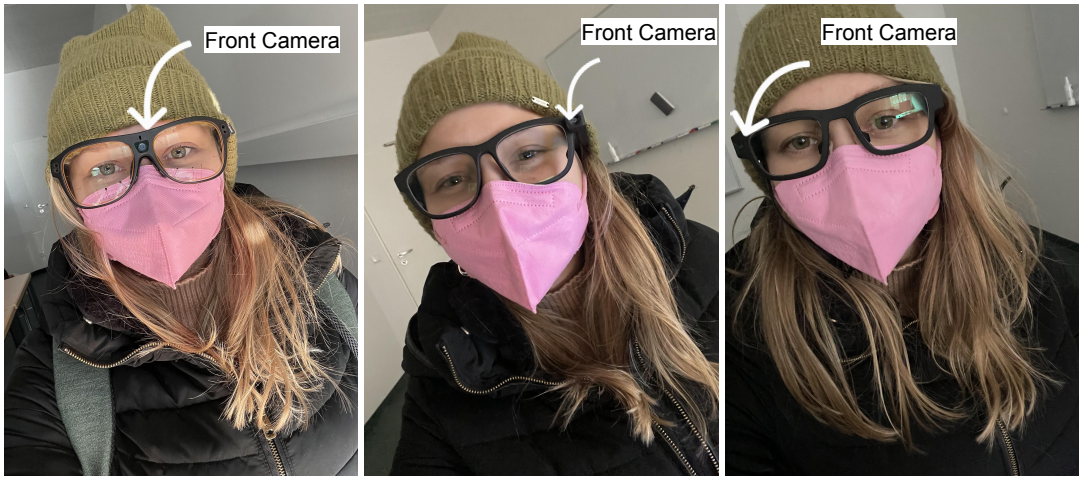


Figure 8.1: Comparison between three eye tracking glasses with AdHawk MindLink being the most comfortable and unobtrusive. At the time of the study, a face mask was required to enter public transport.

Table 8.2: A comparison between the most common Webcam-based eye tracking Software programs.

Software	Programming Language	Output Data	Calibration	Pros	Cons	License Required
TurkerGazer [388]	C#, Javascript	(x, y) coordinates, Timestamps	Self-calibrating	Real-time tracking' Easy setup	Non-commercial use only	No
XLabs [387]	Javscript, C++	None (user can not obtain the data)	Mouse-click calibration	Real-time tracking	No data analysis' No stimulus presentation	Yes (ID token)
WebGazer [135]	JavaScript8	(x, y) coordinates, Timestamps	Self-calibration from clicks and cursor movements	Real-time gaze prediction' Swappable components for eye detection' Gaze prediction models'	The prediction of gaze data is based on the interaction between the user and the web page	No
PyGaze [87]	Python	Pupil diameters	Self-calibration	Acts as a wrapper around several existing packages e.g., PyGame, Tobii SDK' compatible with different eye trackers'	Only pupil data obtained	No
OpenGazer [268]	C++, Python	(x, y) coordinates	Self-calibrating	Heat map' shows looking direction'	The user has to keep his/her head very still during the whole procedure}	No
Haytham tracker [251]	C#	Pupil center (x, y), pupil diameters, timestamps	Self-calibrating	Eye-based head gestures supported	Single eye recorded	No
GazeRecorder [125]	JavaScript	(X, Y) coordinates, timestamps	Self-calibrating	Easy setup real time tracking	Long calibration process	Yes

3 stages calibrations like GazeRecorder. The calibration step is crucial for tracking accuracy as we described in section 2.1.2, however, in most cases, it affects the usability of the prototype. Hence, designers should take care of the tradeoff between accuracy and usability.

Data Storage

Data storage is a critical consideration when gathering behavioral data. In our work, we investigated both storing and not storing the acquired task data - passwords in this case. We noticed that when the collected data was not saved but rather processed on the

run, users demonstrated more unique behaviors. This allowed users to behave naturally while also establishing trust between the researcher and the users, allowing for more distinctive behavior elicitation, and improving the quality of the collected gaze data. Another aspect is where to store the collected gaze data, client-side, provider-side, or on the cloud. This concern is more on the privacy and ethical consideration level as depending on where the gaze data is stored, it can have an impact on users' behavior and consent for data collection. Different research is being conducted to have a benchmark for collecting, storing, and processing gaze data in a privacy-preserving way.

8.1.3 Empirical Considerations

Finally, to complete the cycle of understanding users' behavior, we must discuss aspects affecting empirical evaluations. Here, we reflect on recruitment, participants, and screening.

There are different channels to recruit participants/users for the data collection such as online platforms like Amazon MTurk[19], and Prolific[20]. Although such platforms include a wide pool of participants with different backgrounds, experiences, etc., however, from our experiences, the collected data contains a lot of noise due to several interruptions, and most of the time they fail attention checks that we add throughout the study. This led to discarding their data and recruiting more participants which waste a lot of time and money. Other ways such as using university mailing lists, social media, or word of mouth usually provide better results as participants are more attentive however it can limit the pool of participants.

Another aspect affecting data collection is task proficiency. From our findings in chapter 6, we show that individual differences such as task proficiency affect the collected behavioral data. For example, if it is a typing task, then how frequently users use a similar keyboard is a huge factor as it will affect for example their gaze shift between the screen and keyboard affecting the overall model. Another example is if it's a reading task or email categorization, then how frequently users read on screens and how many emails they send or receive each day are crucial aspects of the participants' demographics.

[19] https://www.mturk.com/

[20] https://www.prolific.co/

8.2 Modeling User Behavior

The second step for employing gaze behavior in security systems is classification and modeling. Below we reflect on the different methodological, technical, and empirical considerations.

8.2.1 Methodological Considerations

There are different aspects to consider while modeling users' behavioral gaze data such as 1) classifiers, 2) areas of interest, and 3) feature set.

Classification Models

It is important to highlight that there is a wide spectrum of models that could be used for classification. Throughout our work, we only used off-the-self machine learning models however deep learning is also very promising when enough datasets are collected. In our work, we did not follow one classifier but rather ran different ones and picked up the best accuracy/AUC. An important aspect here is data labeling and having a ground truth to compare to. We manually labeled the data for tasks where were asked the participants about the passwords at a later stage as in chapter 7 and we had automatic labeling when participants were asked to create a weak or a strong password as reported in chapter 6. It is also advised to create user-dependent and user-independent models as with behavioral data, user-dependent models usually result in better accuracy due to the individual differences as we discussed earlier in section 8.1.3.

Areas of Interest (AOI)

Another important aspect of modeling users' behavior is highlighting the areas of interest. The AOIs differ from one aim to another according to what is being investigated. Such areas of interest can be and are not limited to certain interface areas, mobile phones, screens, users in a scene, etc. Mapping the areas of interest is different from one setup to another and from hardware to another. For example, collecting data in VR is easier to label the AOIs as you have control over every aspects participant see in the virtual environment. On the other side, running in-the-wild studies is the most complicated one as both the user and the environment are moving. In such scenarios, one can highlight the AOI using computer vision algorithms for object detection such as YOLO[21]. Running in the lab studies whether with stationary or wearable eye trackers lays in the middle between the two earlier extremes, where one can have AOI tagged in an interface or have QR codes to highlight the AOI in the study room/lab.

[21] https://github.com/jinfagang/yolov7_d2

Feature Set

There are different ways to create feature sets. First, if deep learning or neural networks were used, no need for the feature set it can be automatically generated and reported. However, with ML we need to provide a feature vector for the model to be trained on. One can get the most common features from literature where researchers report on them at the end of each modeling section or start with the straightforward ones we reported in the foundation chapter 2. In addition, one can also introduce new features as they found appropriate for example ratio between fixations on the keyboard and fixations on the screen which can reflect cognitive load and task proficiency. To have a better understanding of which features might work better, it is advised to run some statistical tests to highlight if there is an initial difference between your classes for each feature. This can indicate the data collected is different which might be promising to run classification on.

8.2.2 Technical Considerations

For the technical considerations, we highlight data cleaning and processing requirements as the main considerations.

Data Cleaning

After collecting the data, highlighting the areas of interest, and creating gaze features, some data cleaning have to be done before running the classifiers to not have bias. It is best to first visualize the created scene to understand what happened in it and where did the user look, etc. This can be done in different ways such as creating simple scripts with python or using sophisticated tools such as Ocumen studio[22]. After that, it will become better to understand where are the outliers and exclude them from the datasets. Such outliers can be just a few gaze points from certain participants, it can also be a few participants due to calibration errors/drafts or certain areas of the screen due to calibration issues.

Processing Requirements

This also depends on the amount of data collected and the classification model. Data processing also has an ethical aspect as where the processing is being done can have different privacy implications as participants might not consent that their data is being processed on e.g. a google virtual machine but rather on a university server.

[22] https://developer.tobii.com/xr/solutions/tobii-ocumen/

8.2.3 Empirical Considerations

For the empirical consideration, it is important to highlight that the amount of data needed is critical for the classification model. Few data might lead to overfitting. There is always this question of how much data is needed. We certainly do not have an answer for it, but we can comment that the amount of data needed for ML models is different than for deep learning and neural networks. It is also important that it is a factor in the number of participants. One can have short studies with a huge number of participants or the other way around with longer studies with a smaller pool of participants. In our studies, we did both designs, each one has its pros and cons. For example, having shorter studies attract more participants, however, it can run for longer periods to get enough datasets and one cannot run a user-dependent classifier with a such small mouth of data. On the other side, having longer studies enable running user-dependent classifiers, however, it attracts fewer people.

8.3 Providing Feedback

Finally, after creating the models, and when deploying the models in everyday life or for further research, we need to provide users with feedback. Here, we will reflect on considerations while creating feedback mechanisms.

8.3.1 Methodological Considerations

There are different considerations when it comes to the methodology of providing feedback, here we will highlight two of them, nudging and interventions.

Nudging

The term *"nudging"* was first used by Thaler and Sunstein [229]. hey argued that by exploiting what we know about systemic biases in decision-making, we may encourage people to make the best choices. Nudges are described as *"any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any option or significantly changing their economic incentive"*. For instance, nudge users to create strong passwords. Over the past 10 years, nudging has been applied in several domains, including HCI and security [142, 228, 68]. For instance, Harbach et al. [142] changed the permissions dialogue of the Google Play Store to nudge users to understand the consequences of giving different permissions to apps.

On the other side, one must consider the tradeoff between ″cost and benefit while designing nudges [13]. Some organizations require their staff members to change their passwords every 40 days. This costs the company a lot of money. On average, it takes

users five minutes to create a new password, which leads to a significant financial loss for the organization. Another factor is *nudging fatigue*. Users experience nudging fatigue as a result of everyday exposure to numerous nudges, which makes them ignore the majority of them. To prevent users from getting accustomed to them, a solution to this is to design dynamic nudges [86], which alter their colors, positioning, and text over time. Nudges can also use feat appeal or consequences to show the threat users might face with wrong password choices. It's important to emphasize the significant influence of the nudging strategy on password choice. This was already stated in the foundation chapter section 2.2.2, where we discussed the various visualizations of password strength and their impact on password choice. Also, the use of fear appeal for example can increase users' cognitive load affecting their gaze behavior. Therefore, the design of such a nudging technology might either work or backfire and result in poor decisions that put users at risk.

Interventions

Last but not least. A system that incorporates the prediction model should give different interventions based on the results of the classification models. The precision of the developed model should be used to choose an intervention. Figure 8.2 shows four different two-dimensional scenarios. The first dimension is the user's actual behavior, such as creating a weak password. The second dimension is the system's prediction, of whether the system suspects the user is entering a weak password.

True Negative There is no need for action because this is the best case, the users are not at risk and their behavior reflects the same.

True Positive In this case, the system recommends an intervention that optimally nudges the users to reconsider their decision.

False Positive The best use of nudging techniques is to encourage users to choose better security measures without attempting to force them into it because this negatively affects their user experience. Users should have the ability to activate and disable interventions as they see appropriate, with the hope that by the time they utilize the interventions, their learning curve will have increased and they will be able to independently make better decisions.

False Negative A system would not react in this case. As a result, the user might be at risk of reusing a password or responding to a phishing email. Intervention creators should train better classifiers to limit this case.

8.3.2 Technical Considerations

For the technical considerations, there are two main aspects feedback channel and feedback time.

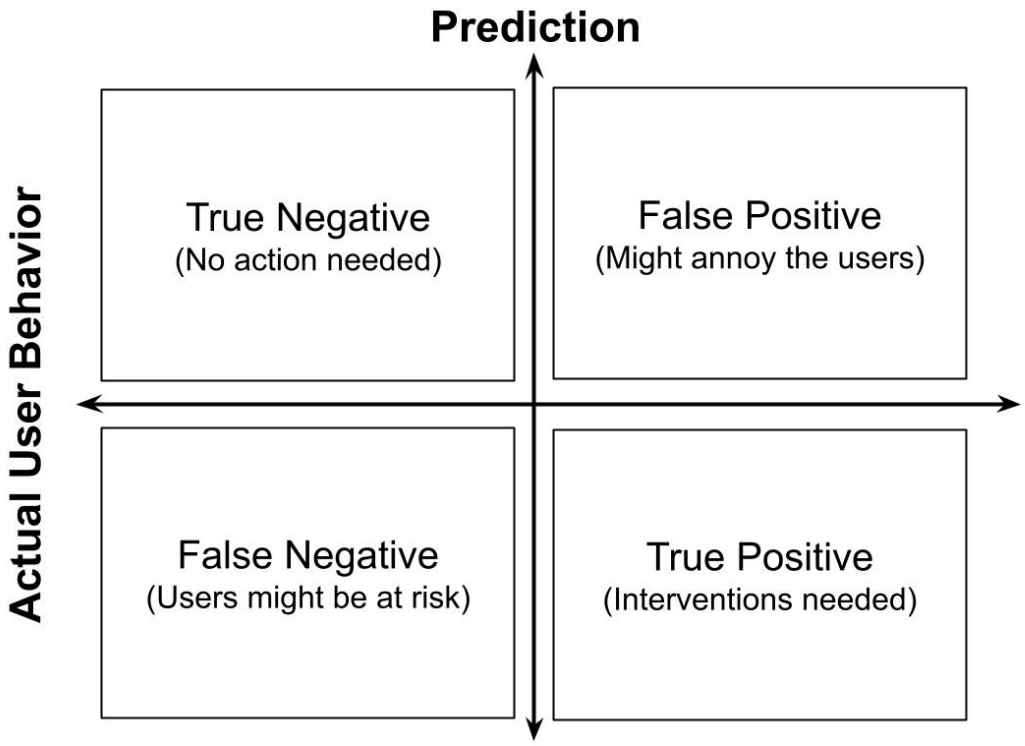


Figure 8.2: Interplay of Actual User Behavior and System Prediction.

Feedback Channel

For the feedback channel, nudges can take different forms, such as text, visual, haptic, or auditory nudges. This separates nudging into interface-dependent and interface-independent aspects. All of the methods previously described in section 8.3.1 and section 2.2.2 depend on the interface. However, one benefit of our strategy in this thesis is that it can identify password reuse and strength without having access to the user's data and on an interface-independent level. This opens up new possibilities for nudging approaches, such as sending nudges to users' smartwatches or smart glasses. We inspect that this will be the case for most of the behavioral eye-tracking models as they6 are less dependent on the interface itself.

Feedback Timing

Last but not least, the timing of the feedback is crucial. Literature has shown that users are less willing to change their decisions in a later stage [139], hence early-stage feedback can save a lot of time and help users make better decisions. For example, in this, we found that password reuse could be detected password reuse before users enter their password. Similarly, in social engineering, nudges can postpone sending emails or sharing fake news. For example, such nudge can be *"We triggered that you have a high cognitive workload at the moment, reflecting on your gaze and mouse movements*

we believe that the email you are responding to is phishing, hence, your response will be delayed by 3 hours for your confirmation later on" or *"Your behavioral analysis reveal that this post might contain misinformation, we will report to the administration for review and meanwhile it can't be shared"*.

8.3.3 Empirical Considerations

Finally, during data collection, we need to communicate data collection to the participants/users. This is mainly an ethical consideration, however, which way the communication is being conducted can affect users' gaze behavior. For example, one can be an always-on feedback where users can see where they are looking or just a small icon indicating that there is data being collected. This can come with the cost of distraction as we mentioned earlier, it also might be enough to inform users at the beginning of the study or with the consent for using the model, however, this needs deeper investigation.

8.4 Ethical Considerations

As noticed, we added ethical considerations as a vertical aspect. As we are collecting such sensitive behavioral measure that reflects different aspects of users [217], data collection, storage, and processing need to be done in a privacy-preserving way to protect users' privacy and prevent misuse or leakage of such sensitive data. Not only the users' privacy but also bystanders if they are visible in the environment and recorded with a wearable eye tracker world view. Due to not consenting as part of the study, bystanders' privacy should be considered, this can be done using face blurring for example to anonymize their identities during data collection. Unfortunately, there are very few works that look into bystanders' privacy in such scenarios.

8.5 Chapter Summary

In this chapter, we presented a framework for employing eye gaze behavior in security systems. We divided the process into 3 phases, understanding users' behavior, modeling users' behavior, and providing feedback. We reflected on three main considerations, methodological, technical, and empirical. Finally, we highlighted some ethical considerations to be taken care of during data collection. The aspects presented in this framework were derived through hands-on experience with conducting several behavioral eye-tracking studies in various security domains during the thesis period. This chapter answers **RQ4**: *What are the considerations for employing gaze behavior*

in security systems?. We hope with this framework that we provided researchers and practitioners with a guide on how behavioral eye gaze data could be collected, modeled, and communicated for enhancing security systems.

Chapter 9

Discussion

In this chapter, we set out to compile and discuss the findings and implications from all conducted user studies presented in this thesis. In particular, we provide a comprehensive analysis of our conclusions.

9.1 Discussion of Findings

In the introduction chapter of this thesis, we presented four research questions that this work aimed to explore during the course of the past four years (cf. Chapter 1, Section 1.1). In this section, we aim to present a synthesis of findings gathered from our evaluations of the research systems and probes, in light of these four questions.

RQ 1: How can security mechanisms benefit from users' eye gaze?

In chapter 3, we laid out the foundation of eye gaze usage in security and privacy applications. We grouped the work done in these areas into three main domains, 1) authentication, 2) privacy protection, and 3) improving security based on gaze monitoring. Each area of them could be further divided into subdomains. For example, work done in authentication was found to be in one of the directions, 1) explicit authentication, 2) implicit authentication and 3) multimodel authentication. Similarly, work done in privacy protection was also found to take two directions either 1) active privacy protection or 2) raising users' awareness. Finally, the last domain *improving security based on gaze monitoring* was found to be the least explored with very few works. From our findings, we can highlight two main challenges that are facing eye gaze in security applications, 1) Accuracy and Speed Trade-off, and 2) Privacy Implications of Eye Tracking.

Challenge 1: Accuracy and Speed Trade-off Implicit gaze-based security applications and gaze monitoring require highly accurate gaze estimates to be truly implicit and work without the user's intervention. To collect highly accurate gaze data, calibration is necessary [249]. For a long time, eye trackers required users to be very still and even required them to use chin rests [98]. While modern eye trackers afford to allow users to move around to an extent, they often need to be recalibrated every time the user's or the setup's state change significantly. But calibration introduces an overhead to the interaction process, and it is perceived to be tedious, unnatural, and time-consuming [370]. There is a lot of research directed at making calibration more of an implicit rather than an explicit procedure by, for example, making it part of the interaction process while reading text or watching videos [194, 286, 347].

Previous studies on implicit calibration addressed general use cases but not implicit authentication. This leaves room for future work on how to calibrate in a way to optimize the performance of implicit authentication and gaze monitoring applications. This requires first understanding the trade-off between calibration time and accuracy in implicit gaze-based authentication.

In contrast, some explicit gaze-based security applications do not require accurate gaze data. For example, many explicit schemes employ calibration-free gaze input methods like gestures [95, 192] and Pursuits [84, 196] which can perform accurately even when using inaccurate gaze data. These techniques require no calibration, as a result of which users can start the authentication process faster. However, calibration-free gaze input techniques often require longer entry times compared to other modalities. For example, in CueAuth [196], users spent 26.35 seconds authenticating using Pursuits, while touch input required only 3.73 seconds.

Challenge 2: Privacy Implications of Eye Tracking As we mentioned earlier in the implications 9.3 that using gaze behavior induces some ethical and privacy concerns that eye-tracking technology itself can be a threat to privacy. For example, a user's mobile device with eye tracking enabled could track the eyes of bystanders without their consent. This raises multiple questions. How can bystanders be made aware that a particular user's device can track their eyes? How can their consent be retrieved? And how can their privacy be protected if they do not wish their eyes to be tracked? Like many ubiquitous technologies [222], eye tracking can reveal many private users attributes [235] such as emotional valence [278], mind wandering [366], personality traits [155], and women's hormonal cycles [221]. Another important challenge is to securely store and process the gaze data without leaking it to third parties. This becomes more problematic if the tracking device uploads eye images to the cloud for processing rather than estimating gaze on the fly.

The privacy implications of pervasive eye tracking were discussed in recent work [187], and a few solutions to address this was proposed. For example, PrivacEYE [342] is a system that integrates a mechanical shutter into a wearable eye tracker. The shutter is activated when a bystander's face is in the camera's view. This protects the

privacy of bystanders and assures them that they are not being tracked. Another line of work applied differential privacy to gaze data by introducing noise to gaze data to prevent user identification without compromising the data's utility. Steil et al. [341] applied their differential privacy approach on gaze interfaces in virtual reality, while Liu et al. [237] applied theirs on heatmaps.

RQ 2: What is the influence of knowledge-based authentication on users' gaze behavior?

In chapter 4 and 5, we started an understanding of users' gaze behavior during authentication. We reflected on two knowledge-based authentication techniques 1) graphical passwords, and 2) text-based passwords. We found that during graphical passwords, users' gaze follows their hand movements which can reflect aspects of their internal state, concentration level, and cognitive load. Similarly, we studied the relation between cognitive load and password creation and found that creating strong passwords induces cognitive load on users reflected in their pupil diameter.

Highlighting that password creation induces cognitive load can be a great addition to workload-aware interfaces. It can be used to optimize users' workloads for better productivity. Moreover, 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 create unique memorable passwords [18]. On the other side, it urges researchers to create better authentication techniques that do not consume users' cognitive load and cannot be easily faked with images, videos, or deep fakes.

RQ 3: How can users' behavior during authentication be modeled?

We found in chapter 6 and 7 that users' behavior while creating and typing passwords is reflected in the gaze behavior and can be modeled to predict password strength and password reuse. However, the prediction depends on different aspects such as the input modality. As we compared users' behavior while entering passwords on a smartphone and on a laptop, the classifiers performed better on smartphone data. Due to the unique way users hold the phone at different distances to the screen, one hand or two hands, all of this affect their perception of the interface which is reflected in their gaze behavior affecting the classification accuracy. Similarly, the sensitivity of the protected data affects users' password choice which is reflected in their gaze behavior. This suggests that although generic classifiers such as user-independent, modality-independent, and interface-independent classifiers will have acceptable accuracies, personalized classifiers will better reflect individual user behavior which will have higher accuracies as they are customized. However, the question is, how

many classifiers are enough? Unfortunately, we do not have an answer for this, however, as a start, modality-dependent classifiers might be a good starting point until we achieve a fully independent classifier that can work across interfaces, modalities, and users.

Although Aristotle said *"The Whole is Greater than the Sum of its Parts"*, which is true in most cases, however, in the development of our models, this was not true. We show in this thesis that dissecting authentication into phases gave a better understanding of what happens during authentication, how users behave, when do they chose their passwords, and which aspects affect their behavior.

RQ 4: What are the considerations for employing gaze behavior in security systems?

From our work throughout the thesis, we draw a framework for deploying eye gaze behavior in security systems in chapter 8. We divide the framework into 3 main phases, understanding users' gaze, modeling users' gaze, and providing feedback. We reflect on different considerations namely methodological, technical, empirical, and ethical. Our framework shows that although employing behavioral gaze data to security systems provides insightful findings and gives a better understating of the task in hand, it comes with a list of implications to be taken care of, not to affect data collection. We showed that different interface aspects and interruptions affect users' gaze behavior consequently adding noise to the collected data and that interfaces should be adaptive. We also showed that the choice of eye tracker can also affect data collection as wearable eye trackers are perceived to be more obtrusive than screen-based eye trackers. Finally, with the recent survey conducted by Kroger et al. [217] reflecting on the different aspects that can be drawn from analyzing users' gaze data, we reflect on users' privacy and how the community needs to find ways to preserve users' privacy throughout collected gaze data. However, such approaches usually have a trade-off between the amount of data collection and classification accuracy. Here, it is important to highlight that the general population is not aware of the risk of collecting, analyzing, and sharing their gaze data according to Steil et al. [341]. This opens the door for different approaches to raise users' awareness to be able to choose which eye-tracking data to be collected and decline the rest having in mind the trade-off mentioned earlier.

9.2 Gaze Behavior Beyond Authentication

How our methodology and approach can be used for purposes other than authentication is a further topic for discussion. We believe that our approach can be extended in similar contexts because we found that creating passwords induces cognitive load

which could be modeled in our use case. Defending consumers from social engineering attacks, particularly phishing emails, is one possible direction. Phishing emails can take many various forms, but the most prevalent ones are emails with phishing links [236]. In such emails, attackers frequently pressure users to act within a certain time window. According to our hypothesis, the use of such approach will exert more cognitive stress on users, which will be reflected in their gaze movements. However, the issue is, how to track such activity in a natural environment without influencing users' gaze movements. Also, how users' eye behavior affected by reading such emails needs to be studied.

Eye gaze can also be useful in preventing users, particularly younger individuals, from being deceived or exposed to online harassment. Statistics done by Squicciarini et al. [337] shows that 61% of 13-17-year-old teenagers have a personal profile on social networking sites and around 44% of teenagers with profiles have been contacted by a stranger.In addition, MySpace removes 25k profiles each week for users under the age of 14, and between 2007 and 2009, 90k accounts belonging to registered sex offenders were deleted by the website. All of this emphasizes how essential user protection is. Although deep learning on social media post analysis can reveal a lot about each user [348], we believe that the use of behavioral aspects is a strong dimension. As opposed to evaluating users' or harassers' behavior after the fact, we can do it while they are conducting out the action. As a result, we can protect users from potential risk at an early stage. The work by Buker et al. investigated into this on the level of typing behavior. However, we believe that eye gaze behavior can reveal more about the attacker, improving the classification accuracy [62, 61].

Moreover, users' gaze behavior can reveal if they are struggling or having reading difficulties [310]. This for example can be used in privacy cookies and consents to show users simpler text, or provide more elaboration or even take the opportunity to raise their awareness about this privacy concern.

9.3 Ethics and User Privacy

Data privacy is the most crucial aspect of utilizing behavioral gaze data. We discovered from the literature that a different set of information, including the user's age, gender, and mental health, can be derived from gaze data [217]. Figure 9.1 shows the information that may be extracted from users' gaze data. Our work in this thesis contributes a new branch which can be called *Password Choices* highlighted in green. Under this branch and from this thesis we can add password strength and password reuse. But these are not the only security indications that we can extract we believe future work needs to investigate this direction in-depth and also investigate other security aspects such as responding to phishing emails. Moreover, Steil et al. [342] showed that users are ready to share their gaze data anonymously. However, they do

not know what can be extracted from their gaze data. This highlights the need to better communicate to the users the privacy risks from collecting and using eye gaze data similar to other behavioral privacy aspects or website cookies.

From our work, we can highlight three privacy aspects of eye gaze data, 1) how the data is saved, 2) where it is analyzed, and 3) user consent of what to be extracted from it. Currently, gaze data is saved as x and y positions or feature vectors, however, if this data is leaked then users might be at risk of identity theft for example. Hence, how gaze data is saved is a huge aspect that needs to be investigated. For example, we need to investigate the feasibility of gaze data hashing. In addition, recent work showed promising different ways of applying differential privacy mechanisms to gaze data [56, 237].

Another aspect is where the analysis is being carried out. Yet, most gaze-aware applications carry the analysis either on the cloud or on the server side. However, carrying the analysis on the users' side might build trust relation and also make users feel safer when using their gaze on the different interfaces. Finally, we need to raise users' awareness regarding their gaze data privacy and give them the possibility to accept or reject collecting certain types of eye gaze data. For example, users might accept collecting fixations only instead of the raw data. This can influence their experience however more research is needed here to balance data collection with users' experience.

9.4 Towards Gaze-based Behavioral Security Systems

Recent work showed that users' behavior is being heavily studied to better understand and protect them from online threats [265]. Because we can better protect users if we have a better understanding of their behavior, this underlines the importance of the research presented in this thesis. This thesis has provided examples of how to employ user gaze behavior to enhance knowledge-based passwords. Why, then, do we not currently have gaze behavioral systems? The answer to this question is twofold, 1) privacy side and 2) technology side. From a privacy perspective, we discussed in section 9.3 and 8.4 the different privacy implications of collecting, analyzing, modeling, and communicating behavioral gaze data. We also showed that different works are being held to create privacy-preserving gaze systems. Employing behavioral eye gaze into systems, and specifically, security systems is not yet ready until these problems are solved or until there is an ethical code that everyone must abide by. Additionally, users must be comfortable with the technology and provide their consent to data collection while being aware of the necessity, significance, and drawbacks of doing so.

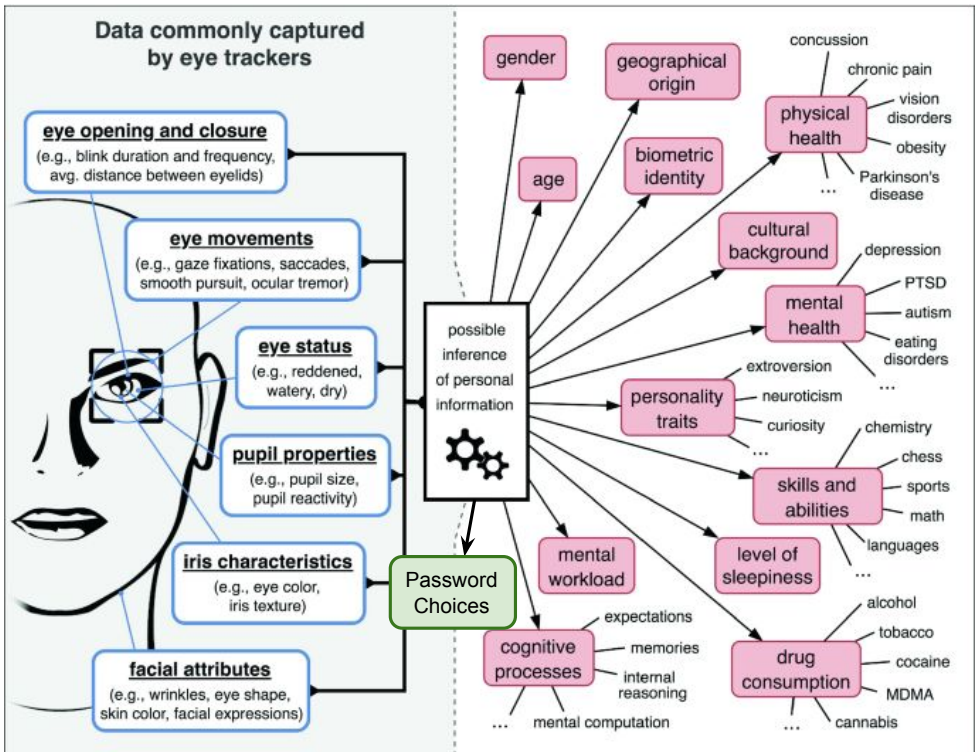


Figure 9.1: *Password Choices* branch addition to "What can be Extracted from Gaze Data" reported by Kröger et al. [217].

On the technological side, data collection is still needed for the long term and for multiple user profiles with different eye-tracking framerates and precision. Additionally, we must devise various interventions based on the precision of the prediction and find out how to take into account distractions and noise from the observed gaze patterns. By accommodating these two main aspects, employing behavioral eye gaze data in security systems can be achieved in the near future.

9.5 Chapter Summary

In this chapter, we discussed and reflected on the main findings of the thesis, discussed how gaze behavior can be used to enhance other security domains other than authentication, and finally, discussed different privacy implications of adopting eye gaze behavior.

Chapter 10

Conclusion

This thesis explored the use of gaze behavior as a tool to enhance existing security mechanisms with a focus on knowledge-based passwords. We utilized a user-centered design approach and a set of developed research probes to understand users' gaze behavior and enhance existing security mechanisms. In this chapter, we summarize our research contributions and provide a conclusion and future outlook for using gaze behavior in security mechanisms.

10.1 Summary of Contributions

Overall, this thesis provides four main contributions. First, we a systematic literature analysis of the use of eye gaze in security and privacy applications. Second, we present two systems that helped us better understand users' gaze behavior during knowledge-based authentication. Third, we present two prototypes with their respective machine-learning classifiers to predict password strength and reuse from users' gaze behavior. Finally, we present a set of implications reflecting on the study method, design, and ethics that will help others carry out behavioral gaze-based data collection studies.

10.1.1 Systematization Of Knowledge on the Usage of Gaze in Security

We contributed a systematic literature review in chapter 3 reflecting on how eye gaze has been used in the security field over the last 25 years. We analyzed more than 200 papers from top conferences and grouped the results into three main categories 1) authentication, 2) privacy protection, and 3) gaze monitoring. The findings of this

chapter motivate the rest of the work and why we focused on gaze monitoring to enhance security mechanisms.

10.1.2 Understanding Users' Gaze Behavior During Knowledge-based Authentication9n

In chapter 4 and 5, we presented two systems for eye gaze data collection while users are creating knowledge-based authentication. Each system is presented with its technical implementation and empirical evaluation. We found from chapter 4 that users follow their hands while performing touch gestures which we found to be promising to reflect on users' cognitive load. In chapter 5, we studied the relationship between cognitive load and password strength. Both systems with their conceptual, technical, and empirical evaluations built an understanding of users' gaze behavior during authentication.

10.1.3 Eye Gaze Modeling to Enhance Knowledge-based Authentication

We extended the previous findings by building two other systems that focused on quantifying users' gaze movements to predict password strength and reuse. Chapters 6 and 7 present these systems with their technical implementation, classification models, and empirical evaluations. The results show that users' gaze data can be modeled to train classifiers that predict password strength and reuse. This finding contributes to the user privacy domain where we show that users' passwords choice could be collected and modeled from their eye gaze data.

10.1.4 Framework for Employing Gaze Behavior in Security Systems

Evaluating the developed research probes uncovered different implications of running behavioral gaze-based studies. We highlighted a set of implications in a framework reflecting on the different aspects namely, methodological, technical, empirical and ethical for employing eye gaze behavior in security systems.

10.2 Directions for Future Research

Throughout this thesis, we came across a variety of ideas that could serve as the foundation for future research on the use of gaze in security systems.

10.2.1 Raising Users' Awareness and Implement Privacy Mitigation Techniques

A promising direction is looking into different nudging techniques that change according to user behavior. In section 8.3.1, we discussed several considerations to consider when creating nudging techniques. We believe that behavioral nudges that adjust to user profiles, such as personality (e.g., being more persuasive), cognitive load (e.g., delaying phishing email responses if a user is cognitively loaded), or heart rate (e.g., when under attack), can be a significant improvement to current security measures.

There are many approaches to increase users' awareness, including user training and a knowledge of the potential uses and ramifications of the data collected, in addition to nudges. We discussed in section 9.3, that one of the implications that needs to be addressed is user privacy and data collection. As a result, we propose another future direction that can look into various strategies for increasing users' awareness of their collected gaze data and teaching them which metrics to enable collecting and which not. Additionally, service providers should offer a variety of mitigation strategies, such as letting users pick which data are collected and offering additional information about how that data is used. For instance, gathering and maintaining raw eye gaze data is more critical than collecting fixations alone since various aspects of the user may be exposed from the raw gaze. As a result, service providers need to determine how to provide users varying levels of their service depending on the amount of data accessed to their services while also explaining to them how these levels would influence their usability, security, and privacy.

10.2.2 Exploring Different Physiological, Behavioral, and Cognitive Measurements

Human behavior is unique and reflects several aspects of an individual's internal state. Different physiological, behavioral, and cognitive measures, could be used to enhance existing security mechanisms. As we showed in this thesis that knowledge-based authentication induces cognitive load, and that was reflected in users' gaze behavior, we hypothesize that this could also be triggered by other behavioral and physiological metrics. For example, users' heart rate, body temperature, mouse movement, emotions,

facial expressions, personality traits, and facial temperature. Such measures have been found in the literature to be promising for accurately detecting users' internal states.

For example, heart rate has been shown to reflect users' anxiety [258, 385, 344], cognitive workload [282, 55], emotions [297, 403]. Similarly, thermal imaging has shown a promising direction of detecting users' emotions [206], cognitive workload [4] and multitasking [318]. Moreover, facial expressions were found to predict e.g. users' personality traits [124, 49], emotions [205, 164, 114], task performance [34, 326], cognitive profiles [115], cognitive load [156]. Last but not least, unconscious mouse movements were found to reveal users' cognitive load [133, 322], and stress level [379]. All of these aspects can be used with off-the-shelf equipment. For example, one can get a thermal camera for 250 Euros [23], and facial expressions and user emotions can be detected from web cameras. Heart rate can be detected from users' smartwatches which are becoming ubiquitous over time. Hence, a promising direction is to investigate leveraging the different physiological and cognitive measure to enhnace different security systems.

10.2.3 Exploring Different Security Domains

Another promising research direction is exploring other security mechanisms or security threats that could be enhanced using different behavioral aspects. Such threats can be other types of authentication techniques such as biometrics. Although fingerprints and FaceID are adopted by most users nowadays, they face different threats such as unauthorized access due to false positives [169], deep fakes using images and videos for FaceID [357], and unconscious authentication with fingerprints especially if the user is asleep [252]. Hence, we find behavioral aspects can add another level of security to such authentication techniques. Another direction is social engineering, we briefly discussed this in section 9.2, we believe that this field can benefit from behavioral data and not only eye gaze behavior but also typing and mouse movements behavior in addition to heart rate for example. This can also be extended beyond phishing emails to SMS phishing or vishing[24]. It can also be beneficial in detecting deception and online harassment to protect young adults as we discussed in section 9.2.

10.2.4 In-the-Wild Behavioral Gaze Data Collection

In all of our studies, we collected users' behavior in the lab. In addition, users knew they were in a study where their main aim was to create a password. However, in real life, password creation is not the primary task, it is usually a secondary task where

[23] Flir: `https://amzn.eu/d/21oIOMA`

[24] Voice phishing

users want to respond to an email however they have to change their password e.g. it exceeded its validity window. Hence, studying users' behavior at this stage might reveal different insights that were not noticed in the lab studies and will also extend the generalizability of our results when password creation is not the primary task.

10.3 Closing Remarks

This thesis investigates the use of eye gaze behavior to make password creation more secure. It gives password meters a new behavioral dimension. As we found that password creation induces cognitive load which changes users' gaze behavior reflecting different password aspects such as password strength and reuse. We also presented a framework for the usage of eye gaze behavior in security systems. We discussed how can eye gaze behavior be used to protect users beyond authentication and some ethical considerations to be tackled. From this thesis, we learned 3 main lessons, 1) Gaze behavior can be used to enhance security mechanisms, 2) Using gaze behavior allows us to evaluate security systems without accessing the content which is important for users' privacy, 3) Gaze behavior adds a new dimension to evaluate security systems and passwords in particular. We envision that with more eye trackers being embedded into existing devices and smart glasses, this approach will be a powerful tool for deploying aware-user interfaces (AWI).

V

BIBLIOGRAPHY

Bibliography

[1] J. Abbott, D. Calarco, and L. J. Camp. Factors influencing password reuse: A case study. In *Telecommunications Policy Research Conference on Communications, Information and Internet Policy (TPRC 46). DOI: http://dx. doi. org/10.2139/ssrn*, volume 3142270, 2018.

[2] Y. Abdelrahman, M. Khamis, S. Schneegass, and F. Alt. Stay cool! understanding thermal attacks on mobile-based user authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3751–3763, New York, NY, USA, 2017. ACM.

[3] Y. Abdelrahman, A. Khan, J. Newn, E. Velloso, S. Safwat, J. Bailey, A. Bulling, F. Vetere, and A. Schmidt. Classifying attention types with thermal imaging and eye tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):69:1–69:27, Sept. 2019.

[4] Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere. Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), sep 2017.

[5] Y. Abdrabou, Y. Abdelrahman, M. Khamis, and F. Alt. *Think Harder! Investigating the Effect of Password Strength on Cognitive Load during Password Creation*. Association for Computing Machinery, New York, NY, USA, 2021.

[6] Y. Abdrabou, M. Khamis, R. Eisa, S. Ismail, and A. Elmougy. Just gaze and wave: Exploring the use of gaze and gestures for shoulder-surfing resilient authentication. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19, New York, NY, USA, 2019. ACM.

[7] Y. Abdrabou, K. Pfeuffer, M. Khamis, and F. Alt. Gazelockpatterns: Comparing authentication using gaze and touch for entering lock patterns. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA 2020 Short Papers, New York, NY, USA, 2020. Association for Computing Machinery.

[8] Y. Abdrabou, S. R. Rivu, T. Ammar, J. Liebers, A. Saad, C. Liebers, U. Gruenefeld, P. Knierim, M. Khamis, V. Mäkelä, S. Schneegass, and F. Alt. Understanding shoulder surfer behavior and attack patterns using virtual reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, AVI 2022, New York, NY, USA, 2022. Association for Computing Machinery.

[9] Y. Abdrabou, J. Schütte, A. Shams, K. Pfeuffer, D. Buschek, M. Khamis, and F. Alt. "your eyes tell you have used this password before": Identifying password reuse from gaze and keystroke dynamics. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI 2022, New York, NY, USA, 2022. Association for Computing Machinery.

[10] Y. Abdrabou, A. Shams, M. O. Mantawy, A. Ahmad Khan, M. Khamis, F. Alt, and Y. Abdelrahman. *GazeMeter: Exploring the Usage of Gaze Behaviour to Enhance Password Assessments*. Association for Computing Machinery, New York, NY, USA, 2021.

[11] E. Abdulin and O. Komogortsev. Person verification via eye movement-driven text reading model. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, USA, Sep. 2015. IEEE.

[12] N. Abe, S. Yamada, and T. Shinzaki. A novel local feature for eye movement authentication. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, USA, Sep. 2016. IEEE.

[13] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, Y. Wang, and S. Wilson. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Comput. Surv.*, 50(3), Aug. 2017.

[14] A. Adams and M. A. Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, dec 1999.

[15] A. Adams, M. A. Sasse, and P. Lunt. Making passwords secure and usable. In *People and Computers XII*, pages 1–19. Springer, 1997.

[16] S. K. Ahern. Activation and intelligence: Pupillometric correlates of individual differences in cognitive abilities. 1979.

[17] K. Ahuja, R. Islam, V. Parashar, K. Dey, C. Harrison, and M. Goel. Eyespyvr: Interactive eye sensing using off-the-shelf, smartphone-based vr headsets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):57:1–57:10, July 2018.

[18] M. N. Al-Ameen, M. Wright, and S. Scielzo. 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*, CHI '15, page 2315–2324, New York, NY, USA, 2015. Association for Computing Machinery.

[19] W. Albert. Do web users actually look at ads? a case study of banner ads and eye-tracking technology. In *Proceedings of the 11th Annual Conference of the Usability Professionals' Association*, 2002.

[20] M. Ali, A. Anwar, I. Ahmed, T. Hashem, L. Kulik, and E. Tanin. Protecting mobile users from visual privacy attacks. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 1–4, New York, NY, USA, 2014. ACM.

[21] J. Almeida and F. Rodrigues. Facial expression recognition system for stress detection with deep learning. In *ICEIS (1)*, pages 256–263, 2021.

[22] H. Almoctar, P. Irani, V. Peysakhovich, and C. Hurter. Path word: A multimodal password entry method for ad-hoc authentication based on digits' shape and smooth pursuit eye movements. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, ICMI '18, pages 268–277, New York, NY, USA, 2018. ACM.

[23] I. M. A. Alseadoon. *The impact of users' characteristics on their ability to detect phishing emails*. PhD thesis, Queensland University of Technology, 2014.

[24] S. A. Alsuhibany, M. Almushyti, N. Alghasham, and F. Alkhudhayr. The impact of using different keyboards on free-text keystroke dynamics authentication for arabic language. *Information & Computer Security*, 2019.

[25] S. A. Alsuhibany, M. Almushyti, N. Alghasham, and F. Alkhudier. Analysis of free-text keystroke dynamics for arabic language using euclidean distance. In *2016 12th International Conference on Innovations in Information Technology (IIT)*, pages 1–6. IEEE, 2016.

[26] F. Alt and S. Schneegass. Beyond passwords—challenges and opportunities of future authentication. *IEEE Security & Privacy*, 2021.

[27] P. Andriotis, T. Tryfonas, and G. Oikonomou. Complexity metrics and user strength perceptions of the pattern-lock graphical authentication method. In *Proceedings of the Second International Conference on Human Aspects of Information Security, Privacy, and Trust - Volume 8533*, pages 115–126, Berlin, Heidelberg, 2014. Springer-Verlag.

[28] A. Antoinette and S. Buvaneswari. Real-time eye tracking for password authentication using matlab.

[29] M. Arianezhad, L. Camp, T. Kelley, and D. Stebila. Comparative eye tracking of experts and novices in web single sign-on. In *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*, CODASPY '13, pages 105–116, New York, NY, USA, 2013. ACM.

[30] M. Arianezhad, D. Stebila, and B. Mozaffari. Usability and security of gaze-based graphical grid passwords. In A. Adams, M. Brenner, and M. Smith, editors, *Financial Cryptography and Data Security*, pages 17–33, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[31] S. M. Asish, A. K. Kulshreshth, and C. W. Borst. User identification utilizing minimal eye-gaze features in virtual reality applications. In *Virtual Worlds*, volume 1, pages 42–61. MDPI, 2022.

[32] A. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. Smith. Smudge attacks on smartphone touch screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies*, WOOT '10, pages 1–7, Berkeley, CA, USA, 2010. USENIX Association.

[33] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies*, WOOTâ€™10, page 1â€“7, USA, 2010. USENIX Association.

[34] A. R. Babu, A. Rajavenkatanarayanan, J. R. Brady, and F. Makedon. Multimodal approach for cognitive task performance prediction from body postures, facial expressions and eeg signal. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*, MCPMD '18, New York, NY, USA, 2018. Association for Computing Machinery.

[35] M. Bâce, A. Saad, M. Khamis, S. Schneegass, and A. Bulling. Privacyscout: Assessing vulnerability to shoulder surfing on mobile devices. *Proceedings on Privacy Enhancing Technologies*, 1:21, 2022.

[36] T. Bader and J. Beyerer. Putting gaze into context: A framework for analyzing gaze behavior in interactive and dynamic environments. In *Proceedings of the 2010 Workshop on Eye Gaze in Intelligent Human Machine Interaction*, EGIHMI '10, page 20–27, New York, NY, USA, 2010. Association for Computing Machinery.

[37] T. Bafna, J. P. P. Hansen, and P. Baekgaard. Cognitive load during eye-typing. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA 2020 Full Papers, New York, NY, USA, 2020. Association for Computing Machinery.

[38] N. Banholzer, S. Feuerriegel, E. Fleisch, G. F. Bauer, and T. Kowatsch. Computer mouse movements as an indicator of work stress: Longitudinal observational field study. *J Med Internet Res*, 23(4):e27121, Apr 2021.

[39] G. Bargary, J. Bosten, P. Goodbourn, A. Lawrance-Owen, R. Hogg, and J. Mollon. Individual differences in human eye movements: An oculomotor signature? *Vision Research*, 141:157–169, 2017.

[40] R. Baskerville and M. Siponen. An information security meta-policy for emergent organizations. *Logistics Information Management*, 2002.

[41] A. Bayat and M. Pomplun. Biometric identification through eye-movement patterns. In D. N. Cassenti, editor, *Advances in Human Factors in Simulation and Modeling*, pages 583–594, Cham, 2018. Springer International Publishing.

[42] J. Beatty and B. Lucero-Wagoner. The pupillary system in t. cacioppo, l. tassinary & g. berntson (eds.), handbook of psychophsyiology (pp. 142-162), 2000.

[43] R. Bednarik, T. Kinnunen, A. Mihaila, and P. Fränti. Eye-movements as a biometric. In H. Kalviainen, J. Parkkinen, and A. Kaarna, editors, *Image Analysis*, pages 780–789, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[44] S. Berkovsky, R. Taib, I. Koprinska, E. Wang, Y. Zeng, J. Li, and S. Kleitman. Detecting personality traits using eye-tracking data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery.

[45] D. Best and A. Duchowski. A rotary dial for gaze-based pin entry. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, pages 69–76, New York, NY, USA, 2016. ACM.

[46] S. Bhagavatula, L. Bauer, and A. Kapadia. (how) do people change their passwords after a breach? *arXiv preprint arXiv:2010.09853*, 2020.

[47] O. S. Bhatti, M. Barz, and D. Sonntag. Eyelogin - calibration-free authentication method for public displays using eye gaze. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA 2021 Short Papers, New York, NY, USA, 2021. Association for Computing Machinery.

[48] R. Biedert, M. Frank, I. Martinovic, and D. Song. Stimuli for gaze based intrusion detection. In J. J. (Jong Hyuk) Park, V. Leung, C. Wang, and T. Shon, editors, *Future Information Technology, Application, and Service*, pages 757–763, Dordrecht, 2012. Springer Netherlands.

[49] J.-I. Biel, L. Teijeiro-Mosquera, and D. Gatica-Perez. Facetube: Predicting personality from facial expressions of emotion in online conversational video. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ICMI '12, page 53–56, New York, NY, USA, 2012. Association for Computing Machinery.

[50] BMW. Bmw camera keeps an eye on the driver. https://www.autonews.com/article/20181001/OEM06/181009966/bmw-camera-keeps-an-eye-on-the-driver, 2018. accessed 19 December 2019.

[51] A. Boehm, D. Chen, M. Frank, L. Huang, C. Kuo, T. Lolic, I. Martinovic, and D. Song. Safe: Secure authentication with face and eyes. In *2013 International Conference on Privacy and Security in Mobile Systems (PRISMS)*, pages 1–8, USA, June 2013. IEEE.

[52] J. Bonneau. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *2012 IEEE Symposium on Security and Privacy*, pages 538–552, 2012.

[53] J. Bonneau, C. Herley, P. Van Oorschot, and F. Stajano. Passwords and the evolution of imperfect authentication. *Communications of the ACM*, 58(7):78–87, July 2015. Copyright: Copyright 2018 Elsevier B.V., All rights reserved.

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

[55] V. Borisov, E. Kasneci, and G. Kasneci. Robust cognitive load detection from wrist-band sensors. *Computers in Human Behavior Reports*, 4:100116, 2021.

[56] E. Bozkir, O. Günlü, W. Fuhl, R. F. Schaefer, and E. Kasneci. Differential privacy for eye tracking with temporal correlations. *PLOS ONE*, 16(8):e0255979, aug 2021.

[57] J. B. Brookings, G. F. Wilson, and C. R. Swain. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology*, 42(3):361–377, 1996.

[58] F. Brudy, D. Ledo, S. Greenberg, and A. Butz. Is anyone looking? mitigating shoulder surfing on public displays through awareness and protection. In *Proceedings of The International Symposium on Pervasive Displays*, PerDis '14, pages 1:1–1:6, New York, NY, USA, 2014. ACM.

[59] D. Bruneau, M. A. Sasse, and J. D. McCarthy. The eyes never lie: the use of eyetracking data in hci research. ACM, 2002.

[60] T. Buch, A. Cotoranu, E. Jeskey, F. Tihon, and M. Villani. An enhanced keystroke biometric system and associated studies. *Proc. CSIS Research Day, Pace Univ*, 2008.

[61] A. Buker and A. Vinciarelli. Who is typing? automatic gender recognition from interactive textual chats using typing behaviour. In A. E. Hassanien, A. Darwish, S. M. Abd El-Kader, and D. A. Alboaneen, editors, *Enabling Machine Learning Applications in Data Science*, pages 3–15, Singapore, 2021. Springer Singapore.

[62] A. A. N. Buker, G. Roffo, and A. Vinciarelli. Type like a man! inferring gender from keystroke dynamics in live-chats. *IEEE Intelligent Systems*, 34(6):53–59, 2019.

[63] A. Bulling, F. Alt, and A. Schmidt. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 3011–3020, New York, NY, USA, 2012. Association for Computing Machinery.

[64] M. D. Byrne, J. R. Anderson, S. Douglass, and M. Matessa. Eye tracking the visual search of click-down menus. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 402–409, 1999.

[65] J. Campbell, W. Ma, and D. Kleeman. Impact of restrictive composition policy on user password choices. *Behaviour & Information Technology*, 30(3):379–388, 2011.

[66] V. Cantoni, C. Galdi, M. Nappi, M. Porta, and D. Riccio. Gant: Gaze analysis technique for human identification. *Pattern Recognition*, 48(4):1027–1038, 2015.

[67] V. Cantoni, T. Lacovara, M. Porta, and H. Wang. A study on gaze-controlled pin input with biometric data analysis. In *Proceedings of the 19th International Conference on Computer Systems and Technologies*, CompSysTech'18, pages 99–103, New York, NY, USA, 2018. Association for Computing Machinery.

[68] A. Caraban, E. Karapanos, D. Gonçalves, and P. Campos. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–15, New York, NY, USA, 2019. Association for Computing Machinery.

[69] J. R. Carl and R. S. Gellman. Human smooth pursuit: stimulus-dependent responses. *Journal of Neurophysiology*, 57(5):1446–1463, 1987. PMID: 3585475.

[70] D. Cazzato, M. Leo, A. Evangelista, and C. Distante. Soft biometrics by modeling temporal series of gaze cues extracted in the wild. In S. Battiato, J. Blanc-Talon, G. Gallo, W. Philips, D. Popescu, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, pages 391–402, Cham, 2015. Springer International Publishing.

[71] I. Chen, C.-C. Chang, et al. Cognitive load theory: An empirical study of anxiety and task performance in language learning. 2009.

[72] L. Chen, W. Cai, D. Yan, and S. Berkovsky. Eye-tracking-based personality prediction with recommendation interfaces. *User Modeling and User-Adapted Interaction*, pages 1–37, 2022.

[73] S. Chen and J. Epps. Using task-induced pupil diameter and blink rate to infer cognitive load. *Human–Computer Interaction*, 29(4):390–413, 2014.

[74] S. Chen, J. Epps, and F. Chen. Automatic and continuous user task analysis via eye activity. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI '13, page 57–66, New York, NY, USA, 2013. Association for Computing Machinery.

[75] S. Chen, J. Epps, N. Ruiz, and F. Chen. Eye activity as a measure of human mental effort in hci. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, IUI '11, page 315–318, New York, NY, USA, 2011. Association for Computing Machinery.

[76] E. Cherepovskaya and A. Lyamin. An evaluation of biometrie identification approach on low-frequency eye tracking data. In *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pages 123–128, USA, Jan 2017. IEEE.

[77] S. Chiasson, P. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *15th USENIX Security Symposium (USENIX Security 06)*, Vancouver, B.C. Canada, July 2006. USENIX Association.

[78] C. Code. C# - gestural pattern draw lock screen control (from android devices) [ro], Feb 2016.

[79] A. Constantinides, M. Belk, C. Fidas, and A. Pitsillides. An eye gaze-driven metric for estimating the strength of graphical passwords based on image hotspots. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, IUI '20, page 33–37, New York, NY, USA, 2020. Association for Computing Machinery.

[80] D. Conway, I. Dick, Z. Li, Y. Wang, and F. Chen. The effect of stress on cognitive load measurement. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson,

and M. Winckler, editors, *Human-Computer Interaction – INTERACT 2013*, pages 659–666, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[81] B. Coskun and C. Herley. Can "something you know" be saved? In *International Conference on Information Security*, pages 421–440. Springer, 2008.

[82] L. Cowen, L. Ball, and J. Delin. *An Eye Movement Analysis of Webpage Usability.* Springer-Verlag Ltd., 2002.

[83] N. Cuong, V. Dinh, and L. Ho. Mel-frequency cepstral coefficients for eye movement identification. In *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, pages 253–260, USA, Nov 2012. IEEE.

[84] D. Cymek, A. Venjakob, S. Ruff, O. Lutz, S. Hofmann, and M. Roetting. Entering pin codes by smooth pursuit eye movements. *Journal of Eye Movement Research*, 7(4), 2014.

[85] P. D, S. N, S. P, and S. V. N. Fdmca: A novel authentication technique using face detection and gaze-based morse code entry. In *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, pages 716–722, 2021.

[86] S. Dalecke and R. Karlsen. Designing dynamic and personalized nudges. In *Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics*, WIMS 2020, page 139–148, New York, NY, USA, 2020. Association for Computing Machinery.

[87] E. Dalmaijer. Pygaze: Open-source toolbox for eye tracking in python, Accessed on 2.Dec.2020.

[88] A. Dantcheva, N. Erdogmus, and J. Dugelay. On the reliability of eye color as a soft biometric trait. In *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 227–231, USA, Jan 2011. IEEE.

[89] A. Darwish and E. Bataineh. Eye tracking analysis of browser security indicators. In *2012 International Conference on Computer Systems and Industrial Informatics*, pages 1–6, USA, Dec 2012. IEEE.

[90] A. Darwish and M. Pasquier. Biometric identification using the dynamic features of the eyes. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, USA, Sep. 2013. IEEE.

[91] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang. The tangled web of password reuse. In *NDSS*, volume 14, pages 23–26, 2014.

[92] I. Das, R. Das, S. Singh, A. Banerjee, M. G. Mohiuddin, and A. Chowdhury. Design and implementation of eye pupil movement based pin authentication system. In *2020 IEEE VLSI DEVICE CIRCUIT AND SYSTEM (VLSI DCS)*, pages 1–6, 2020.

[93] N. Davies, S. Clinch, and F. Alt. Pervasive displays: Understanding the future of digital signage. *Synthesis Lectures on Mobile and Pervasive Computing*, 8(1):1–128, Apr. 2014.

[94] X. de Carné de Carnavalet and M. Mannan. From very weak to very strong: Analyzing password-strength meters. In *Network and Distributed System Security Symposium (NDSS 2014)*. Internet Society, 2014.

[95] A. De Luca, M. Denzel, and H. Hussmann. Look into my eyes!: Can you guess my password? In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, pages 7:1–7:12, New York, NY, USA, 2009. ACM.

[96] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch me once and i know it's you!: Implicit authentication based on touch screen patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 987–996, New York, NY, USA, 2012. ACM.

[97] A. De Luca, E. von Zezschwitz, and H. Hussmann. Vibrapass: Secure authentication based on shared lies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 913–916, New York, NY, USA, 2009. ACM.

[98] A. De Luca, R. Weiss, and H. Drewes. Evaluation of eye-gaze interaction methods for security enhanced pin-entry. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces*, OZCHI '07, pages 199–202, New York, NY, USA, 2007. ACM.

[99] A. De Luca, R. Weiss, H. Hussmann, and X. An. Eyepass - eye-stroke authentication for public terminals. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3003–3008, New York, NY, USA, 2008. ACM.

[100] V. Dhakal, A. M. Feit, P. O. Kristensson, and A. Oulasvirta. *Observations on Typing from 136 Million Keystrokes*, page 1–12. Association for Computing Machinery, New York, NY, USA, 2018.

[101] H. Drewes, M. Khamis, and F. Alt. Dialplates:enabling pursuits-based user interfaces with large target numbers. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, MUM '19, New York, NY, USA, 2019. ACM.

[102] H. Drewes and A. Schmidt. Interacting with the computer using gaze gestures. In C. Baranauskas, P. Palanque, J. Abascal, and S. Barbosa, editors, *Human-Computer Interaction – INTERACT 2007: 11th IFIP TC 13 International Conference, Rio de Janeiro, Brazil, September 10-14, 2007, Proceedings, Part II*, pages 475–488. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[103] A. Duchowski. T.,(2003),"eye tracking methodology: Theory and practice".

[104] A. T. Duchowski, K. Krejtz, I. Krejtz, C. Biele, A. Niedzielska, P. Kiefer, M. Raubal, and I. Giannopoulos. The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery.

[105] P. Dunphy, A. Fitch, and P. Olivier. Gaze-contingent passwords at the atm. In *4th Conference on Communication by Gaze Interaction (COGAIN)*, pages 59–62, Prague,Czech Republic, 2008. COGAIN.

[106] M. Dupuis and F. Khan. Effects of peer feedback on password strength. In *2018 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–9, 2018.

[107] R. Düzgün, N. Noah, P. Mayer, S. Das, and M. Volkamer. Sok: A systematic literature review of knowledge-based authentication on augmented reality head-mounted displays. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*, ARES '22, New York, NY, USA, 2022. Association for Computing Machinery.

[108] D. Eargle, J. Godfrey, H. Miao, S. Stevenson, R. Shay, B. Ur, and L. Cranor. You can do better—motivational statements in password-meter feedback. *Proc. SOUPS Posters*, 2015.

[109] S. Eberz, K. Rasmussen, V. Lenders, and I. Martinovic. Looks like eve: Exposing insider threats using eye movement biometrics. *ACM Transactions on Privacy and Security*, 19(1):1:1–1:31, June 2016.

[110] M. K. Eckstein, B. Guerra-Carrillo, A. T. Miller Singley, and S. A. Bunge. Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 25:69–91, 2017. Sensitive periods across development.

[111] S. Egelman, S. Jain, R. S. Portnoff, K. Liao, S. Consolvo, and D. Wagner. Are you ready to lock? In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS â€™14, page 750â€“761, New York, NY, USA, 2014. Association for Computing Machinery.

[112] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley. 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*, CHI '13, page 2379–2388, New York, NY, USA, 2013. Association for Computing Machinery.

[113] M. Eiband, M. Khamis, E. von Zezschwitz, H. Hussmann, and F. Alt. Understanding shoulder surfing in the wild: Stories from users and observers. In *Proceedings of the 35th Annual ACM Conference on Human Factors in Computing Systems*, CHI '17, pages 4254–4265, New York, NY, USA, 2017. ACM.

[114] P. Ekman and H. Oster. Facial expressions of emotion. *Annual review of psychology*, 30(1):527–554, 1979.

[115] M. El Kerdawy, M. El Halaby, A. Hassan, M. Maher, H. Fayed, D. Shawky, and A. Badawi. The automatic detection of cognition using eeg and facial expressions. *Sensors*, 20(12):3516, 2020.

[116] S. Eraslan, Y. Yesilada, and S. Harper. Eye tracking scanpath analysis on web pages: How many users? In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, pages 103–110, New York, NY, USA, 2016. ACM.

[117] C. Fidas, M. Belk, G. Hadjidemetriou, and A. Pitsillides. Influences of mixed reality and human cognition on picture passwords: An eye tracking study. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris, editors, *Human-Computer Interaction – INTERACT 2019*, pages 304–313, Cham, 2019. Springer International Publishing.

[118] D. Florencio and C. Herley. A large-scale study of web password habits. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, page 657–666, New York, NY, USA, 2007. Association for Computing Machinery.

[119] A. Forget, S. Chiasson, and R. Biddle. Input precision for gaze-based graphical passwords. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 4279–4284, New York, NY, USA, 2010. ACM.

[120] A. Forget, S. Chiasson, and R. Biddle. Shoulder-surfing resistance with eye-gaze entry in cued-recall graphical passwords. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1107–1110, New York, NY, USA, 2010. ACM.

[121] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle. Persuasion for stronger passwords: Motivation and pilot study. In H. Oinas-Kukkonen,

P. Hasle, M. Harjumaa, K. Segerståhl, and P. Øhrstrøm, editors, *Persuasive Technology*, pages 140–150, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[122] W. Fuhl, M. Tonsen, A. Bulling, and E. Kasneci. Pupil detection for head-mounted eye tracking in the wild: An evaluation of the state of the art. *Machine Vision and Applications*, 27(8):1275–1288, Nov 2016.

[123] C. Galdi, M. Nappi, D. Riccio, V. Cantoni, and M. Porta. A new gaze analysis based soft-biometric. In J. Carrasco-Ochoa, J. Martínez-Trinidad, J. Rodríguez, and G. di Baja, editors, *Pattern Recognition*, pages 136–144, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[124] M. Gavrilescu. Study on determining the big-five personality traits of an individual based on facial expressions. In *2015 E-Health and Bioengineering Conference (EHB)*, pages 1–6, 2015.

[125] GazeRecorder. Gazerecorder app, Accessed on 2.Dec.2020.

[126] D. R. Gentner. *Keystroke Timing in Transcription Typing*, pages 95–120. Springer New York, New York, NY, 1983.

[127] A. George and A. Routray. A score level fusion method for eye movement biometrics. *Pattern Recognition Letters*, 82:207–215, 2016.

[128] C. George, D. Buschek, A. Ngao, and M. Khamis. Gazeroomlock: Using gaze and head-pose to improve the usability and observation resistance of 3d passwords in virtual reality. In L. T. De Paolis and P. Bourdot, editors, *Augmented Reality, Virtual Reality, and Computer Graphics*, pages 61–81, Cham, 2020. Springer International Publishing.

[129] C. George, M. Khamis, D. Buschek, and H. Hussmann. Investigating the third dimension for authentication in immersive virtual reality and in the real world. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 277–285, USA, March 2019. IEEE.

[130] J. H. Goldberg and X. P. Kotval. Computer interface evaluation using eye movements: methods and constructs. *International journal of industrial ergonomics*, 24(6):631–645, 1999.

[131] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky. Eye tracking in web search tasks: Design implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, page 51–58, New York, NY, USA, 2002. Association for Computing Machinery.

[132] J. Gray, V. N. L. Franqueira, and Y. Yu. Forensically-sound analysis of security risks of using local password managers. In *2016 IEEE 24th International Requirements Engineering Conference Workshops (REW)*, pages 114–121, 2016.

[133] G. M. Grimes and J. S. Valacich. Mind over mouse: The effect of cognitive load on mouse movement behavior. In T. A. Carte, A. Heinzl, and C. Urquhart, editors, *Proceedings of the International Conference on Information Systems - Exploring the Information Frontier, ICIS 2015, Fort Worth, Texas, USA, December 13-16, 2015*. Association for Information Systems, 2015.

[134] I. Griswold-Steiner, Z. Fyke, M. Ahmed, and A. Serwadda. Morph-a-dope: Using pupil manipulation to spoof eye movement biometrics. In *2018 9th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pages 543–552, USA, Nov 2018. IEEE.

[135] B. H. Group. Webgazer.js, Accessed on 2.Dec.2020.

[136] D. Guitton and M. Volle. Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of Neurophysiology*, 58(3):427–459, 1987. PMID: 3655876.

[137] D. Gunetti, C. Picardi, and G. Ruffo. Keystroke analysis of different languages: A case study. In A. F. Famili, J. N. Kok, J. M. Peña, A. Siebes, and A. Feelders, editors, *Advances in Intelligent Data Analysis VI*, pages 133–144, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[138] T. Halevi, J. Lewis, and N. Memon. A pilot study of cyber security and privacy related behavior and personality traits. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, page 737–744, New York, NY, USA, 2013. Association for Computing Machinery.

[139] A. Hanamsagar, S. S. Woo, C. Kanich, and J. Mirkovic. *Leveraging Semantic Transformation to Investigate Password Habits and Their Causes*, page 1–12. Association for Computing Machinery, New York, NY, USA, 2018.

[140] S. T. Haque, M. Wright, and S. Scielzo. A study of user password strategy for multiple accounts. In *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*, CODASPY '13, page 173–176, New York, NY, USA, 2013. Association for Computing Machinery.

[141] M. Harbach, A. De Luca, and S. Egelman. The anatomy of smartphone unlocking: A field study of android lock screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4806–4817, New York, NY, USA, 2016. ACM.

[142] M. Harbach, M. Hettig, S. Weber, and M. Smith. Using personal examples to improve risk communication for security & privacy decisions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 2647–2656, New York, NY, USA, 2014. Association for Computing Machinery.

[143] M. Harbach, E. Von Zezschwitz, A. Fichtner, A. De Luca, and M. Smith. It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception. In *Proceedings of the Tenth USENIX Conference on Usable Privacy and Security*, SOUPS '14, pages 213–230, USA, 2014. USENIX Association.

[144] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[145] G. Hauland. Measuring team situation awareness by means of eye movement data. In *Human-Centered Computing*, pages 230–234. CRC Press, 2019.

[146] J. J. Hendrickson. Performance, preference, and visual scan patterns on a menu-based system: Implications for interface design. *SIGCHI Bull.*, 20(SI):217–222, mar 1989.

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

[148] J. B. Hirschberg, F. Enos, S. Benus, R. L. Cautin, M. Graciarena, and E. Shriberg. Personality factors in human deception detection: Comparing human to machine performance. 2006.

[149] T. Hirzle, J. Gugenheimer, F. Geiselhart, A. Bulling, and E. Rukzio. A design space for gaze interaction on head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 625:1–625:12, New York, NY, USA, 2019. ACM.

[150] B. Hoanca and K. Mock. Secure graphical password system for high traffic public areas. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, ETRA '06, pages 35–35, New York, NY, USA, 2006. ACM.

[151] B. Hoanca and K. Mock. Methods and systems for multiple factor authentication using gaze tracking and iris scanning, July 2011. US Patent 7,986,816.

[152] C. Holland and O. Komogortsev. Biometric identification via eye movement scanpaths in reading. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–8, USA, Oct 2011. IEEE.

[153] C. Holland and O. Komogortsev. Biometric verification via complex eye movements: The effects of environment and stimulus. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 39–46, USA, Sep. 2012. IEEE.

[154] C. Holland and O. Komogortsev. Complex eye movement pattern biometrics: The effects of environment and stimulus. *IEEE Transactions on Information Forensics and Security*, 8(12):2115–2126, Dec 2013.

[155] S. Hoppe, T. Loetscher, S. A. Morey, and A. Bulling. Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience*, 12:105, 2018.

[156] M. S. Hussain, R. A. Calvo, and F. Chen. Automatic cognitive load detection from face, physiology, task performance and fusion during affective interference. *Interacting with computers*, 26(3):256–268, 2014.

[157] S. Hussain, S. Chen, R. A. Calvo, and F. Chen. Classification of cognitive load from task performance & multichannel physiology during affective changes. In *Conference on Multimodal Interaction*, pages 1–4, 2011.

[158] A. Huth, M. Orlando, and L. Pesante. Password security, protection, and management. *United States Computer Emergency Readiness Team*, 2012.

[159] C. S. Ikehara and M. Crosby. Assessing cognitive load with physiological sensors. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pages 295a–295a, 2005.

[160] P. G. Inglesant and M. A. Sasse. The true cost of unusable password policies: Password use in the wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 383–392, New York, NY, USA, 2010. Association for Computing Machinery.

[161] S. T. Iqbal, P. D. Adamczyk, X. S. Zheng, and B. P. Bailey. 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*, CHI '05, page 311–320, New York, NY, USA, 2005. Association for Computing Machinery.

[162] ISO9241-210:2019. Ergonomics of human-system interaction - part 210 human-centred design for interactive systems, 2019.

[163] Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework. Standard, International Organization for Standardization, Geneva, CH, Apr 2006.

[164] C. E. Izard. Facial expressions and the regulation of emotions. *Journal of personality and social psychology*, 58(3):487, 1990.

[165] L. IzsÓ and E. Láng. Heart period variability as mental effort monitor in human computer interaction. *Behaviour & Information Technology*, 19(4):297–306, 2000.

[166] R. Jacob. What you look at is what you get: Eye movement-based interaction techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 11–18, New York, NY, USA, 1990. ACM.

[167] R. J. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*, pages 573–605. Elsevier, 2003.

[168] J. L. Jenkins, M. Grimes, J. G. Proudfoot, and P. B. Lowry. Improving password cybersecurity through inexpensive and minimally invasive means: Detecting and deterring password reuse through keystroke-dynamics monitoring and just-in-time fear appeals. *Information Technology for Development*, 20(2):196–213, 2014.

[169] M. Johnson. Biometrics and the threat to civil liberties. *Computer*, 37(4):90–92, 2004.

[170] M. Juhola, Y. Zhang, and J. Rasku. Biometric verification of a subject through eye movements. *Computers in Biology and Medicine*, 43(1):42–50, 2013.

[171] M. A. Just and P. A. Carpenter. Eye fixations and cognitive processes. *Cognitive psychology*, 8(4):441–480, 1976.

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

[173] C. Jyotsna and J. Amudha. Eye gaze as an indicator for stress level analysis in students. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1588–1593. IEEE, 2018.

[174] P. Kasprowski. The impact of temporal proximity between samples on eye movement biometric identification. In K. Saeed, R. Chaki, A. Cortesi, and S. Wierzchoń, editors, *Computer Information Systems and Industrial Management*, pages 77–87, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[175] P. Kasprowski and K. Harężlak. Cheap and easy pin entering using eye gaze. *Annales Universitatis Mariae Curie-Skłodowska. Sectio AI, Informatica*, 14(1):75–84, 2014.

[176] P. Kasprowski and K. Harężlak. Using dissimilarity matrix for eye movement biometrics with a jumping point experiment. In I. Czarnowski, A. M. Caballero, R. Howlett, and L. Jain, editors, *Intelligent Decision Technologies 2016*, pages 83–93, Cham, 2016. Springer International Publishing.

[177] P. Kasprowski and K. Harężlak. Biometric identification using gaze and mouse dynamics during game playing. In S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, and D. Kostrzewa, editors, *Beyond Databases, Architectures and Structures. Facing the Challenges of Data Proliferation and Growing Variety*, pages 494–504, Cham, 2018. Springer International Publishing.

[178] P. Kasprowski and K. Harężlak. Fusion of eye movement and mouse dynamics for reliable behavioral biometrics. *Pattern Analysis and Applications*, 21(1):91–103, Feb 2018.

[179] P. Kasprowski and J. Ober. Eye movements in biometrics. In D. Maltoni and A. Jain, editors, *Biometric Authentication*, pages 248–258, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[180] K. Kassem, J. Salah, Y. Abdrabou, M. Morsy, R. El-Gendy, Y. Abdelrahman, and S. Abdennadher. 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*, MUM '17, page 273–278, New York, NY, USA, 2017. Association for Computing Machinery.

[181] C. Katsini, Y. Abdrabou, G. E. Raptis, M. Khamis, and F. Alt. *The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions*, page 1–21. Association for Computing Machinery, New York, NY, USA, 2020.

[182] C. Katsini, C. Fidas, M. Belk, G. Samaras, and N. Avouris. A human-cognitive perspective of users' password choices in recognition-based graphical authentication. *International Journal of Human-Computer Interaction*, 25(19):1800–1812, 2019.

[183] C. Katsini, C. Fidas, G. Raptis, M. Belk, G. Samaras, and N. Avouris. Eye gaze-driven prediction of cognitive differences during graphical password composition. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, pages 147–152, New York, NY, USA, 2018. ACM.

[184] C. Katsini, C. Fidas, G. Raptis, M. Belk, G. Samaras, and N. Avouris. Influences of human cognition and visual behavior on password strength during picture password composition. In *Proceedings of the 2018 CHI Conference on Human*

Factors in Computing Systems, CHI '18, pages 87:1–87:14, New York, NY, USA, 2018. ACM.

[185] C. Katsini, G. Raptis, C. Fidas, and N. Avouris. Towards gaze-based quantification of the security of graphical authentication schemes. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18, pages 17:1–17:5, New York, NY, USA, 2018. ACM.

[186] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *2012 IEEE Symposium on Security and Privacy*, pages 523–537, 2012.

[187] M. Khamis, F. Alt, and A. Bulling. The past, present, and future of gaze-enabled handheld mobile devices: Survey and lessons learned. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18, New York, NY, USA, 2018. ACM.

[188] M. Khamis, F. Alt, M. Hassib, E. von Zezschwitz, R. Hasholzner, and A. Bulling. Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2156–2164, New York, NY, USA, 2016. ACM.

[189] M. Khamis, L. Bandelow, S. Schick, D. Casadevall, A. Bulling, and F. Alt. They are all after you: Investigating the viability of a threat model that involves multiple shoulder surfers. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, MUM '17, pages 31–35, New York, NY, USA, 2017. ACM.

[190] M. Khamis, M. Eiband, M. Zürn, and H. Hussmann. Eyespot: Leveraging gaze to protect private text content on mobile devices from shoulder surfing. *Multimodal Technologies and Interaction*, 2(3), 2018.

[191] M. Khamis, R. Hasholzner, A. Bulling, and F. Alt. Gtmopass: Two-factor authentication on public displays using gazetouch passwords and personal mobile devices. In *Proceedings of the 6th International Symposium on Pervasive Displays*, PerDis '17, New York, NY, USA, 2017. ACM.

[192] M. Khamis, M. Hassib, E. von Zezschwitz, A. Bulling, and F. Alt. Gazetouchpin: Protecting sensitive data on mobile devices using secure multimodal authentication. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ICMI 2017, pages 446–450, New York, NY, USA, 2017. ACM.

[193] M. Khamis, C. Oechsner, F. Alt, and A. Bulling. Vrpursuits: Interaction in virtual reality using smooth pursuit eye movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, AVI '18, pages 18:1–18:8, New York, NY, USA, 2018. ACM.

[194] M. Khamis, O. Saltuk, A. Hang, K. Stolz, A. Bulling, and F. Alt. Textpursuits: Using text for pursuits-based interaction and calibration on public displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 274–285, New York, NY, USA, 2016. ACM.

[195] M. Khamis, T. Seitz, L. Mertl, A. Nguyen, M. Schneller, and Z. Li. Passquerade: Improving error correction of text passwords on mobile devices by using graphic filters for password masking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 686:1–686:8, New York, NY, USA, 2019. ACM.

[196] M. Khamis, L. Trotter, V. Mäkelä, E. v. Zezschwitz, J. Le, A. Bulling, and F. Alt. Cueauth: Comparing touch, mid-air gestures, and gaze for cue-based authentication on situated displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):174:1–174:22, Dec. 2018.

[197] M. Khamis, L. Trotter, M. Tessmann, C. Dannhart, A. Bulling, and F. Alt. Eyevote in the wild: Do users bother correcting system errors on public displays? In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, MUM '16, pages 57–62, New York, NY, USA, 2016. ACM.

[198] I. A. Khan, W.-P. Brinkman, N. Fine, and R. M. Hierons. Measuring personality from keyboard and mouse use. In *Proceedings of the 15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction*, ECCE '08, New York, NY, USA, 2008. Association for Computing Machinery.

[199] P. Kiefer, I. Giannopoulos, A. Duchowski, and M. Raubal. Measuring cognitive load for map tasks through pupil diameter. In J. A. Miller, D. O'Sullivan, and N. Wiegand, editors, *Geographic Information Science*, pages 323–337, Cham, 2016. Springer International Publishing.

[200] T. Kinnunen, F. Sedlak, and R. Bednarik. Towards task-independent person authentication using eye movement signals. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 187–190, New York, NY, USA, 2010. ACM.

[201] P. A. Kirschner. Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction*, 12(1):1–10, 2002.

[202] P. A. Kirschner and F. Kirschner. *Mental Effort*, pages 2182–2184. Springer US, Boston, MA, 2012.

[203] J. Klingner. *Measuring Cognitive Load During Visual Tasks by Combining Pupillometry and Eye Tracking*. Ph.d. dissertation, Stanford University, Department of Computer Science, May 2010.

[204] T. Kocejko and J. Wtorek. Gaze pattern lock for elders and disabled. In E. Piętka and J. Kawa, editors, *Information Technologies in Biomedicine*, pages 589–602, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[205] C. G. Kohler, T. Turner, N. M. Stolar, W. B. Bilker, C. M. Brensinger, R. E. Gur, and R. C. Gur. Differences in facial expressions of four universal emotions. *Psychiatry Research*, 128(3):235–244, 2004.

[206] A. Kolli, A. Fasih, F. A. Machot, and K. Kyamakya. Non-intrusive car driver's emotion recognition using thermal camera. In *Proceedings of the Joint INDS'11 & ISTET'11*, pages 1–5, 2011.

[207] 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*, CHI '11, page 2595–2604, New York, NY, USA, 2011. Association for Computing Machinery.

[208] O. Komogortsev, C. Holland, and A. Karpov. Template aging in eye movement-driven biometrics. In *Biometric and Surveillance Technology for Human and Activity Identification XI*, volume 9075, USA, 2014. SPIE.

[209] O. Komogortsev, S. Jayarathna, C. Aragon, and M. Mahmoud. Biometric identification via an oculomotor plant mathematical model. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 57–60, New York, NY, USA, 2010. ACM.

[210] O. Komogortsev, A. Karpov, and C. Holland. CUE: Counterfeit-resistant usable eye movement-based authentication via oculomotor plant characteristics and complex eye movement patterns. In *Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX*, volume 8371, USA, 2012. SPIE.

[211] O. Komogortsev, A. Karpov, C. Holland, and H. Proença. Multimodal ocular biometrics approach: A feasibility study. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 209–216, USA, Sep. 2012. IEEE.

[212] O. Komogortsev, A. Karpov, L. Price, and C. Aragon. Biometric authentication via oculomotor plant characteristics. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 413–420, USA, March 2012. IEEE.

[213] O. Komogortsev and I. Rigas. Bioeye 2015: Competition on biometrics via eye movements. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, USA, Sep. 2015. IEEE.

[214] T. Kosch, M. Hassib, D. Buschek, and A. Schmidt. 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*, CHI EA '18, page 1–6, New York, NY, USA, 2018. Association for Computing Machinery.

[215] T. Kosch, M. Hassib, R. Reutter, and F. Alt. Emotions on the go: Mobile emotion assessment in real-time using facial expressions. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI 2020, New York, NY, USA, 2020. Association for Computing Machinery.

[216] T. Kosch, M. Hassib, P. W. Woundefinedniak, D. Buschek, and F. Alt. Your eyes tell: Leveraging smooth pursuit for assessing cognitive workload. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery.

[217] J. L. Kröger, O. H.-M. Lutz, and F. Müller. What does your gaze reveal about you? on the privacy implications of eye tracking. In *IFIP International Summer School on Privacy and Identity Management*, pages 226–241. Springer, 2019.

[218] A. Kumar, L.-H. Lee, J. Chauhan, X. Su, M. A. Hoque, S. Pirttikangas, S. Tarkoma, and P. Hui. Passwalk: Spatial authentication leveraging lateral shift and gaze on mobile headsets. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 952–960, New York, NY, USA, 2022. Association for Computing Machinery.

[219] C. Kumar, D. Akbari, R. Menges, S. MacKenzie, and S. Staab. Touchgazepath: Multimodal interaction with touch and gaze path for secure yet efficient pin entry. In *2019 International Conference on Multimodal Interaction*, ICMI â€™19, page 329â€“338, New York, NY, USA, 2019. Association for Computing Machinery.

[220] M. Kumar, T. Garfinkel, D. Boneh, and T. Winograd. Reducing shoulder-surfing by using gaze-based password entry. In *Proceedings of the 3rd Symposium on Usable Privacy and Security*, SOUPS '07, pages 13–19, New York, NY, USA, 2007. ACM.

[221] B. Laeng and L. Falkenberg. Women's pupillary responses to sexually significant others during the hormonal cycle. *Hormones and Behavior*, 52(4):520–530, Nov. 2007.

[222] M. Langheinrich. Privacy in ubiquitous computing. In J. Krumm, editor, *Ubiquitous Computing Fundamentals*. Chapman & Hall / CRC, NW, USA, 2009.

[223] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos. Toward emotion recognition from physiological signals in the wild: approaching the methodological issues in real-life data collection. *Frontiers in psychology*, 11:1111, 2020.

[224] K. LaRubbio, J. Wright, B. David-John, A. Enqvist, and E. Jain. Who do you look like? - gaze-based authentication for workers in vr. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 744–745, 2022.

[225] P. Lawson, C. J. Pearson, A. Crowson, and C. B. Mayhorn. Email phishing and signal detection: How persuasion principles and personality influence response patterns and accuracy. *Applied ergonomics*, 86:103084, 2020.

[226] S. E. Lea, P. Fischer, and K. M. Evans. The psychology of scams: Provoking and committing errors of judgement. 2009.

[227] D. LeBlanc, A. Forget, and R. Biddle. Guessing click-based graphical passwords by eye tracking. In *2010 Eighth International Conference on Privacy, Security and Trust*, pages 197–204, USA, Aug 2010. IEEE.

[228] M. K. Lee, S. Kiesler, and J. Forlizzi. Mining behavioral economics to design persuasive technology for healthy choices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 325–334, New York, NY, USA, 2011. Association for Computing Machinery.

[229] T. C. Leonard. Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness, 2008.

[230] C. Li, J. Xue, C. Quan, J. Yue, and C. Zhang. Biometric recognition via texture features of eye movement trajectories in a visual searching task. *PLOS ONE*, 13(4):1–24, Apr 2018.

[231] N. Li, Q. Wu, J. Liu, W. Hu, B. Qin, and W. Wu. Eyesec: A practical shoulder-surfing resistant gaze-based authentication system. In J. K. Liu and P. Samarati, editors, *Information Security Practice and Experience*, pages 435–453, Cham, 2017. Springer International Publishing.

[232] Z. Li, W. He, D. Akhawe, and D. Song. The Emperor's new password manager: Security analysis of web-based password managers. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 465–479, San Diego, CA, Aug. 2014. USENIX Association.

[233] Z. Li, M. Li, P. Mohapatra, J. Han, and S. Chen. itype: Using eye gaze to enhance typing privacy. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, USA, May 2017. IEEE.

[234] J. Liebers, P. Horn, C. Burschik, U. Gruenefeld, and S. Schneegass. Using gaze behavior and head orientation for implicit identification in virtual reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, VRST '21, New York, NY, USA, 2021. Association for Computing Machinery.

[235] D. Liebling and S. Preibusch. Privacy considerations for a pervasive eye tracking world. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 1169–1177, New York, NY, USA, 2014. ACM.

[236] T. Lin, D. E. Capecci, D. M. Ellis, H. A. Rocha, S. Dommaraju, D. S. Oliveira, and N. C. Ebner. Susceptibility to spear-phishing emails: Effects of internet user demographics and email content. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(5):1–28, 2019.

[237] A. Liu, L. Xia, A. Duchowski, R. Bailey, K. Holmqvist, and E. Jain. Differential privacy for eye-tracking data. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19, pages 28:1–28:10, New York, NY, USA, 2019. ACM.

[238] D. Liu, B. Dong, X. Gao, and H. Wang. Exploiting eye tracking for smartphone authentication. In T. Malkin, V. Kolesnikov, A. B. Lewko, and M. Polychronakis, editors, *Applied Cryptography and Network Security*, pages 457–477, Cham, 2015. Springer International Publishing.

[239] S. Liu, J. Wilson, and Y. Xia. Eye gazing passcode generation crossing augmented reality (ar) and virtual reality (vr) devices, Nov. 2017. US Patent 9,824,206.

[240] D. Lockton. Attitudes, meaning, emotion and motivation in design for behaviour change. *Available at SSRN 2123495*, 2012.

[241] M. Loge, M. Duermuth, and L. Rostad. On user choice for android unlock patterns. In *European Workshop on Usable Security, ser. EuroUSEC*, volume 16, 2016.

[242] N. Lord. Uncovering password habits: Are users' password security habits improving? Accessed: 2019-03-19.

[243] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pages 4768–4777, 2017.

[244] A. Lyamin and E. Cherepovskaya. Biometric student identification using low-frequency eye tracker. In *2015 9th International Conference on Application of Information and Communication Technologies (AICT)*, pages 191–195, Red Hook, NY, USA, Oct 2015. IEEE.

[245] A. Lyamin and E. Cherepovskaya. An approach to biometric identification by using low-frequency eye tracker. *IEEE Transactions on Information Forensics and Security*, 12(4):881–891, April 2017.

[246] Z. Ma, X. Wang, R. Ma, Z. Wang, and J. Ma. Integrating gaze tracking and head-motion prediction for mobile device authentication: A proof of concept. *Sensors*, 18(9):2894, Aug 2018.

[247] A. Maeder and C. Fookes. A visual attention approach to personal identification. In *Eighth Australian and New Zealand Intelligent Information Systems Conference*, ANZIIS 2003, pages 55–60, Brisbane, QLD, December 2003. The Australian Pattern Recognition Society.

[248] A. Maeder, C. Fookes, and S. Sridharan. Gaze based user authentication for personal computer applications. In *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004*, pages 727–730, USA, Oct 2004. IEEE.

[249] P. Majaranta and A. Bulling. Eye tracking and eye-based human–computer interaction. In S. Fairclough and K. Gilleade, editors, *Advances in Physiological Computing*, pages 39–65. Springer London, London, 2014.

[250] V. Mäkelä, M. Khamis, L. Mecke, J. James, M. Turunen, and F. Alt. Pocket transfers: Interaction techniques for transferring content from situated displays to mobile devices. In *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems*, CHI '18, New York, NY, USA, 2018. ACM.

[251] D. Mardanbegi. Haytham gaze tracker, Accessed on 2.Dec.2020.

[252] M. V. Martin, K. Jóhannsdottir, G. B. Reynaga, J. Tashiro, and M. A. Garcia-Ruiz. Unconscious mind: Authenticating with something you don't know? or just an infallible liveness test? In *CCECE 2010*, pages 1–6, 2010.

[253] R. Martins and J. Carvalho. Eye blinking as an indicator of fatigue and mental load—a systematic review. *Occupational safety and hygiene III*, 10:231–235, 2015.

[254] F. Mathis, J. H. Williamson, K. Vaniea, and M. Khamis. Fast and secure authentication in virtual reality using coordinated 3d manipulation and pointing. *ACM Trans. Comput.-Hum. Interact.*, 28(1), jan 2021.

[255] Y. Matsubara, T. Samura, H. Nishimura, et al. Keyboard dependency of personal identification performance by keystroke dynamics in free text typing. *Journal of Information Security*, 6(03):229, 2015.

[256] C. Mello-Thoms, C. F. Nodine, and H. L. Kundel. What attracts the eye to the location of missed and reported breast cancers? In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, page 111–117, New York, NY, USA, 2002. Association for Computing Machinery.

[257] M. Mihajlov, M. Trpkova, and S. Arsenovski. Eye tracking recognition-based graphical authentication. In *2013 7th International Conference on Application of Information and Communication Technologies*, pages 1–5, USA, Oct 2013. IEEE.

[258] D. Miranda, M. Calderón, and J. Favela. Anxiety detection using wearable monitoring. In *Proceedings of the 5th Mexican Conference on Human-Computer Interaction*, MexIHC '14, page 34–41, New York, NY, USA, 2014. Association for Computing Machinery.

[259] D. Miyamoto, T. Iimura, G. Blanc, H. Tazaki, and Y. Kadobayashi. Eyebit: Eye-tracking approach for enforcing phishing prevention habits. In *2014 Third International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, pages 56–65, USA, Sep. 2014. IEEE.

[260] F. Monrose and A. D. Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation computer systems*, 16(4):351–359, 2000.

[261] P. Morgado, N. Sousa, and J. J. Cerqueira. The impact of stress in decision making in the context of uncertainty. *Journal of Neuroscience Research*, 93(6):839–847, 2015.

[262] P. R. Mosaly, L. M. Mazur, F. Yu, H. Guo, M. Derek, D. H. Laidlaw, C. Moore, L. B. Marks, and J. Mostafa. Relating task demand, mental effort and task difficulty with physicians' performance during interactions with electronic health records (ehrs). *International Journal of Human–Computer Interaction*, 34(5):467–475, 2018.

[263] R. Moskovitch, C. Feher, A. Messerman, N. Kirschnick, T. Mustafic, A. Camtepe, B. Löhlein, U. Heister, S. Möller, L. Rokach, and Y. Elovici. Identity theft, computers and behavioral biometrics. In *Proceedings of the 2009 IEEE International Conference on Intelligence and Security Informatics*, ISI'09, page 155–160. IEEE Press, 2009.

[264] M. Motlagh and P. Bours. User identification based on eye gaze data. In *Proceedings of Norwegian Information Security Conference 2014*, NISK 2014, pages 1–9, Trondheim, Norway, 2014. Trondheim: Akademika.

[265] A. A. Moustafa, A. Bello, and A. Maurushat. The role of user behaviour in improving cyber security management. *Frontiers in Psychology*, page 1969, 2021.

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

[267] S. Mukhopadhyay and S. Nandi. Lpitrack: Eye movement pattern recognition algorithm and application to biometric identification. *Machine Learning*, 107(2):313–331, Feb 2018.

[268] E.-M. Nel. Opengazer: open-source gaze tracker for ordinary webcams, Accessed on 2.Dec.2020.

[269] J. Nielsen and K. Pernice. *Eye Tracking Web Usability*. New Riders, Berkeley, CA, USA, 2010.

[270] M. Nishigaki and D. Arai. A user authentication based on human reflexes using blind spot and saccade response. *International Journal of Biometrics*, 1(2):173–190, 2008.

[271] G. Notoatmodjo and C. Thomborson. Passwords and perceptions. In *Proceedings of the Seventh Australasian Conference on Information Security-Volume 98*, pages 71–78. Citeseer, 2009.

[272] N. Nugrahaningsih and M. Porta. Pupil size as a biometric trait. In V. Cantoni, D. Dimov, and M. Tistarelli, editors, *Biometric Authentication*, pages 222–233, Cham, 2014. Springer International Publishing.

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

[274] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, Dec 2003.

[275] F. G. Paas and J. J. Van Merriënboer. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors*, 35(4):737–743, 1993.

[276] S. Pabst, M. Brand, and O. T. Wolf. Stress and decision making: A few minutes make all the difference. *Behavioural Brain Research*, 250:39–45, 2013.

[277] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, page 141–144, New York, NY, USA, 2010. Association for Computing Machinery.

[278] T. Partala, M. Jokiniemi, and V. Surakka. Pupillary responses to emotionally provocative stimuli. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, pages 123–129, New York, NY, USA, 2000. ACM.

[279] K. Passyn and M. Sujan. Self-Accountability Emotions and Fear Appeals: Motivating Behavior. *Journal of Consumer Research*, 32(4):583–589, 03 2006.

[280] S. Pearman, J. Thomas, P. E. Naeini, H. Habib, L. Bauer, N. Christin, L. F. Cranor, S. Egelman, and A. Forget. Let's go in for a closer look: Observing passwords in their natural habitat. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, page 295–310, New York, NY, USA, 2017. Association for Computing Machinery.

[281] S. Pearman, S. A. Zhang, L. Bauer, N. Christin, and L. F. Cranor. Why people (don't) use password managers effectively. In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security*, SOUPS'19, page 319–338, USA, 2019. USENIX Association.

[282] V. Pejović, T. Matkovič, and M. Ciglarič. Wireless ranging for contactless cognitive load inference in ubiquitous computing. *International Journal of Human–Computer Interaction*, 37(19):1849–1873, 2021.

[283] K. Pfeffel, P. Ulsamer, and N. H. Müller. Where the user does look when reading phishing mails – an eye-tracking study. In P. Zaphiris and A. Ioannou, editors, *Learning and Collaboration Technologies. Designing Learning Experiences*, pages 277–287, Cham, 2019. Springer International Publishing.

[284] K. Pfeuffer, Y. Abdrabou, A. Esteves, R. Rivu, Y. Abdelrahman, S. Meitner, A. Saadi, and F. Alt. Artention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics*, 95:1–12, 2021.

[285] K. Pfeuffer, M. J. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt. Behavioural biometrics in vr: Identifying people from body motion and relations in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI â€™19, New York, NY, USA, 2019. Association for Computing Machinery.

[286] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 261–270, New York, NY, USA, 2013. ACM.

[287] A. Poole and L. Ball. *Eye tracking in human-computer interaction and usability research: Current status and future prospects*, pages 211–219. 01 2006.

[288] A. Poole, L. J. Ball, and P. Phillips. In search of salience: A response-time and eye-movement analysis of bookmark recognition. In *People and computers XVIII—Design for life*, pages 363–378. Springer, 2005.

[289] M. Porta, P. Dondi, N. Zangrandi, and L. Lombardi. Gaze-based biometrics from free observation of moving elements. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(1):85–96, 2022.

[290] C. Quintana-Nevárez, F. López-Orozco, and R. Florencia-Juárez. Biometric authentication based on eye movements by using scan-path comparison algorithms. In *Proceedings of the RCCS-SPIDTEC2 Workshop on International Regional Consortium for Foundations, Research and Spread of Emerging Technologies in Computing Sciences*, volume 2031, pages 33–38, Hannover, Germany, 2017. CEUR-WS.

[291] R. Radach, J. Hyona, and H. Deubel. *The mind's eye: Cognitive and applied aspects of eye movement research*. Elsevier, 2003.

[292] R. Radiah, V. Mäkelä, S. Prange, S. D. Rodriguez, R. Piening, Y. Zhou, K. Köhle, K. Pfeuffer, Y. Abdelrahman, M. Hoppe, A. Schmidt, and F. Alt. Remote vr studies: A framework for running virtual reality studies remotely via participant-owned hmds. *ACM Trans. Comput.-Hum. Interact.*, 28(6), nov 2021.

[293] K. Ragozin, Y. Pai, O. Augereau, K. Kise, J. Kerdels, and K. Kunze. Private reader: Using eye tracking to improve reading privacy in public spaces. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '19, New York, NY, USA, 2019. ACM.

[294] V. Rajanna and T. Hammond. Can gaze beat touch? a fitts' law evaluation of gaze, touch, and mouse inputs. *arXiv preprint arXiv:2208.01248*, 2022.

[295] V. Rajanna, A. Malla, R. Bhagat, and T. Hammond. Dygazepass: A gaze gesture-based dynamic authentication system to counter shoulder surfing and video analysis attacks. In *2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, pages 1–8, USA, Jan 2018. IEEE.

[296] V. Rajanna, S. Polsley, P. Taele, and T. Hammond. A gaze gesture-based user authentication system to counter shoulder-surfing attacks. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 1978–1986, New York, NY, USA, 2017. ACM.

[297] R. Rakshit, V. R. Reddy, and P. Deshpande. Emotion detection and recognition using hrv features derived from photoplethysmogram signals. In *Proceedings of the 2nd Workshop on Emotion Representations and Modelling for Companion Systems*, ERM4CT '16, New York, NY, USA, 2016. Association for Computing Machinery.

[298] G. Raptis, C. Katsini, M. Belk, C. Fidas, G. Samaras, and N. Avouris. Using eye gaze data and visual activities to infer human cognitive styles: Method and feasibility studies. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, pages 164–173, New York, NY, USA, 2017. ACM.

[299] V. Raudonis, G. Dervinis, A. Vilkauskas, A. Paulauskaitė-Taraševičienė, and G. Keršulytė-Raudonė. Evaluation of human emotion from eye motions. *International journal of advanced computer science and applications*, 4(8):79–84, 2013.

[300] K. P. Rayner. A.(1989). the psychology of reading englewood cliffs, 1989.

[301] Real User. Passfaces: Two factor authentication for the enterprise. Webpage, 2005. Retrieved August 20, 2019.

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

[303] K. Renaud and M. Warkentin. Using intervention mapping to breach the cyber-defense deficit. In *Proceedings of the 12th Annual Symposium on Information Assurance*, ASIA'17, pages 14–22, 2017.

[304] I. Rigas, E. Abdulin, and O. Komogortsev. Towards a multi-source fusion approach for eye movement-driven recognition. *Information Fusion*, 32:13–25, 2016. SI: Information Fusion in Biometrics.

[305] I. Rigas, G. Economou, and S. Fotopoulos. Biometric identification based on the eye movements and graph matching techniques. *Pattern Recognition Letters*, 33(6):786–792, 2012.

[306] I. Rigas and O. Komogortsev. Biometric Recognition via Fixation Density Maps. In I. Kakadiaris, W. Scheirer, and C. Busch, editors, *Biometric and Surveillance Technology for Human and Activity Identification XI*, volume 9075, pages 154–163, USA, 2014. International Society for Optics and Photonics, SPIE.

[307] I. Rigas and O. Komogortsev. Biometric recognition via probabilistic spatial projection of eye movement trajectories in dynamic visual environments. *IEEE Transactions on Information Forensics and Security*, 9(10):1743–1754, Oct 2014.

[308] I. Rigas, O. Komogortsev, and R. Shadmehr. Biometric recognition via eye movements: Saccadic vigor and acceleration cues. *ACM Transactions on Applied Perception*, 13(2):6:1–6:21, Jan. 2016.

[309] C. Rinn, K. Summers, E. Rhodes, J. Virothaisakun, and D. Chisnell. Password creation strategies across high-and low-literacy web users. *Proceedings of the Association for Information Science and Technology*, 52(1):1–9, 2015.

[310] R. Rivu, Y. Abdrabou, Y. Abdelrahman, K. Pfeuffer, D. Kern, C. Neuert, D. Buschek, and F. Alt. Did you understand this? leveraging gaze behavior to assess questionnaire comprehension. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '21 Short Papers, New York, NY, USA, 2021. Association for Computing Machinery.

[311] R. Romance, A. Nielsen-Rodríguez, J. Benítez-Porres, J. L. Chinchilla-Minguet, and H. Morente-Oria. Cognitive effects and educational possibilities of physical activity in sustainable cities. *Sustainability*, 10(7):2420, 2018.

[312] J. Rose, Y. Liu, and A. Awad. Biometric authentication using mouse and eye movement data. *Journal of Cyber Security and Mobility*, 6(1):1–16, 2017.

[313] V. Roth, K. Richter, and R. Freidinger. A pin-entry method resilient against shoulder surfing. In *Proceedings of the 11th ACM Conference on Computer and Communications Security*, CCS '04, page 236–245, New York, NY, USA, 2004. Association for Computing Machinery.

[314] A. Saad, M. Chukwu, and S. Schneegass. Communicating shoulder surfing attacks to users. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, pages 147–152, New York, NY, USA, 2018. ACM.

[315] A. Saad, D. H. Elkafrawy, S. Abdennadher, and S. Schneegass. Are they actually looking? identifying smartphones shoulder surfing through gaze estimation. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA 2020 Adjunct, New York, NY, USA, 2020. Association for Computing Machinery.

[316] U. Saeed. Eye movements during scene understanding for biometric identification. *Pattern Recognition Letters*, 82:190–195, 2016. SI: An Insight on Eye Biometrics.

[317] D. Sakai, M. Yamamoto, T. Nagamatsu, and S. Fukumori. Enter your pin code securely!: Utilization of personal difference of angle kappa. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, pages 317–318, New York, NY, USA, 2016. ACM.

[318] J. Salah, Y. Abdelrahman, Y. Abdrabou, K. Kassem, and S. Abdennadher. Exploring the usage of commercial bio-sensors for multitasking detection. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, page 265–277, New York, NY, USA, 2018. Association for Computing Machinery.

[319] H. Salehifar, P. Bayat, and M. Majd. Eye gesture blink password: A new authentication system with high memorable and maximum password length. *Multimedia Tools and Applications*, 78(12):16861–16885, Jun 2019.

[320] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, pages 71–78, New York, NY, USA, 2000. Association for Computing Machinery.

[321] T. Samura and H. Nishimura. Influence of keyboard difference on personal identification by keystroke dynamics in japanese free text typing. In *2012 Fifth International Conference on Emerging Trends in Engineering and Technology*, pages 30–35, 2012.

[322] S. Scherbaum and M. Dshemuchadse. Psychometrics of the continuous mind: Measuring cognitive sub-processes via mouse tracking. *Memory & Cognition*, 48(3):436–454, 2020.

[323] M. Seetharama, V. Paelke, and C. Röcker. Safetypin: Secure pin entry through eye tracking. In T. Tryfonas and I. Askoxylakis, editors, *Human Aspects of Information Security, Privacy, and Trust*, pages 426–435, Cham, 2015. Springer International Publishing.

[324] S. Seha, G. Papangelakis, D. Hatzinakos, A. Zandi, and F. Comeau. Improving eye movement biometrics using remote registration of eye blinking patterns. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP 2019, pages 2562–2566, USA, May 2019. IEEE.

[325] T. Seitz, M. Hartmann, J. Pfab, and S. Souque. Do differences in password policies prevent password reuse? In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 2056–2063, New York, NY, USA, 2017. Association for Computing Machinery.

[326] K. Sharma, E. Niforatos, M. Giannakos, and V. Kostakos. Assessing cognitive performance using physiological and facial features: Generalizing across contexts. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(3), sep 2020.

[327] R. Shay, L. Bauer, N. Christin, L. F. Cranor, A. Forget, S. Komanduri, M. L. Mazurek, W. Melicher, S. M. Segreti, and B. Ur. 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*, CHI '15, page 2903–2912, New York, NY, USA, 2015. Association for Computing Machinery.

[328] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. Can long passwords be secure and usable? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2927–2936, New York, NY, USA, 2014. Association for Computing Machinery.

[329] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. Designing password policies for strength and usability. *ACM Trans. Inf. Syst. Secur.*, 18(4), may 2016.

[330] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, and L. F. Cranor. Encountering stronger password requirements: User attitudes and behaviors. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, New York, NY, USA, 2010. Association for Computing Machinery.

[331] L. E. Sibert and R. J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 281–288, New York, NY, USA, 2000. ACM.

[332] D. Silver and A. Biggs. Keystroke and eye-tracking biometrics for user identification. In *Proceedings of International Conference on Artificial Intelligence*, ICAI 2006, pages 344–348, USA, 2006. CSREA Press.

[333] D. Silver, S. Jana, D. Boneh, E. Chen, and C. Jackson. Password managers: Attacks and defenses. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 449–464, San Diego, CA, Aug. 2014. USENIX Association.

[334] I. Sluganovic, M. Roeschlin, K. Rasmussen, and I. Martinovic. Using reflexive eye movements for fast challenge-response authentication. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 1056–1067, New York, NY, USA, 2016. ACM.

[335] I. Sluganovic, M. Roeschlin, K. B. Rasmussen, and I. Martinovic. Analysis of reflexive eye movements for fast replay-resistant biometric authentication. *ACM Transactions on Privacy and Security*, 22(1):4:1–4:30, Nov 2018.

[336] C. Song, A. Wang, K. Ren, and W. Xu. Eyeveri: A secure and usable approach for smartphone user authentication. In *IEEE International Conference on Computer Communication*, INFOCOM'16, pages 1–9, USA, April 2016. IEEE.

[337] A. C. Squicciarini, M. Shehab, and F. Paci. Collective privacy management in social networks. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, page 521–530, New York, NY, USA, 2009. Association for Computing Machinery.

[338] A. Srivastava. Biometric identification system using eye movement analysis. *International Journal of Engineering Science & Advanced Research*, 3(1):77–83, 2017.

[339] N. Srivastava, U. Agrawal, S. Roy, and U. Tiwary. Human identification using linear multiclass svm and eye movement biometrics. In *Eighth International Conference on Contemporary Computing*, IC3 2015, pages 365–369, USA, Aug 2015. IEEE.

[340] J. M. Stanton, K. R. Stam, P. Mastrangelo, and J. Jolton. Analysis of end user security behaviors. *Comput. Secur.*, 24(2):124–133, Mar. 2005.

[341] J. Steil, I. Hagestedt, M. Huang, and A. Bulling. Privacy-aware eye tracking using differential privacy. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19, pages 27:1–27:9, New York, NY, USA, 2019. ACM.

[342] J. Steil, M. Koelle, W. Heuten, S. Boll, and A. Bulling. Privaceye: Privacy-preserving head-mounted eye tracking using egocentric scene image and eye movement features. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19, pages 26:1–26:10, New York, NY, USA, 2019. ACM.

[343] N. Steinfeld. "i agree to the terms and conditions": (how) do users read privacy policies online? an eye-tracking experiment. *Computers in Human Behavior*, 55:992–1000, 2016.

[344] S. H. Stewart, S. E. Buffett-Jerrott, and R. Kokaram. Heartbeat awareness and heart rate reactivity in anxiety sensitivity: A further investigation. *Journal of Anxiety Disorders*, 15(6):535–553, 2001.

[345] E. Stobert and R. Biddle. The password life cycle: User behaviour in managing passwords. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 243–255, Menlo Park, CA, July 2014. USENIX Association.

[346] B. Stock and M. Johns. Protecting users against xss-based password manager abuse. In *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*, ASIA CCS '14, page 183–194, New York, NY, USA, 2014. Association for Computing Machinery.

[347] Y. Sugano, X. Zhang, and A. Bulling. Aggregaze: Collective estimation of audience attention on public displays. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 821–831, New York, NY, USA, 2016. ACM.

[348] C. Sumner, A. Byers, R. Boochever, and G. J. Park. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *2012 11th International Conference on Machine Learning and Applications*, volume 2, pages 386–393, 2012.

[349] M. P. E. S. SUNKARA. Pupil dilation as an indicator of cognitive workload in human-computer interaction. HCII, 2003.

[350] J. Sweller. Cognitive load theory. In *Psychology of learning and motivation*, volume 55, pages 37–76. Elsevier, 2011.

[351] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel. Influence of mental stress on heart rate and heart rate variability. In J. Vander Sloten, P. Verdonck, M. Nyssen, and J. Haueisen, editors, *4th European Conference of the International Federation for Medical and Biological Engineering*, pages 1366–1369, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[352] M. Tanaka, A. Ishii, and Y. Watanabe. Effects of mental fatigue on brain activity and cognitive performance: a magnetoencephalography study. *Anat Physiol*, 4:1–5, 2015.

[353] V. Taneski, M. Heričko, and B. Brumen. Password security—no change in 35 years? In *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1360–1365. IEEE, 2014.

[354] C. C. Tappert, M. Villani, and S.-H. Cha. Keystroke biometric identification and authentication on long-text input. In *Behavioral biometrics for human identification: Intelligent applications*, pages 342–367. IGI global, 2010.

[355] K. Thomas, J. Pullman, K. Yeo, A. Raghunathan, P. G. Kelley, L. Invernizzi, B. Benko, T. Pietraszek, S. Patel, D. Boneh, et al. Protecting accounts from credential stuffing with password breach alerting. In *28th {USENIX} Security Symposium ({USENIX} Security 19)*, pages 1556–1571, 2019.

[356] A. Tiwari and R. Pal. Gaze-based graphical password using webcam. In V. Ganapathy, T. Jaeger, and R. Shyamasundar, editors, *Information Systems Security*, pages 448–461, Cham, 2018. Springer International Publishing.

[357] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia. Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64:131–148, 2020.

[358] J. Turner, J. Alexander, A. Bulling, D. Schmidt, and H. Gellersen. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, editors, *Human-Computer Interaction – INTERACT 2013*, pages 170–186, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[359] J. Turner, A. Bulling, J. Alexander, and H. Gellersen. Eye drop: An interaction concept for gaze-supported point-to-point content transfer. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, MUM '13, pages 37:1–37:4, New York, NY, USA, 2013. ACM.

[360] J. Turner, A. Bulling, J. Alexander, and H. Gellersen. Cross-device gaze-supported point-to-point content transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, pages 19–26, New York, NY, USA, 2014. ACM.

[361] S. Uellenbeck, M. Dürmuth, C. Wolf, and T. Holz. Quantifying the security of graphical passwords: The case of android unlock patterns. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS â€™13, page 161â€“172, New York, NY, USA, 2013. Association for Computing Machinery.

[362] B. Ur, F. Alfieri, M. Aung, L. Bauer, N. Christin, J. Colnago, L. F. Cranor, H. Dixon, P. Emami Naeini, H. Habib, N. Johnson, and W. Melicher. Design and evaluation of a data-driven password meter. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3775–3786, New York, NY, USA, 2017. Association for Computing Machinery.

[363] B. Ur, J. Bees, S. M. Segreti, L. Bauer, N. Christin, and L. F. Cranor. Do users' perceptions of password security match reality? In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 3748–3760, New York, NY, USA, 2016. Association for Computing Machinery.

[364] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor. How does your password measure up? the effect of strength meters on password creation. In *21st USENIX Security Symposium (USENIX Security 12)*, pages 65–80, Bellevue, WA, Aug. 2012. USENIX Association.

[365] B. Ur, F. Noma, J. Bees, S. M. Segreti, R. Shay, L. Bauer, N. Christin, and L. F. Cranor. "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)*, pages 123–140, Ottawa, July 2015. USENIX Association.

[366] S. Uzzaman and S. Joordens. The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition*, 20(4):1882–1886, 2011.

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

[368] F. Vella, I. Infantino, and G. Scardino. Person identification through entropy oriented mean shift clustering of human gaze patterns. *Multimedia Tools and Applications*, 76(2):2289–2313, Jan 2017.

[369] J. Vickers. Gaze control in basketball foul shooting. In J. M. Findlay, R. Walker, and R. W. Kentridge, editors, *Eye Movement Research*, volume 6 of *Studies in Visual Information Processing*, pages 527–541. North-Holland, 1995.

[370] M. Vidal, A. Bulling, and H. Gellersen. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 439–448, New York, NY, USA, 2013. ACM.

[371] D. Vitonis and D. Hansen. Person identification using eye movements and post saccadic oscillations. In *Tenth International Conference on Signal-Image Technology and Internet-Based Systems*, pages 580–583, USA, Nov 2014. IEEE.

[372] L. M. Vizer, L. Zhou, and A. Sears. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, 67(10):870–886, 2009.

[373] Volvo. Volvo cars to deploy in-car cameras and intervention against intoxication distraction. `https://www.media.volvocars.com/global/en-gb/media/pressreleases/250015/volvo-cars-to-deploy-in-car-cameras-and-intervention-\against-intoxication-distraction`, 2019. accessed 19 December 2019.

[374] E. von Zezschwitz. *Risks and Potentials of Graphical and Gesture-based Authentication for Touchscreen Mobile Devices Balancing Usability and Security through User-centered Analysis and Design*. PhD

dissertation, Der FakultÃ¤t fÃ¼r Mathematik, Informatik und Statistik der Ludwig-Maximilians-UniversitÃ¤t MÃ¼nchen, 2016.

[375] E. von Zezschwitz, A. De Luca, B. Brunkow, and H. Hussmann. Swipin: Fast and secure pin-entry on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1403–1406, New York, NY, USA, 2015. ACM.

[376] E. von Zezschwitz, A. De Luca, and H. Hussmann. Survival of the shortest: A retrospective analysis of influencing factors on password composition. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, editors, *Human-Computer Interaction – INTERACT 2013*, pages 460–467, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[377] E. von Zezschwitz, A. De Luca, P. Janssen, and H. Hussmann. Easy to draw, but hard to trace? on the observability of grid-based (un)lock patterns. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI â€™15, page 2339â€“2342, New York, NY, USA, 2015. Association for Computing Machinery.

[378] E. Von Zezschwitz, P. Dunphy, and A. De Luca. Patterns in the wild: a field study of the usability of pattern and pin-based authentication on mobile devices. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 261–270, 2013.

[379] J. Wang, M. X. Huang, G. Ngai, and H. V. Leong. Are you stressed? your eyes and the mouse can tell. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 222–228, 2017.

[380] R. Wash, E. Rader, R. Berman, and Z. Wellmer. Understanding password choices: How frequently entered passwords are re-used across websites. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 175–188, Denver, CO, June 2016. USENIX Association.

[381] J. Weaver, K. Mock, and B. Hoanca. Gaze-based password authentication through automatic clustering of gaze points. In *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pages 2749–2754, USA, Oct 2011. IEEE.

[382] T. F. Wechsler, L.-M. Bahr, and A. Mühlberger. Can gaze behavior predict stress response and coping during acute psychosocial stress?–a virtual reality based eye tracking study (p. 764). *Nursing*, 20(5):697–706, 2019.

[383] M. Weir, S. Aggarwal, M. Collins, and H. Stern. 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*, CCS

'10, pages 162–175, New York, NY, USA, 2010. Association for Computing Machinery.

[384] R. Weiss and A. De Luca. Passshapes: Utilizing stroke based authentication to increase password memorability. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges*, NordiCHI â€™08, page 383â€“392, New York, NY, USA, 2008. Association for Computing Machinery.

[385] W. Wen, G. Liu, Z.-H. Mao, W. Huang, X. Zhang, H. Hu, J. Yang, and W. Jia. Toward constructing a real-time social anxiety evaluation system: Exploring effective heart rate features. *IEEE Transactions on Affective Computing*, 11(1):100–110, 2020.

[386] D. L. Wheeler. zxcvbn: Low-budget password strength estimation. In *25th {USENIX} Security Symposium ({USENIX} Security 16)*, pages 157–173, 2016.

[387] XLabs. Xlabs webcame eye tracker, Accessed on 2.Dec.2020.

[388] P. Xu, K. A. Ehinger, Y. Zhang, A. Finkelstein, S. R. Kulkarni, and J. Xiao. Turkergaze: Crowdsourcing saliency with webcam based eye tracking, Accessed on 2.Dec.2020.

[389] Y. Yamato and S. Takahashi. Gaze-based authentication method using graphical passwords featuring keypoints. In *Conference on Human-Computer Interaction (OzCHI'21)*, 2021.

[390] S. C.-H. Yang, D. M. Wolpert, and M. Lengyel. Theoretical perspectives on active sensing. *Current opinion in behavioral sciences*, 11:100–108, 2016.

[391] G. Ye, Z. Tang, D. Fang, X. Chen, K. I. Kim, B. Taylor, and Z. Wang. Cracking android pattern lock in five attempts. In *Proceedings of the 2017 Network and Distributed System Security Symposium 2017 (NDSS 17)*. Internet Society, 2017.

[392] G. Ye, Z. Tang, D. Fang, X. Chen, W. Wolff, A. Aviv, and Z. Wang. A video-based attack for android pattern lock. *ACM Transactions on Privacy and Security*, 21(4):19:1–19:31, July 2018.

[393] J. Yin, J. Sun, J. Li, and K. Liu. An effective gaze-based authentication method with the spatiotemporal feature of eye movement. *Sensors*, 22(8):3002, 2022.

[394] H. Yoon, T. Carmichael, and G. Tourassi. Gaze as a Biometric. In C. Mello-Thoms and M. Kupinski, editors, *Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment*, volume 9037, pages 39–45, USA, 2014. International Society for Optics and Photonics, SPIE.

[395] W. H. Zangemeister and L. Stark. Gaze latency: Variable interactions of head and eye latency. *Experimental Neurology*, 75(2):389–406, 1982.

[396] Y. Zhang, M. K. Chong, J. Müller, A. Bulling, and H. Gellersen. Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing*, 19(5):967–981, 2015.

[397] Y. Zhang, W. Hu, W. Xu, C. T. Chou, and J. Hu. Continuous authentication using eye movement response of implicit visual stimuli. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):177:1–177:22, Jan. 2018.

[398] Y. Zhang, J. Laurikkala, and M. Juhola. Biometric verification of a subject with eye movements, with special reference to temporal variability in saccades between a subject's measurements. *International Journal of Biometrics*, 6(1):75, 2014.

[399] Y. Zhang, J. Rasku, and M. Juhola. Biometric verification of subjects using saccade eye movements. *International Journal of Biometrics*, 4(4):317–337, Oct. 2012.

[400] R. Zhao and C. Yue. All your browser-saved passwords could belong to us: A security analysis and a cloud-based new design. In *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*, CODASPY '13, page 333–340, New York, NY, USA, 2013. Association for Computing Machinery.

[401] H. Zhou, V. Ferreira, T. Alves, K. Hawkey, and D. Reilly. Somebody is peeking!: A proximity and privacy aware tablet interface. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 1971–1976, New York, NY, USA, 2015. ACM.

[402] H. Zhou, K. Tearo, A. Waje, E. Alghamdi, T. Alves, V. Ferreira, K. Hawkey, and D. Reilly. Enhancing mobile content privacy with proxemics aware notifications and protection. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1362–1373, New York, NY, USA, 2016. ACM.

[403] J. Zhu, L. Ji, and C. Liu. Heart rate variability monitoring for emotion and disorders of emotion. *Physiological measurement*, 40(6):064004, 2019.

[404] M. E. Zurko and R. T. Simon. User-centered security. In *Proceedings of the 1996 workshop on New security paradigms*, pages 27–33. ACM, 1996.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12. Juli 2011, § 8, Abs. 2 Pkt. 5.)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

München, den 20.12.2022

Yasmeen Essam Abdrabou Mahmoud